

5-1-2016

# The Environment Constrains Successful Search Strategies in Natural Distributed Systems

Drew Levin

Follow this and additional works at: [https://digitalrepository.unm.edu/cs\\_etds](https://digitalrepository.unm.edu/cs_etds)

---

## Recommended Citation

Levin, Drew. "The Environment Constrains Successful Search Strategies in Natural Distributed Systems." (2016).  
[https://digitalrepository.unm.edu/cs\\_etds/32](https://digitalrepository.unm.edu/cs_etds/32)

This Dissertation is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Computer Science ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

Drew Faller Levin

---

Candidate

Computer Science

---

Department

This dissertation is approved, and it is acceptable in quality and form  
for publication:

*Approved by the Dissertation Committee:*

Stephanie Forrest, Chairperson

---

Melanie Moses

---

Lydia Tapia

---

Judy Cannon

# The Environment Constrains Successful Search Strategies in Natural Distributed Systems

by

**Drew Faller Levin**

B.S., Computer Science, Harvey Mudd College, 2002

DISSERTATION

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

Doctor of Philosophy  
Computer Science

The University of New Mexico

Albuquerque, New Mexico

March, 2016

©2016, Drew Faller Levin

# Dedication

*To my wife, Meghan, who made everything possible.*

# Acknowledgments

The path to this dissertation has been long and I have received support from so many along the way. I would not have finished without the support of my family, my friends, and my colleagues.

First, I thank my advisor, Stephanie Forrest, for her support and her belief in me. I end my time under your tutelage an improved writer, researcher, and scientist. Next, I thank the rest of my committee: Melanie Moses, Lydia Tapia, and Judy Cannon. Thank you for your time, energy, and support. I would not be here without you. I also thank all of my co-authors of my published papers. I single out Fred Koster for his continued support and encouragement throughout this process. Of my peers, my coauthors Hugh Mitchell, Soumya Banerjee, and Josh Hecker have been instrumental in helping me produce my published works.

My friends and my peers at UNM, in Albuquerque, and across the globe have helped make this experience so memorable. Thanks to my fellow members of the Adaptive Lab: Josh Karlin, George Bezerra, ThanhVu Nguyen, Michael Groat, Ben Edwards, George Stelle, and Eric Schulte. Thank you to my many other friends at UNM in Computer Science and Physics including David Mohr, Blake Anderson, Mark Olah, Mark Scully, Vamsi Potluru, Jonas Anderson, and Travis McIntyre. And thanks to my lifelong friends from Harvey Mudd: Vipul Lugade, Eric Verner, and Mark Unemori.

Finally, and most importantly, I give thanks to my family. First and foremost, thank you to my wife Meghan; without your sacrifices, none of this would have been possible. Thank you to my two children, Norah and Drake, for bringing joy to my life. And thank you to my parents and my brother Todd for your love and encouragement during this difficult journey. I love you all.

# The Environment Constrains Successful Search Strategies in Natural Distributed Systems

by

**Drew Faller Levin**

B.S., Computer Science, Harvey Mudd College, 2002

PhD., Computer Science, University of New Mexico, 2016

## Abstract

This dissertation investigates two natural systems that use distributed search algorithms and tests the hypothesis that the searchers' environment is a key constraint on an optimal algorithm. Natural instances of distributed autonomous systems of simple components exist in both biology and social systems. These systems have been honed through eons of evolution by natural selection to perform well in their environment. I examine two specific systems that use distributed methods to search and recruit individuals to locations of interest: T cells' search for pathogens in the human body and ants searching for food. Both systems are examples of time-constrained processes that require the distributed coordination of simple autonomous agents and interaction with their environment.

Taking common principles from both domains, the dissertation examines three distributed search strategies: uninformed random search, origin-based local recruitment, and chemical-based pheromone recruitment. Using both numerical and agent-based models, it evaluates the effectiveness of these strategies across two environ-

mental factors: the spatial clustering and temporal volatility of resources. The results demonstrate that both recruitment-based strategies (origin-based and chemical-based) suffer in environments of high resources dispersion and volatility. Conversely, uninformed random search performs better in these environments.

The results are relevant to certain algorithmic issues in swarm robotics. For example, it is expensive to implement chemical trails in a distributed physical system, and the dissertation shows that strategies using only local recruitment perform similarly in all environments. Also, origin-only algorithms are much easier to implement in a robotics context. Further, because each strategy examined in this dissertation performs best at one extreme of resource spatial distribution, the results establish that the most difficult environments for search are likely those with intermediate levels of clustering. Finally, the dissertation classifies the exact nature of the environmental trade-offs and presents methods to determine the best search strategy given knowledge of the environment.



# Contents

<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xvi</b>
<b>Glossary</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Overview . . . . .	4
1.2.1 T Cell Search for Pathogens . . . . .	5
1.2.2 Ant Foraging for Food . . . . .	6
1.3 Organization . . . . .	7
1.4 Contributions . . . . .	10
<b>2 Background</b>	<b>11</b>
2.1 Model Design Techniques . . . . .	12

Contents

2.1.1	Analytical Models . . . . .	13
2.1.2	Differential Equations . . . . .	13
2.1.3	Agent-Based Models . . . . .	15
2.2	Model Calibration . . . . .	16
2.2.1	Regression Techniques . . . . .	16
2.2.2	Genetic Algorithms . . . . .	17
2.3	Uncertainty Analysis . . . . .	17
2.3.1	Model Stochasticity (Aleatory Uncertainty) . . . . .	18
2.3.2	Bootstrapping (Epistemic Uncertainty) . . . . .	19
2.3.3	Sensitivity Analysis (Epistemic Uncertainty) . . . . .	19
2.4	Related Work . . . . .	20
2.4.1	Immunology . . . . .	21
2.4.2	Social Insects . . . . .	22
2.4.3	Robotics and Computation . . . . .	24
2.5	Summary . . . . .	25
<b>3</b>	<b>Preliminary Models of Influenza Infection in Lung Tissue</b>	<b>27</b>
3.1	Influenza Infection Model Design . . . . .	28
3.1.1	Modeling Decisions . . . . .	30
3.1.2	Model Populations and Parameters . . . . .	31
3.1.3	Experimental Data of <i>In Vitro</i> Influenza Infection . . . . .	32

## Contents

3.2	Model Implementations . . . . .	33
3.2.1	Delay Differential Equation Model . . . . .	33
3.2.2	Cellular Automata Model: <code>ma_virions</code> . . . . .	35
3.2.3	Agent-Based Model: <code>CyCells</code> . . . . .	37
3.3	Model Parameterization . . . . .	37
3.3.1	DDE Model Fits . . . . .	38
3.3.2	CA Model Fits . . . . .	40
3.4	Comparison of Influenza Strains . . . . .	41
3.4.1	DDE Results . . . . .	41
3.4.2	CA Results . . . . .	43
3.5	Extended Models to Parameterize T cell and Chemokine Secretion Rates	45
3.5.1	Estimating Chemokine Secretion Rates . . . . .	45
3.5.2	Estimating the T Cell Production Rate . . . . .	47
3.6	Discussion . . . . .	52
3.7	Summary . . . . .	54
<b>4</b>	<b>Chemokine Directed T Cell Search in the Human Lung</b>	<b>55</b>
4.1	Spatial Model of Chemokine Directed T Cell Search . . . . .	56
4.1.1	Model Implementation . . . . .	57
4.1.2	Model Parameters . . . . .	61
4.1.3	<code>CyCells</code> Implementation . . . . .	63

## Contents

4.1.4	Modeling Decisions . . . . .	64
4.2	Results . . . . .	65
4.2.1	T cell sensitivity to chemokine . . . . .	65
4.2.2	Chemokine combinations . . . . .	66
4.2.3	Primary Model Behavior and Variance . . . . .	68
4.2.4	Spatial Effects . . . . .	68
4.2.5	Sensitivity Analysis . . . . .	71
4.2.6	Window of Control . . . . .	73
4.3	Discussion . . . . .	78
4.4	Summary . . . . .	80
<b>5</b>	<b>Environmental Determination of Ant Foraging Strategies</b>	<b>82</b>
5.1	Spatial Model of Ant Foraging . . . . .	83
5.1.1	Foraging Strategies . . . . .	85
5.1.2	Foraging Environments . . . . .	89
5.1.3	Model Design and Implementation . . . . .	89
5.2	Experimental Results . . . . .	91
5.2.1	Fixed Parameters . . . . .	94
5.2.2	Variable Volatility . . . . .	96
5.3	Discussion . . . . .	98
5.3.1	Ant Foraging Dynamics . . . . .	98

*Contents*

5.3.2	Implications For Robot Swarms . . . . .	99
5.3.3	Comparison To Biological Ants . . . . .	100
5.3.4	Caveats . . . . .	101
5.4	Summary . . . . .	103
<b>6</b>	<b>Analytical Model of Ant Foraging Behavior</b>	<b>104</b>
6.1	Analytical Model of Ant Foraging . . . . .	105
6.1.1	How Well Can Ants Do? . . . . .	107
6.1.2	Random Walks . . . . .	108
6.1.3	Ballistic Model of Uninformed Search . . . . .	109
6.1.4	Analytical Model of Nest Recruitment . . . . .	112
6.2	Model Results . . . . .	115
6.2.1	Auto-Correlated Random Walk Versus Ballistic Motion . . . . .	115
6.2.2	Scaling of the Uninformed Foraging Rate . . . . .	117
6.2.3	Analytical Predictions of Ants Using Nest Recruitment . . . . .	120
6.2.4	Robustness of the Scout Population . . . . .	123
6.3	Assumptions and Conclusions of the Analytical Model . . . . .	125
6.4	Summary . . . . .	126
<b>7</b>	<b>Conclusion</b>	<b>128</b>
7.1	Summary of Results . . . . .	129

*Contents*

7.2	Future Work . . . . .	130
7.3	Final Remarks . . . . .	133
	<b>Appendices</b>	<b>134</b>
<b>A</b>	<b>CyCells Sensitivity Analysis</b>	<b>135</b>
A.1	One-factor-at-a-time sensitivity analysis . . . . .	135
A.1.1	Stable Parameters . . . . .	136
A.1.2	Difference in Peak Only . . . . .	145
A.1.3	Sensitive Parameters . . . . .	147
A.2	Partial rank correlation coefficient sensitivity analysis . . . . .	151
A.2.1	PRCC Design and Implementation . . . . .	151
A.2.2	PRCC Results . . . . .	152

# List of Figures

3.1	Diagram of Model Parameters . . . . .	29
3.2	Empirical viral and cytokine titers for three strains of influenza . . . . .	32
3.3	Best fit results for DDE and CA models . . . . .	42
3.4	Comparison of agent-based models . . . . .	44
3.5	Model 3.1 fits to data . . . . .	48
4.1	Visual representation of the model . . . . .	58
4.2	Varying T cell sensitivity to chemokine . . . . .	66
4.3	Effects of different chemokine combinations . . . . .	67
4.4	Model results . . . . .	69
4.5	Simulated sH1N1 infection . . . . .	70
4.6	Simulated infections of aH5N1, sH1N1, and pH1N1 . . . . .	71
4.7	Simulated infections with instant cell death . . . . .	73
5.1	Ant Foraging Model Conceptualization . . . . .	84
5.2	Ant Foraging Model States . . . . .	86

*List of Figures*

5.3	Search Performance vs. Volatility and Spatial Distribution . . . . .	92
5.4	Fixed Parameters Applied to Other Volatility Rates . . . . .	93
5.5	Fixed Parameters Applied to Other Pile Sizes . . . . .	95
5.6	Search Performance in Random Environments . . . . .	97
6.1	Analytical model design . . . . .	106
6.2	Success of correlated random walks . . . . .	116
6.3	Expected instantaneous foraging rate of uninformed ants . . . . .	118
6.4	Random Search Improves as the square root of the number of piles .	120
6.5	Expected instantaneous foraging rate of ants using nest recruitment	122
6.6	Foraging rate in a stable environment . . . . .	124
A.1	aH5N1 sensitivity analysis . . . . .	137
A.2	sH1N1 sensitivity analysis . . . . .	139
A.3	pH1N1 sensitivity analysis . . . . .	141
A.4	PRCC sensitivity analysis . . . . .	150
A.5	PRCC sensitivity analysis for the viral secretion rate . . . . .	153
A.6	Exhaustive Growth . . . . .	154



# List of Tables

3.1	Input Parameters for Cellular Automata Model. . . . .	35
3.2	Best Fit Parameter Values for CA Model . . . . .	36
3.3	Best Fit Parameter Values for DDE Model . . . . .	39
3.4	Empirical cytokine titers for three strains of influenza . . . . .	46
3.5	Root mean squared error of extended model fits . . . . .	48
3.6	Parameters for Extended DDE Model 3.3 (Equation 3.3) . . . . .	49
3.7	Parameters for T cell production DDE model 3.4 (Equation 3.4) . . . . .	51
3.8	Strain-specific parameters . . . . .	53
4.1	CyCells model parameters . . . . .	60
4.2	Sensitivity Analysis Results . . . . .	62
4.3	Linear fits to model runs for $I(t)$ and $D(t)$ . . . . .	76
5.1	Ant Parameters Tuned by the GA . . . . .	87
5.2	Ant Model Environment Parameters . . . . .	91

*List of Tables*

6.1	Scaling relationships of ant foraging . . . . .	121
A.1	Latin hypercube sampling distributions . . . . .	152

# Glossary

ABM	Agent-based model. A modeling approach where each actor or ‘agent’ is modeled explicitly.
aH5N1	Avian influenza strain A/Hong Kong/483/97. It is known for its high rates of morbidity and low rates of transmissibility.
alveoli	Tiny air sacs that make up the volume of the lung.
antibody	Immune protein used to recognize foreign substances such as antigens and pathogens.
antigen	Molecules recognized by immune antibodies that when detected will stimulate an immune response.
apoptosis	The process of cellular death.
CA	Cellular automata model. A modeling approach where each site on a grid or lattice-type structure is defined by a state. A site’s transition to a different state is often governed by the current states of neighboring sites.
chemokine	Describes a family of proteins secreted by epithelial cells infected by a foreign pathogen. Responsible for signaling the immune

## *Glossary*

	response and directing the movement of immune agents called leukocytes.
chemotaxis	The process of a leukocyte climbing a chemokine gradient.
CPFA	Central-place foraging algorithm. An agent-based model of ant foraging developed by Hecker (2015) [1]. Central-place refers to the existence of a single nest location.
CyCells	Three dimensional agent-based modeling environment of immune processes developed by Warrender (2006)
cytokine	A family of immune system signaling molecules. Chemokines make up a subset of cytokine molecules. [2] and extended by me ( <a href="https://github.com/drewlevin/cycells">github.com/drewlevin/cycells</a> ).
DDE	Delay differential equation. A differential equation that includes delay terms that reference past population levels.
dendritic cell	Immune cells that bind to pathogen and transport it to lymph nodes where it can be presented to T cells for detection.
dispersion	In the context of this dissertation, dispersion refers to a spread spatial distribution of resources. The opposite of a clustered environment.
epithelial cell	A generic healthy cell, vulnerable to infection. Organized in monolayers.
FOI	Focus of infection. The location of a spreading viral or bacterial infection.

## Glossary

GA	Genetic algorithm. A function optimization process that uses biological techniques such as genetic crossover and mutation to ‘breed’ an optimal solution.
IgM	Innate immune response molecule Immunoglobulin M. A specific type of antibody.
<i>in silico</i>	A process simulated in a computer.
<i>in vitro</i>	Cellular processes in a laboratory setting, usually in a small well of epithelial cells.
<i>in vivo</i>	Cellular processes inside a living organism.
interferon	A specific cytokine signaling molecule known to inhibit viral replication.
IP-10	A specific type of chemokine, also known as CXCL10.
leukocyte	A generic term for immune cells including T cells and B cells.
LHS	Latin hypercube sampling. A method for generating a set of points to efficiently sample a multidimensional space.
local interaction	In the context of this dissertation, local interaction refers to direct communication between two actors near each other.
lymph node	Regions in the body where lymphatic cells such as T cells originate and mature. The lymph network travels through the body’s lymph nodes.
ma_virions	Cellular automata virology modeling environment, part of the MASyV software suite [3].

## *Glossary*

MASyV	General purpose cellular automata modeling suite developed by Beauchemin (2006) [3]. Contains the <code>ma_virions</code> virology module.
monolayer	A contiguous surface of epithelial cells.
MOI	Multiplicity of infection. Refers to the virus particle to cell ratio. Used to describe the size of an initial infection.
nest recruitment	In the context of this dissertation, nest recruitment refers to an ant foraging strategy where a portion of the ant population remains in the nest ‘in reserve’ to be recruited later by informed scouts.
ODE	Ordinary differential equations. An equation that relates the rate of change of a quantity to its current value.
OFAT	One-factor-at-a-time sensitivity analysis. The method of varying a single model parameter in isolation to test its effects on a model.
pathogen	A foreign substance (usually bacterial or viral) inside the body.
peptide	A small chain of amino acids capable of being detected by the immune system.
PFU	Plaque forming units. A unit of measurement equal to the number of virus particles.
pH1N1	2009 pandemic influenza strain A/California/04/2009. Known for its high rates of virulence and morbidity in otherwise healthy individuals.

## *Glossary*

pheromone	A chemical signaling mechanism. In the context of this dissertation, ants that find promising sources of food can lay a pheromone trail back to the nest, allowing future ants to follow the trail back to the food.
PRCC	Partial rank correlation coefficient. A measurement relating the strength of a model's change in output to variation of a single parameter. Related to Spearman's $\rho$ rank correlation coefficient.
$R_0$	The viral replication rate. The measurement of how many new infected cells a single infected cell will create. Alternatively, the number of new virus particles a single virion will create.
random search	In the context of this dissertation, random search refers to ants searching for food using a simple random walk with no memory, communication, or recruitment. Also referred to as an 'uninformed search'.
RANTES	A specific type of chemokine, also known as CCL5.
resource	In the context of this dissertation, a resource is the goal of a search process. For T cell search, the 'resource' is a virus-infected cell. For ant foraging, a resource is food.
site fidelity	An individual ant's ability to remember and return to a promising food location.
stigmergy	A process of communication where agents signal each other by altering their environment rather than communicating directly. Ant foraging with pheromone trails is an example of stigmergy.
T cell	A immune cell responsible for recognizing and killing epithelial cells infected by antigen.

## *Glossary*

sH1N1	Seasonal influenza strain A/New Caledonia/20/1999. One of many seasonal influenza strains that cause moderate but controllable sickness.
uninformed search	In the context of this paper, uninformed search refers to ant and T cell search for resources using a random walk without communication or recruitment. Also referred to as a ‘random search’.
virion	A single virus particle.
volatility	In the context of this dissertation, volatility refers to the rate at which resources move inside the environment. I measure volatility in terms of round trips: specifically, how many times can an agent return to the resource’s location before it moves.



# Chapter 1

## Introduction

*"It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."*

— Albert Einstein

Designing scalable distributed autonomous systems is challenging. This thesis examines natural examples of search processes as inspiration for the design and evaluation of candidate distributed algorithms for use in computation. Specifically, I examine two known examples of distributed search in nature: T cell search for pathogens in the human lung and ant colony foraging for food resources in the vicinity of the nest. Using both analytical and computational modeling, I demonstrate how spatial and temporal challenges to efficient search processes arise in the context of specific environments.

## 1.1 Motivation

The past decade has borne witness to the rise of ubiquitous computing devices. Cars are monitored and controlled by multiple internal computers [4]. Home appliances such as refrigerators, ovens, and washers and dryers use small on-board computers to optimize performance [5, 6]. Cell phones have become mobile computers [7], and computers now use cell networks to communicate [8].

The increasing number of interconnected devices presents new computational challenges for designing the algorithms that communicate with and control them. Embedded system design focuses on reducing energy consumption to make cheaper, longer lasting devices [9–11]. Meanwhile, the increasing number of components communicating with each other strains conventional systems of control. Energy limitations of modern processor design can be avoided in part through multi-core architecture, yet communication between cores creates new bottlenecks [12]. The increased bandwidth and complexity of internet traffic requires new approaches to routing infrastructure [13]. As computers proliferate to more devices involved in our daily lives we will need new strategies to coordinate their behavior.

Most proposed solutions describe systems of centralized control. These systems usually have the advantages of complete information and global communication with each subcomponent. However, centralized mechanisms have several drawbacks. First, a control mechanism must be created as a separate entity from the devices it controls. This implies that the devices must be designed to interact with a central controller. A centralized control structure can also create an information bottleneck, where the amount of computation required at the central controller increases infeasibly with each device added to the system. Avoiding this bottleneck using techniques such as step-wise data aggregation forces computation back onto the individual power-constrained devices [14–16]. Finally, centralized structures represent

## *Chapter 1. Introduction*

a single point of failure, leaving them vulnerable to adversarial attacks or malfunctions.

This dissertation will explore systems of distributed communication with the goal of discovering efficient, scalable, and robust approaches to these emerging computational challenges. Distributed systems have several possible advantages. Independent devices may not have to increase their computation capability with the number of other devices in the system, allowing them to remain size invariant as the scale of the overall system increases. Because the devices can be kept small and cheap it is possible to have many redundant components, avoiding a single point of failure. Finally, decentralized algorithms allow devices to be created independently of a centralized controller. Thus, enabling deployments of new devices without a preexisting infrastructure for a unique controller.

Distributed systems are not a new concept in computer science. Publications regarding parallel computation can be found as early as 1958 [17]. Today there are numerous algorithms designed for parallel processing on multi-core machines, computing clusters, and super-computers [18–20]. These distributed algorithms have limitations specific to their design and intended domain. For example, canonical parallel algorithms for common problems such as shortest path search (Bellman-Ford) and sorting (Bitonic sort) are adaptations of preexisting serial algorithms. In these cases, the distributed algorithms were chosen over more established serial algorithms (Dijkstra’s algorithm and A\* for shortest path search, Quicksort and Merge sort for sorting) because they are more easily broken down into independent components. Further, because these algorithms are designed to aggregate their solution, they follow Amdahl’s Law [21] which states that the speedup of a parallel algorithm is limited by its serial portion. Finally, most parallel algorithms are designed without considering the computational limitations of individual components [22], an unrealistic assumption in the age of power limited devices such as sensor networks and

## *Chapter 1. Introduction*

swarm robotics.

In this thesis, I evaluate natural examples of scalable, robust, and efficient decentralized algorithms. Evolution has discovered many effective distributed solutions for complex tasks. Natural distributed systems use simple resource constrained components and do not rely on global control functions. I focus on two known examples of distributed search processes in nature: T cell search for pathogens and ant colony search for food resources, both in the context of their environment. Using a combination of analytical, numerical, and agent-based models, this dissertation addresses the following questions:

1. How do natural distributed search processes use communication to facilitate search?
2. How does the specific environment of the search process affect its performance?
3. Given knowledge of the environment, what are candidate efficient and scalable distributed search strategies?

## **1.2 Overview**

This dissertation examines two natural examples of autonomous distributed search processes: T cell search for pathogens in the human lung and ant search for food resources. Both systems are examples of time-constrained processes that require distributed coordination of simple autonomous agents and interaction with their environment. Because the goal of this dissertation is to discover common principles of natural distributed search processes and their environments, I examine these two systems in terms of their similarities. Specifically, I investigate three general distributed search strategies: random search, local recruitment, and chemical recruitment, and study how these strategies are affected by two different environmental conditions:

## Chapter 1. Introduction

resource distribution and volatility. I define these terms in the context of T cell search and ant foraging below.

### 1.2.1 T Cell Search for Pathogens

The immune response to infection from foreign pathogens is a highly general and complex process involving numerous countless components, each with a myriad of reactions to external stimuli. To clarify my studies and simplify the system, I focus specifically on the T cell response to influenza infection in the human lung. However, this process is generalizable to the immune response to other types of infection.

Upon infection, the immune system must locate and eradicate replicating foreign pathogens rapidly to ensure the survival of the host. A key component of this process is the recruitment and migration of T cells from the lymph node to the focus of infection (FOI). To navigate from the lymph node to the lung, T cells travel through the highly branching vascular network. At the FOI, cells infected with the foreign pathogen secrete a variety of chemical signals that diffuse across the cell monolayer (the two dimensional cellular surface). A subset of these molecules, called chemokines, play a strong role in managing the immune response to the spreading infection by allowing T cells to perform chemically-directed motion called *chemotaxis*.

Once in the lung tissue, T cells search for infected cells by performing what resembles an uninformed random walk. Cells infected with virus present peptides of the virus externally which are detectable by nearby T cells. Because the lung is large and T cells can only detect these infected cells locally, the search process requires chemical recruitment to help T cells locate the infection in a timely manner. Because chemokine molecules can diffuse across the cell monolayer, the secretion of these molecules by infected cells results in a wide gradient of chemical concentration with peaks at or near the FOI. Once detected, T cells can use a process called chemotaxis to climb this chemical gradient and locate the source of infection. This type of

## Chapter 1. Introduction

chemical recruitment is a stigmergic process in that agents do not communicate directly, but rather by leaving signals in the environment for other agents to follow.

In the context of T cell search for infection, the term *resource* refers to an infected cell. The spatial distribution of the infected cells is generally highly clustered around the initial point. Infection may begin from as little as a single virus, and subsequent infection expands locally from that point. On the other hand, the volatility of *resources* can vary depending on the type of infection. In the case of virulent infections, the location of the infected cells may be highly volatile as new cells become infected faster than T cells can locate them. More benign infections may be considered stable as the spread of infection may be effectively static when compared to the time scale of the T cell response.

In this dissertation, I examine infection by three different strains of influenza chosen for their observed phenotypic differences in virulence: H5N1 avian influenza (aH5N1), H1N1 seasonal influenza (sH1N1), and 2009 pandemic H1N1 influenza (pH1N1). I attempt to differentiate these strains using their replication rate,  $R_0$ , which represents the number of new infected cells a single infected cell will create.

### 1.2.2 Ant Foraging for Food

Ants must locate food resources in a timely and efficient manner. Not only must ants locate food, they must retrieve it back to their nest. To aid in this process, individual ants may coordinate their search with other ants from the same colony. Ant species have been observed to use a multitude of search strategies when foraging for food [23]. For simplicity, I focus on three specific strategies that subsume most observed strategies found in nature: random search, local (or nest) recruitment, and chemical-based pheromone recruitment.

While pheromone recruitment is often considered the canonical form of ant forag-

## *Chapter 1. Introduction*

ing, many ant species use other methods to locate their food. In areas of high resource availability, individual ants forage using correlated random walks with no inter-agent communication at all. Other ant species have been known to use a strategy I call nest recruitment. In this case, ants communicate only locally at the entrance to their nest. A sub-population of ants remain inactive in reserve at the nest, ready to be recruited. When an active ‘scout’ ant locates a patch of food, it returns to the nest and leads a proportion of the reserve population back to the food where they can retrieve it efficiently. Pheromone recruitment involves ants leaving a chemical trail of pheromones back to the nest from a promising location of food. Other ants may follow this trail back to the food location. As more ants follow a trail successfully, the trail becomes reinforced by subsequent trips. Once the food is collected in full, the trail is no longer reinforced and the pheromone decays through evaporation. Similar to how T cells follow a chemokine gradient, ants following a pheromone trail is an example of stigmergy, where information is shared by affecting the search environment.

Ants forage for many different types of resources, including plants, seeds, and even other insects. These food sources have different spatial profiles. Plant resources may be highly dispersed in space, while the location of other insects may be in highly clustered colonies. Similarly, the volatility distribution of resource may vary as well. Seeds may blow away at any moment in a short gust of wind, while nutritious leaves will stay attached to their tree for months on end. I use these environmental classifications to evaluate the advantages and limitations of the above foraging strategies.

### **1.3 Organization**

Analytical, numerical, and computational modeling let us consider processes that can’t be observed directly. To best reflect the behavior of the real system, models of biological systems are commonly parameterized by fitting the parameters so the

## *Chapter 1. Introduction*

models match empirical data. Once calibrated through this fitting process, these models can then be evaluated using uncertainty analysis techniques. As my thesis relies heavily on several types of modeling, I review the relevant modeling approaches, data fitting techniques, and uncertainty analysis methods in the first half Chapter 2.

T cell search for pathogens in the human lung and ant foraging for food resources are both complicated biological processes. Understanding of the underlying methods governing these search processes requires specific knowledge of the biological agents and mechanisms involved. I review relevant work specific to these fields in the second half of Chapter 2.

My thesis first considers the environmental causes of differences seen in the infection of various strains of influenza. While ant foraging can be observed naturally, T cell search in the human lung presents challenges for environment determination. To better understand the infection process, I examine the infection of the human lung by three different strains of influenza in Chapter 3. I use three different models to evaluate the functional differences between H5N1 avian influenza, H1N1 seasonal influenza, and 2009 pandemic H1N1 influenza. The models are consistent in predicting differing viral replication rates for each strain of influenza, and demonstrate how avian influenza may proliferate by suppressing the innate immune response.

Rapid recruitment of T cells to the focus of an infection is a complex, time-sensitive process that relies on chemotaxis. Extending the work from Chapter 3, I next examine the dynamics of the T cell response to infections of the same three strains of influenza in Chapter 4. 2009 pandemic influenza is known to present a unique challenge to the human immune system, sometimes proving fatal, and I attempt to determine what makes the pH1N1 influenza infection process different from that of aH5N1 and sH1N1. To do so, I design, calibrate, and use a spatial agent-based model (ABM) to examine the temporal and spatial effects of the T cell response to infection. The model shows how rapid expansion and spatial dispersion



## *Chapter 1. Introduction*

of virus presents challenges to T cell search dependent on chemotaxis.

T cell search for pathogens and ant colony foraging for food resources share many similarities. Both are examples of natural distributed systems using information to coordinate autonomous agents' search for resources, where the resources in question are food in the context of ants and the infection itself in the context of the T cell response. Like T cells using chemotaxis to locate high density areas of infection, ants often use pheromones to locate and exploit areas with a high density of food resources. Because work in Chapter 4 suggests that recruitment of T cells through chemotaxis can be detrimental in environments of high dispersion and volatility, we examine the effects of the spatial distribution and volatility of resources on ant foraging rates in Chapter 5. To do this, I adapt an established agent-based model of ant foraging behavior to include resource volatility and add a new non-pheromone nest recruitment method to compare to uninformed and pheromone-directed foraging. The model demonstrates that methods based on recruitment are less effective in environments of high spatial dispersion and high volatility of resources where simple random search is optimal. Further, nest recruitment, which relies only on local agent interactions, performs comparably to pheromone recruitment in all environments, suggesting a less computationally intensive approach to the design of ant colony optimization algorithms.

While simulations are useful for examining complex systems across a range of conditions, they do not allow for direct inspection of the modeled system's underlying scaling principles. In Chapter 6, I develop a simple analytical model of foraging that predicts foraging rates based on input parameters describing the environment. The model shows that random search is an optimal search strategy in a bounded space. The model confirms and explains an observed square root relationship between the uninformed foraging rate and the spatial dispersion of resources in ant foraging. Further, in the design of an analytical model of foraging using nest recruitment,

## *Chapter 1. Introduction*

I develop an equation to predict the optimal ratio between scouts and reserve ants given an environment. Finally, the model of foraging using nest recruitment confirms the results of the CPFA model and demonstrates that the most difficult environments for search are neither those of high or low dispersion, but those in the middle of the spectrum.

I close with a short section of concluding remarks in Chapter 7.

## **1.4 Contributions**

This dissertation examines how environmental factors affect natural distributed search processes. Through careful modeling and analysis, I show

1. Chemical-based search can perform poorly if the resources move more quickly than the chemical signal can diffuse or decay.
2. Increased volatility and spatial dispersion of resources is detrimental to both nest and pheromone recruitment-based search, while uninformed random search benefits from high resource dispersion.
3. Both local interactions and stigmergic signaling are effective communication strategies to facilitate distributed search. Further, nest recruitment performs comparably to pheromone recruitment in all environments while only using local interactions.
4. Comparison of observed strategy environment pairs in nature demonstrates that natural systems generally have evolved to use an optimal strategy given their environment.
5. Because each strategy examined in this dissertation prefers resource spatial distribution of one extreme, the most difficult environments for search are likely those of medium clustering.

# Chapter 2

## Background

Modeling provides a tool for understanding natural systems without the cost and constraints of a wet lab or a field site. Modeling also can be used to interpret and understand experimental data collected in the field. Models allow us to change parameters and measure values that are impossible to collect in an empirical study. For example, a model of the human immune system can estimate population values of numerous acting agents [24], values that would be impossible to measure in a wet lab without destroying the experiment. Similarly, a model of ant behavior can potentially track and measure every individual ant in a way that is impossible at a field site [25].

My work with the immune system and ant colonies uses models based on experimental data. Thus, it is important to design and calibrate the models in such a way that they reflect observed behavior. This is a three-step process. First, I design one or more models that capture the relevant mechanisms of the system of interest. Next, I use observed empirical data to choose values for model parameters such that the model dynamics match the observed behaviors. Finally, I use established uncertainty analysis techniques to examine the model dynamics and quantify the sensitivity of

the model output to the values of estimated parameters.

## 2.1 Model Design Techniques

The first step of model design is choosing the level of detail to represent the system of interest. There are several factors to consider during this process. Detailed complex models incorporate more details and also have the potential to recreate the observed behavior more closely. On the other hand, complex models are often so computationally intensive that evaluating them takes too long to be useful while analysis of complex analytical models proves intractable. Overly complex models also run the risk of ‘over-fitting’ [26], a scenario where the model replicates empirical data perfectly but cannot generalize to new observations. Although simpler models may not suffer from over-fitting or high computational complexity, they can fail to capture the relevant behavior of the system. Model design follows the principle of parsimony. To paraphrase Einstein: “Everything should be made as simple as possible, but no simpler.”

Here, I examine the relative strengths of three different modeling approaches: analytical models, ordinary differential equation (ODE) and agent-based models (ABMs). Analytical models use algebraic equations to describe the relationships between a system’s individual components. ODE models describe relationships between populations and the rate of change of those populations. ABMs track each individual explicitly and describe rules for interactions between individuals. Differential equation models have the advantage of being more computationally efficient than ABMs, but are limited to representing aggregate population behavior in their pure form and do not take spatial information into account. Agent-based models can simulate spatial behavior, but are generally more computationally expensive and can require more extensive parameterization. The trade-offs between equation-based and agent-based models have been examined previously [3, 27], so here I summarize each

type.

### 2.1.1 Analytical Models

While in modern times models are often thought of as complex computational or physical constructs, the use of simpler conceptual analytical models is well established. In this dissertation I use the term ‘analytical model’ to refer to the use of non-differential algebraic equations that describe the static relationships of the system of interest. For example, our basic understanding of kinetic physics can be represented by simple algebraic relationships as witnessed by Newton’s Second Law of Motion which states that the acceleration of an object is directly proportional to the net force acted upon it and inversely proportional to its own mass:

$$\vec{F} = m\vec{a} \tag{2.1}$$

The use of analytical equations to describe a system can be advantageous as the algebraic form of basic equations may lend themselves to quick and intuitive analysis.

Unfortunately, complex systems of interest often consist of time-ordered non-linear interactions between many modular components that cannot be reduced to simple algebraic terms. In these cases, other modeling approaches such as systems of differential equations or computational agent-based models are necessary.

### 2.1.2 Differential Equations

A population-based differential equation model describes one equation for each population type. Each equation relates the rate of change of the given population to the current level of each population. A simple example of an ODE model of two populations is the Lotka-Volterra model of the relationship between a predator and

## Chapter 2. Background

prey population [28]. The model is shown below where  $X$  represents the number of prey and  $Y$  represents the number of predators.

$$\begin{aligned}\frac{dX}{dt} &= \alpha X - \beta XY \\ \frac{dY}{dt} &= \delta XY - \gamma Y\end{aligned}\tag{2.2}$$

The change in the prey population depends positively on the current number of prey ( $\alpha X$ ) and negatively on the product of both populations ( $-\beta XY$ ). This shows that prey breed independently of predators, but are hunted at a rate that increases both with their own numbers and with those of the predator population. Conversely, predators depend on both their own numbers and the availability of prey in order to reproduce ( $\delta XY$ ), while dying off at a rate related only to their own numbers ( $-\gamma Y$ ).

Because the model depends on current population levels to determine future behavior, the model must be seeded with initial values. Once these are supplied, it is possible to extrapolate future population levels using variations of the Runge-Kutta algorithm [29, 30]. Versions of this algorithm are implemented in Matlab and in the GNU Scientific Library package.

Differential equations generally assume the system is well-mixed. In the above example, it is assumed that every predator has equal access to each individual prey. Because ODEs do not model individual interactions, systems that depend on isolated interactions, variable spatial densities, or transportation times are not good candidates to be modeled by ODEs. Because ODEs represent populations as a whole and are continuous, they do not scale down well to smaller discrete populations. For example, it is possible for an ODE model to represent a predator/prey system that contains a fraction of a predator (or prey), something impossible in a real system.

### 2.1.3 Agent-Based Models

Agent based models (ABMs) represent each individual explicitly. This approach allows for local interactions between individuals as well as spatial effects not seen in standard ODE models. ABMs represent some behaviors stochastically and are therefore not necessarily deterministic. Because of these capabilities, ABMs are good candidates to model systems with variable spatial densities, small discrete populations, and unique interactions between individuals. A main drawback of ABMs is the higher computational complexity that arises from representing each individual explicitly. For example, in a system with  $n$  agents where every agent interacts with every other agent, the ABM will have to account for approximately  $n^2$  interactions in each time step. Contrast this quadratic growth with a generalized ODE model that scales in complexity with the square of the number of *different types* of individuals. Further, because ABMs are defined through individual interactions, there is no direct way to derive closed-form analytical equations that describe behavior at the system level. Finally, because ABMs are stochastic by nature, evaluation of the model typically requires numerous repeat evaluations to determine model variance.

ABMs are defined by associating rules of behavior with each individual. These rules may include instructions for how to move in space, how to create new agents, how to remove agents, and how to interact with agents of specific types. Once behavioral rules are defined, models are initialized by arranging agents in space. ABMs are evaluated by following the rules of behavior for each agent before advancing to the next discrete time step.

There are many programs that allow for the definition and evaluation of ABMs. In this thesis, I use the both the MASyV and the CyCells modeling environment for the representation of the human immune system [2], and Joshua Hecker's CPFA swarm robot simulator [1] to implement my agent based models of ant foraging.

## 2.2 Model Calibration

Many models are designed to represent a particular system. In this case, it is important to match model output to the observed behavior of the real system. This is not a straightforward task. If a model contains nonlinear interactions there may be no analytical mapping of input parameters to output values and the only way to match the model's output to data is through extensive searches of the model's parameter space. Because this space expands exponentially with the number of parameters, an exhaustive search of all possibilities is infeasible for all but the most simple of models.

Model fitting can be treated as a function optimization problem. In this case, the function takes the model's free parameters as input and gives as output the difference between the model's behavior and the observed behavior of the real system. The problem of fitting the model is reduced to finding the parameter values that minimize this function. Fortunately, there are established methods for nonlinear function optimization. Common approaches include regression [31], genetic algorithms [32], and simulated annealing [33]. I use multiple methods in my research.

### 2.2.1 Regression Techniques

Regression (or gradient descent) algorithms work by measuring the gradient of the search space and proceeding in the direction of most improvement [31]. This process can be visualized as rolling a ball down a hill. The ball will roll down the hill until it reaches a valley and then stop. Gradient descent algorithms work well on smooth search spaces that have few local minima or maxima as the gradient leads the algorithm directly to the best solution. To implement gradient descent, I use Matlab's `nlinfit` function which is an implementation of the Levenberg-Marquardt gradient descent algorithm for nonlinear systems [34].



When successful, the gradient descent operates efficiently. Many models, however, have complex nonlinear relationships among their parameters. This leads to a complex search space with a large number of local minima, an environment where gradient descent techniques struggle.

### **2.2.2 Genetic Algorithms**

A genetic algorithms (GA) is a randomized function optimization technique modeled after the natural evolutionary process [32]. A GA represents points in the search space as a genome and output of the function being maximized at a specific point as that genome's fitness. A GA seeds itself with random genomes taken from a sampling of points in the search space. These genomes are evaluated and new ones are generated by combining and mutating features of the most successful points from the first generation. This process is repeated for a fixed number of iterations and then the best genome from the entire run is returned as the best-fit parameter setting. GAs do not depend on the local gradient of the search space and are therefore not limited to a single basin of attraction. The mutation and crossover operations of the GA allow the search to escape from local minima and find optima that standard regression techniques cannot. This allows GAs to be effective when searching over nonlinear surfaces where standard regression techniques fail.

## **2.3 Uncertainty Analysis**

Uncertainty in model analysis arises from two sources: aleatory and epistemic uncertainty [35]. Aleatory uncertainty describes the variation between evaluations of a non-deterministic model using the same parameter set. Because the model is non-deterministic, it is not guaranteed to produce the same result in subsequent runs. Non-deterministic models must be evaluated numerous times to obtain a measure of their inherent aleatory stochasticity. Epistemic uncertainty describes the uncertainty

## Chapter 2. Background

of the model results due to the uncertainty of the model's parameter values. There are multiple techniques to quantify the uncertainty of parameter values found using calibration techniques. Further, sensitivity analysis techniques must be employed to quantify the uncertainty of a model's output dependent on the uncertainty of the model parameter values. I detail my use of these methods below.

### 2.3.1 Model Stochasticity (Aleatory Uncertainty)

Deterministic models such as systems of differential equations always produce the same output for a given input. To obtain a measure of the aleatory uncertainty of a non-deterministic model, the model must be evaluated multiple times. Once a set of model outputs has been generated, statistical measurements can be used to quantify the amount of aleatory model stochasticity.

A common statistical measure of model variance is the standard error of the model output. The standard error is equal to the standard deviation of the sample output, divided by the number of samples, and establishes a confidence interval over the model output's true mean. The standard error is an accurate measure of aleatory uncertainty when the outputs are normally distributed. When the distribution of a model's output is not guaranteed to be normal, a common approach is to use the generated output values to establish a credible confidence interval. The model output is sorted and upper and lower sample points are chosen such that the interval defined by those points is centered about the median and contains the desired proportion of overall sample points. I define the aleatory uncertainty of my stochastic models in terms of the 96%<sup>1</sup> credible confidence interval calculated from a sample of 100 model runs.

---

<sup>1</sup>A centered 95% credible interval is impossible to generate from 100 samples.

### 2.3.2 Bootstrapping (Epistemic Uncertainty)

Both regression and GA fitting techniques return a single multi-dimensional point as the best-fit parameter set. In order to estimate confidence intervals for those parameters, a resampling technique is required to generate a distribution of estimates. I use bootstrapping with resampled residuals to generate confidence intervals [36]. Bootstrapping generates new empirical data sets under the assumption that the residuals (error terms) of the original parameter fit are uniformly distributed (the amount that the fit misses each data point is drawn from a uniform distribution). This assumption suggests that a given residual could be associated with any of the original data points. A synthetic dataset is generated by iteratively altering each original data point by a residual of the original fit selected uniformly at random with replacement. Then, the model is refit to the newly generated data and new parameter values are recorded. This process is repeated on the order of 1,000 times generating a new parameter estimate each time. The variance of each parameter value is then calculated from the set of generated parameter estimates.

### 2.3.3 Sensitivity Analysis (Epistemic Uncertainty)

A general sensitivity analysis serves two purposes. First, it allows observation of how the model behavior changes as a single parameter is varied. Second, it identifies which parameters affect the model strongly and which do not. There are many established approaches to both qualitative and quantitative sensitivity analyses. This thesis makes use of two specific types of sensitivity analyses: one-factor-at-a-time (OFAT) analysis [37] and partial rank correlation coefficient (PRCC) analysis [38] using Latin hypercube sampling (LHS) [39].

OFAT analysis is the process of measuring the change in model output while adjusting one model parameter in isolation. There are two main advantages of OFAT

## Chapter 2. Background

analysis over multiple-factor (or factorial) sensitivity approaches. First, manipulating a single parameter in isolation makes it possible to interpret the effect of that specific parameter on the model output. Second, because the parameters are manipulated one at a time, the number of evaluations required scales with to the number of parameters in the model itself and not with all possible combinations. Unfortunately, this second advantage is also a limitation. By varying one parameter at a time, OFAT analysis does a poor job at sampling the response of the model to the full parameter space. For example, if a reaction requires the presence of one of two existing components, varying one in isolation will not affect model output while varying both at the same time would.

PRCC analysis extends OFAT analysis by using LHS sampling to provide comprehensive coverage of the full parameter space. LHS sampling is a method of generating a fixed number of points that provide full and non-overlapping coverage of the parameter space. The resulting PRCC analysis gives the Spearman's  $\rho$  statistic and the  $p$  value of each parameter's rank correlation with the model's output. Because PRCC uses rank correlation instead of the traditional Pearson's linear correlation coefficient, it is more suitable for use with nonlinear systems than the basic OFAT analysis. Models of dynamical systems typically generate a series of output values versus time. PRCC analysis can be applied to full time series by calculating separate values for  $\rho$  and  $p$  at each time point of the model. At each point, the resulting  $p$  value can be used to demonstrate significance of the correlation, and when significant, the  $\rho$  value provides a measurement of the strength and direction of the model's correlation to the specified parameter.

## 2.4 Related Work

Many studies have successfully used both numerical and computational models for the study of natural distributed systems. This section reviews the most relevant

## Chapter 2. Background

prior work to this thesis and clarifies the differences and extensions to these works that are presented in the following chapters.

### 2.4.1 Immunology

The field of computational immunology uses both differential equation and agent-based models. Differential equation models of immune processes date back to as early as 1970 [40], and models of influenza as early as 1994 [41]. Alan Perelson has successfully employed ODE models in the study of HIV infection [42, 43]. Baccam (2006) [44] and Beauchemin (2008) [45] demonstrate the use of ODE models fit to empirical data to describe the dynamics of influenza infection in humans, and several groups published work using ODE models to describe the T cell response to influenza infection [46–49].

ODE models are generally *autonomous*, meaning their dynamics do not change over time and are memoryless, making it difficult to represent delayed effects such as virus incubation. Lee (2009) [50], along with my own work [24] presented in Chapter 3, use delay differential equations (DDEs) to represent time delays in the proposed model. DDE models can generate periodic dynamics that ODE models cannot, a common occurrence in influenza infections which contain incubation delays.

As noted in Section 2.1.2, ordinary and delay differential equation models assume a well-mixed population and usually ignore spatial effects such as diffusion. Therefore, researchers also use computational cellular automata and agent-based models to better explore the spatial constraints in the interactions between pathogens and the immune response.

There are many examples of computational models used in immunological research. Beauchemin (2005) [51] first demonstrates the accuracy of a simple cellular automata model of influenza infection when compared to a more complicated nu-

## Chapter 2. Background

merical model [41]. Beltman (2007) [52] uses a cellular Potts model, a lattice-based physical simulation of cellular structures, to investigate how T cells discover and interact with antigen-presenting dendritic cells (DC) in the lymph node. Zheng (2008) [53] extends this work by creating a lattice-based 3D agent-based model of T cell and DC movement, and Vroomans (2012) [54] uses a cellular potts model to examine the effect of chemotaxis on T cell and DC interactions. Textor (2014) [55] creates a multiscale agent-based model of random T cell movement in the lymph node with a stochastic model of T cell circulation.

While these works demonstrate the effectiveness of computational spatial models in the realm of immunology, none present a true agent-based model of T cell search for influenza infection on the surface of the human lung. I develop such a model and examine how T cells use chemotaxis to locate infections in Chapters 3 and 4.

### 2.4.2 Social Insects

Similar to the field of immunology, the study of social insects has an established history that uses both numerical and computational models to answer questions about collective behavior. Observation of real insect colonies has suggested possible strategies of search, foraging, and recruitment [56]. Because real insects are unpredictable and difficult to observe in the wild, computational models of these processes have proven invaluable for understanding colony level coordination.

Ants use a number of different strategies to share and utilize information gained from exploring their environments and thus improve foraging rates. At an individual level, ants employ *site fidelity* [57], the ability to remember promising locations of high food density and return directly to the same site without search. Ants may also exchange information through *local interactions*, such as direct recruitment of ants at the nest or directional reinforcement of ants on a trail [58]. The best known ant foraging strategy may be the use of stigmergic chemical signals, known as *pheromone*

## Chapter 2. Background

*trails* [59]. When an individual discovers food, it lays trails of pheromone on its way back to the nest. Future ants may follow these trails back to discovered sources of food.

Most research into ant foraging focuses on the use of pheromone trails to direct ants to locations of high food density. However, recent work has shown that pheromone-based foraging strategies may be sub-optimal in the wrong environment. For example, Evison (2008) [60] show that visual landmarks and pheromone trails contribute equally to an ant's ability to locate a previously encountered food source, and that the two may have a complementary function. Pheromone trails may also direct ants to poor food sources if they are found before a richer location can be detected [61, 62], although these traps can be avoided using repellent pheromone [63]. The drawbacks to chemical recruitment are not limited to ant foraging, nor to the problem of reinforcing the wrong path in a stable environment. My work in the domain of immunology shows how T cells that rely on chemical gradients to locate sources of infection can also become stuck when the infection spreads faster than the signal can diffuse (Chapter 4). These limitations of pheromone recruitment suggest the need for a more comprehensive understanding of foraging techniques.

Foraging strategies that work well in one location may work poorly in another. For example, division of labor in insect colonies may not be necessary when food is abundant [64]. Aside from pheromone trails, ants are known to use a wide range of other foraging strategies, as cataloged by Lanan (2014) [23]. The desert ant, *Aphenogaster cockerelli*, forages individually with site fidelity [65], and site fidelity has been shown to be a more effective foraging strategy than pheromone recruitment in some contexts [66]. The acorn ant, *Temnothorax albipennis*, uses tandem running where ants that have located a food source lead naive ants to the food, without using a detectable pheromone trail [67]. Ants such as the *Formica cinerea* establish long-term trunk trails, where massive numbers of ants follow one another to stable

sources of food [68]. The predatory *Pheidologeton diversus* raid smaller ant species and termites when colonies are discovered [69]. Given this diversity, there is a need for more complete investigations of these other foraging strategies, and there is a need for extending previous work examining which environments are best for which strategies [23, 70–72].

Previous work looked at the effect of spatial distribution of resources [25, 73, 74] and temporal availability of resources [75] on foraging strategies. I extend these studies by creating a spatial agent-based model that incorporates food depletion effects and developing an analytical model to classify the scaling relationships between foraging strategies and different aspects of their environment in Chapters 5 and 6.

### 2.4.3 Robotics and Computation

Social insects are notable for their ability to harness large populations of simple individuals without any apparent centralized control to solve complex problems, such as finding food and building nests [56]. Computer scientists have long been interested in ant foraging behaviors, particularly as inspiration for *ant colony optimization* (ACO) search and optimization algorithms [76] based on stigmergic signaling. ACO algorithms have been applied to many computational problems, including network routing [77], the Traveling Salesman Problem [78], task scheduling [79], and the multiple knapsack problem [80].

The use of naturally inspired distributed algorithms is of particular interest for the control of swarm robotic systems. Swarm robotics generally refers to the coordination of many autonomous resource-constrained robots towards the solution of a complex task, a very similar system to social insect colonies. As with naturally inspired algorithm design, many implementations of naturally inspired swarm robotic controllers are based on the concept of pheromones. For example, swarm researchers have constructed elaborate stigmergic communication mechanisms using



## *Chapter 2. Background*

an always-on ink pen and white paper flooring [81]; a tightly-coupled video camera, video projector, and vision processing system [82]; and a phosphorescent-painted floor combined with ultraviolet light emitters [83]. Robot swarms that make use of these stigmergic signals must either limit their domain to areas where they can read and manipulate the environment directly, or maintain a centralized controller that maintains the global state of the environment. Neither option is ideal for the design of a scalable distributed system.

Ants also use simpler, more primitive recruitment strategies such as tandem running and group raids, which include a local recruitment display to stimulate nest mates to return to high-quality food patches [84]. Robot swarms mimic these short-range recruitment strategies using robot-to-robot physical connections [85], nearest-neighbor local communication [86], and robot-chain path formation [87]. These swarms employ relatively simple communication schemes that do not require global coordination or preexisting infrastructure in order to collectively forage for resources or aggregate in target areas.

## **2.5 Summary**

This chapter examined the modeling design and evaluation techniques I will use to examine natural systems of distributed search. By understanding the trade-offs of analytical, differential equation, and agent-based modeling approaches, I am able to select a model type most appropriate to answer my research questions. Knowledge of the scientific methods required to calibrate and evaluate my models lends confidence to my results.

In order to frame the contributions of this dissertation, I reviewed the most relevant work in the field of computational immunology, social insect biology, and naturally inspired computation and robotic techniques. There are many examples of

## *Chapter 2. Background*

successful use of models to investigate immune processes, but no prior model of T cell search in the human lung such as I present in Chapters 3 and 4. Many studies of pheromone-based ant search exist, yet few investigate environmental effects and alternative foraging strategies such as I present in Chapters 5 6.

## Chapter 3

# Preliminary Models of Influenza Infection in Lung Tissue

The immune response to infection is an example of a natural autonomous distributed search process. Upon infection, immune agents must navigate the human vascular and lymph networks to rapidly locate and eliminate the spread of foreign pathogens. In this chapter I examine the processes that lead to different infection profiles resulting from infection of different strains of influenza. I use modeling to determine the underlying factors that cause different infection profiles resulting from infection of three different strains of influenza: avian influenza, seasonal influenza, and 2009 pandemic influenza. Because the *in vivo* infection process is impossible to observe directly, I use both numerical and spatial models fit to *in vitro* empirical data to obtain previously unpublished parameter values such as viral production rates, an infected cell's expected lifespan, and the viral incubation delay time.

The results show that the observed differences between the three strains of influenza are best described by the virus' replication rate and by the response to the body's antiviral defense. Subsequent model fits establish parameter estimates for the

production rates of T cells and for signaling molecules called chemokines. The collective results of this chapter characterize the factors of the environment that determine the effectiveness of the immune response. The T cell search process is examined in light of this information in Chapter 4.

The material described in this chapter is published in Mitchell (2011) [24] and the use of the words we/our refers to my coauthors. This chapter contains excerpts of the paper that is my research and excludes the wet lab experiments that were conducted by other authors.

### **3.1 Influenza Infection Model Design**

This section describes the creation and evaluation of three different models of influenza infection in the human lung. The model focuses on the spread of infection across the surface of bronchial alveoli, the tiny air sacs that make up the volume of the lung. While the volume of the lung consists of spherical alveoli, the spread of influenza throughout the lung occurs through the interconnected bronchiole-alveoli complex, which can be viewed as a connected monolayer (a two-dimensional surface of cells). Thus, we represent the lung as a two-dimensional sheet of healthy epithelial cells in our agent-based models (Section 3.1.1).

Figure 3.1 summarizes the infection process. An influenza infection can be initiated by a single virus particle, or virion. A virion in proximity to a healthy cell can enter the cell and replicate inside its nucleus. After an incubation period, the healthy cell expresses new virus particles that diffuse and infect neighboring cells. Infected cells also secrete an antiviral molecule called interferon which inhibits a virus' ability to replicate, although different virus strains react to the inhibition differently. After a finite window of secretion, the infected cell undergoes apoptosis (dies).

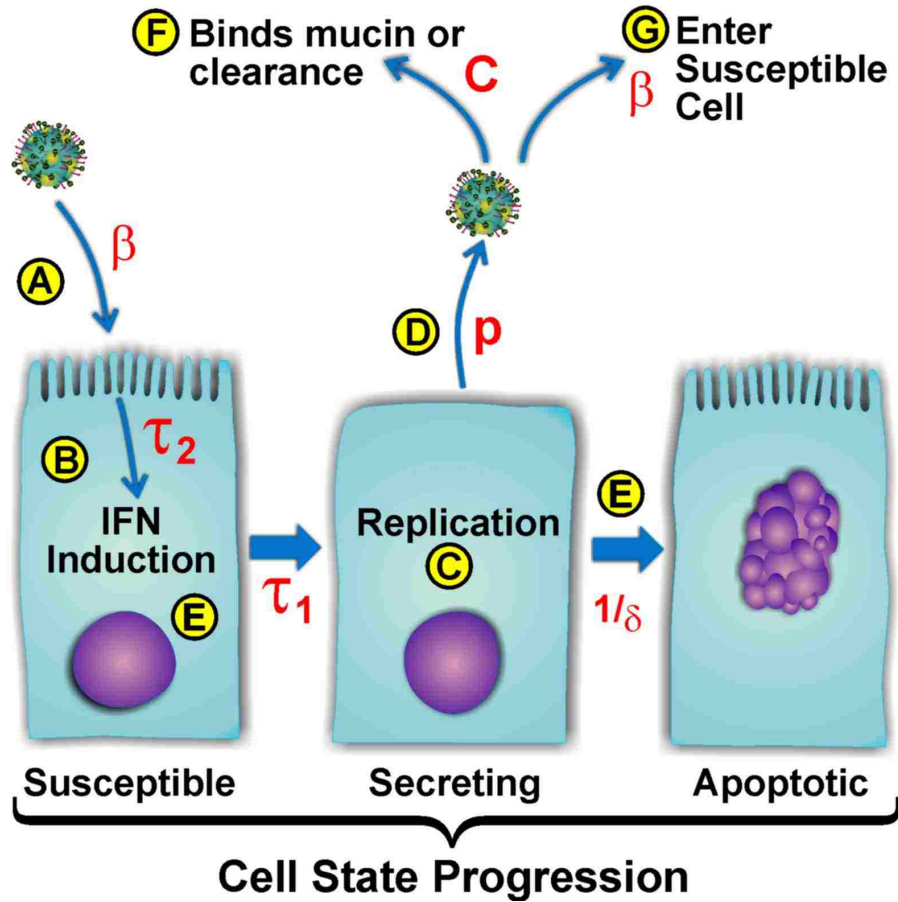


Figure 3.1: **Illustration of Model Parameters.** A: Viruses enter the susceptible cell population according to the infection rate coefficient  $\beta$ . B: cell-mediated inhibition of viral replication is initiated after interval  $\tau_2$ . The inhibition is governed by parameter  $e$ . C: infected cells begin secreting virus after incubation delay  $\tau_1$  fixed at 10h based on experimental data. D: virus is released at virion production rate  $p$ . E: virus secretion by a single cell is limited by  $1/\delta$  (production window). F: Free virus is cleared from the secreted virus pool by the decay rate  $c$  plus the absorption rate  $\beta$ . G: free virus is available to infect susceptible cells at the probability  $\beta$ . Figure taken from Mitchell (2011) [24].

The spread of the infection can be measured by its replication rate,  $R_0$ .  $R_0$  describes the number of new infected cells arising from a single infected cell. An infection where  $R_0$  is greater than one expands, while an infection where  $R_0$  is less than one does not (is controlled).

I examine the dynamics of this system with a delay differential equation (DDE) model and two agent-based models in the following sections. The use of multiple models lends confidence to the independent results of each, and justifies performing a bootstrapping sensitivity analysis only on the more computationally efficient DDE model.

### **3.1.1 Modeling Decisions**

Every model must simplify components of the system being modeled. The DDE model described in Section 3.2.1 has no explicit representation of spatial position and assumes all components of the system exist in the same location. For the purposes of our agent-based models, we represent the lung as a two dimensional sheet of epithelial cells.

Details of the limitations of differential equation models are discussed in Section 2.1.2. A main drawback of ODE and DDE models is their inability to represent spatial effects. Because populations are represented as single quantities, there is no knowledge of the position of individuals of the population and every free virus particle is inherently located on top of every healthy lung cell. The DDE model presented in Section 3.2.1 accounts for this limitation by using a scaling coefficient,  $\beta$ , to balance virus-cell infection events. Because the parameter is static in time, there is no dynamical scaling of the infection rate as spatial profile of the infection changes. For that purpose we complement the DDE model with two spatial models for accuracy.

The agent-based models represent the lung as a two-dimensional system for several reasons. Unlike the lymph node where dendritic cells and T cells navigate a three-dimensional volume, lung infection dynamics are confined to a thin tissue between capillary endothelial cells and alveolar epithelial cells. Although these alveoli are segregated by the lung's acinar structure on a small scale, we do not represent this in the model because the spreading of influenza eventually ignores the boundaries between acinii. Therefore our model does not incorporate this small-scale level of segregation.

### 3.1.2 Model Populations and Parameters

For simplicity, we model only the populations and parameters that affect the infection process (Fig. 3.1). The system includes four cell types and two molecule types.

Cells may either be *susceptible*, *incubating*, *secreting*, or *inactive*. Susceptible cells are functionally inert and may be infected by nearby virions. Once infected, healthy cells transition to the incubating state where they remain inert for a specified period of time equal to the viral incubation delay. After the delay is complete, incubating cells transition to the secreting state where they secrete virus at a rate determined by the virus' replication rate for a specified amount of time. Once the secretion window is complete, the cell dies and transitions to the inactive state where it remains for the remainder of the model.

There are two molecular types represented: virus and interferon. Virus is introduced by secreting cells, after which it diffuses across the cell monolayer until it either encounters and infects a susceptible cell or decays away. If a virion encounters a susceptible cell, infection is determined by the virus' infectivity rate and the concentration of interferon present. Interferon is produced by virus incubating cells and diffuses and decays passively at rates determined by its chosen parameters.

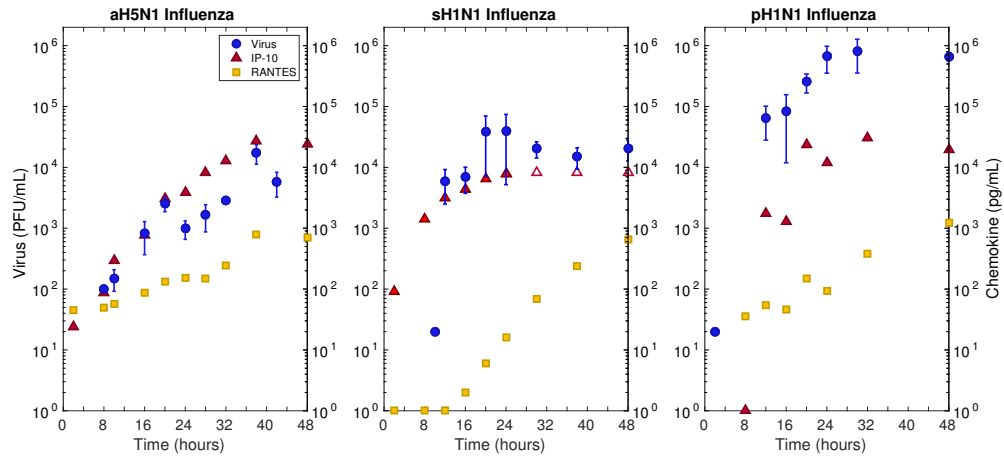


Figure 3.2: **Empirical viral and cytokine titers for three strains of influenza: Avian H5N1, Seasonal sH1N1, and Pandemic pH1N1.** Viral titer (blue circles) is in PFU/mL, and IP-10 (red triangle) and RANTES (yellow square) are shown in pg/mL. sH1N1 IP-10 secretion exceeded measurement accuracy above 8500 pg/mL and these three values (empty red triangles) were not included in the model fitting. An extended differential equation model from [24] was fit to IP-10 and RANTES data (Eq. 3.3, Fig. 3.5). These fits were used to obtain chemokine production values for use in the spatial CyCells model (Table 3.8). Human bronchial epithelial cells were infected at an MOI of 0.01 (10,000 virions) with one of the three strains of influenza. Figure reproduced from Levin (2016) [88].

### 3.1.3 Experimental Data of *In Vitro* Influenza Infection

To generate experimental data for model fitting, we infected human lung cells with three different strains of influenza *in vitro*. Three influenza strains were selected for comparison: avian strain A/Hong Kong/483/97 (aH5N1), human seasonal strain A/New Caledonia/20/1999 (sH1N1), and the A/California/04/2009 (pH1N1) swine-origin pandemic influenza strain. We infected wells of human bronchial cells with either avian, seasonal or pandemic influenza and measured viral titers at 4-8 hour intervals over the next 48 hours (Fig. 3.2).

Initial virus amounts were chosen to achieve an estimated multiplicity of infection



(MOI, i.e. the ratio of virus particles-to-cells) of 0.01 in all experiments, a very low value compared to most studies because we are interested in the early stages of infection. Due to experimentation error, one of two pandemic infections was initiated using a lower MOI of 0.001.

The exact process of is detailed in Mitchell (2011) [24].

## 3.2 Model Implementations

We model the system described in the previous section using a delay differential equation model, a cellular automata model, and a continuous-space agent-based model for completeness. While expressive and computationally efficient, the DDE model does not represent the spatial effects of virus and interferon diffusion. Because the spread of the focus of infection (FOI) is a dynamical spatial process, we complement the DDE model by extending and parameterizing an earlier cellular automata model of infection, `ma_virions` [3]. To confirm the results of the cellular automata model, we use the parameters from the cellular automata model fits to parameterize a new agent-based model designed in the CyCells modeling environment [2].

### 3.2.1 Delay Differential Equation Model

We used a DDE model that contains five populations: target cells, infected cells not yet producing virus, infected cells secreting virus, infectious virus, and an inhibitor representing the cellular antiviral response (Eq. 3.1). Model parameters correspond to viral infectivity, viral particle production rate, death rate of infected cells, strength of the antiviral factor, and the delay before antiviral factor production begins (Fig. 3.1). The structure of the DDE model we used is similar to Baccam (2006) [44]. A subsequent study derived an analytical approximation to fit the same data [89].

The model includes the following variables:  $T$  is the number of susceptible target

Chapter 3. Preliminary Models of Influenza Infection in Lung Tissue

cells;  $I_1$  is the population of infected cells not yet secreting virus,  $I_2$  is the amount of infected cells secreting virus,  $V$ , the free virus population, and  $F$  is the antiviral interferon that suppresses viral production (Fig. 3.1).

The DDE model contains one equation for each population. The rate of change of a population (signified using dot notation) is determined by an equation combining of the specific amounts of the represented populations, including itself. The target cell population,  $T$ , decreases at a rate proportional to the target cell population and the virus concentration ( $-\beta TV$ ), corresponding to the infection process. The Virus-incubating population,  $I_1$ , incorporates the infected target cells ( $\beta TV$ ) and loses cells whose incubation delay is at an end ( $-\beta T[t - \tau_1]V[t - \tau_1]$ ). Virus-secreting cells,  $I_2$ , gain the new cells from  $I_1$  and lose cells through natural decay ( $-\delta I_2$ ). Virus is produced proportional to the number of virus-secreting cells ( $I_2$ ) yet limited by the concentration of interferon ( $F$ ); when a virus infects a cell, it is removed from the system ( $-\beta TV$ ). After a delay, Interferon is produced at a rate proportional to the virus-incubating cell population ( $I_1[t - \tau_2]$ ).

$$\begin{aligned}
 \dot{T} &= -\beta TV \\
 \dot{I}_1 &= \beta TV - \beta T[t - \tau_1]V[t - \tau_1] \\
 \dot{I}_2 &= \beta T[t - \tau_1]V[t - \tau_1] - \delta I_2 \\
 \dot{V} &= \frac{p}{1 + eF} I_2 - \beta TV \\
 \dot{F} &= I_1[t - \tau_2]
 \end{aligned} \tag{3.1}$$

$\beta$  determines the rate at which virus infects target cells;  $p$  is the rate at which infected cells secrete new virus;  $\delta$  is the decay rate of virus secreting cells;  $e$  is the effectiveness of the anti-viral activity in suppressing viral production;  $\tau_2$  is the latent period before infected cells start producing inhibitor.

Input	Default Value	Source
Cell diameter	11 $\mu m$	Measured and [90]
Time step size	2 min.	Selected [90]
Virtual plate radius	45, 60, 525 cells	Measured
Virion diffusion rate	$3.18 \times 10^{-15} m^2/s$	[90]
Diffusion rate of antiviral factor	$3.18 \times 10^{-13} m^2/s$	Selected
MOI	0.001, 0.01	Selected
Length of eclipse phase	10 h	Figure 3.2
Antiviral factor decay rate	0.0	Selected
Infection rate coefficient	35.0	Selected
Viral decay rate	0.0	Measured

Table 3.1: **Input Parameters for Cellular Automata Model.** Parameters were taken directly from `ma_virions` model when possible and translated when necessary.

### 3.2.2 Cellular Automata Model: `ma_virions`

We use a cellular automata model to corroborate the results of the DDE model. Each site in the CA's lattice represents one epithelial cell. For simplicity, there is no cell division or differentiation over the course of the infection. Each location also keeps track of the amount of virus local to the site. Epithelial cells are initialized to the *uninfected* state, i.e., they do not contain any influenza. Virus is introduced either at the outset of the simulation at random locations in the grid with probability determined by the multiplicity of infection (MOI), or secreted by actively infected cells. Once virus enters a cell, the cell state becomes infected (but not secreting) until the end of the incubation period (virion release delay). After the incubation delay, the cell enters the secreting state and secretes virus according to the production rate (virion production rate). Each cell has an infected cell lifespan which is sampled from the exponential distribution, and at the end of its lifespan, the cell becomes inactive, no longer absorbing or secreting virus. Virus diffuses across the grid according to a discretized version of the diffusion equation.

To compute  $R_0$ , and for this case only, a single randomly selected cell was infected,

Data	Expt	lifespan (h)	$\tau_2$ (h)	$p$ (PFU/h)	$e$ (1/h)	$R_0$	MOI
avian	1	17.5	1.5	0.31	0.29	$3.0 \pm 1$	.01
	2	23.6	11.5	0.23	0.76	$3.6 \pm 5$	.01
seasonal	1	27.9	8.7	6.1	58.8	$15.4 \pm 1.4$	.01
	2	23.4	8.8	7.9	51.4	$19.1 \pm 1.8$	.01
pandemic	1	20.3	8.0	28.4	15.0	$62.2 \pm 5.5$	.001
	2	20.7	10.1	25.5	59.4	$40.1 \pm 4.0$	.01

Table 3.2: **Best Fit Parameter Values for Cellular Automata Model.** lifespan - infected cell lifespan,  $\tau_2$  - antiviral molecule release delay,  $p$  - virion release rate,  $e$  - antiviral molecule release rate,  $R_0$  - viral reproduction number, MOI - multiplicity of infection.  $\beta$ , the infection rate coefficient was set to 35.

and only that cell was allowed to secrete virions. In this way, infections resulting from virions produced by secondarily infected cells would not distort results.  $R_0$  was calculated as the mean of 500 independent simulations where each simulation ran for the secretion phase of the infected cell.

Beauchemin (2005) [51] and Beauchemin (2006) [3] give a detailed description of the model, and the software, `ma_virions`, is publicly available at <http://www.adaptive.cs.unm.edu/~mitchell/CA.tar>. `ma_virions` uses a virtual hexagonal grid of cells to approximate the two-dimensional configuration of real cells. We used version 0.10 of `ma_virions` to which we made minor modifications. Each modeled cell consists of a unique grid location and infection state, which can be one of the following: *uninfected*, *infected*, *secreting*, or *inactive*. Each cell tracks the virion and antiviral factor concentration in its immediate environment, and each cell is assigned a lifespan. The infected cell lifespan, virion release delay period, and antiviral release delay period for each cell were selected randomly from a Gaussian distribution whose mean,  $\mu$ , is an input parameter of the simulation and whose variance  $\sigma = \mu/10$ . The cellular antiviral response was implemented as a single diffusible antiviral factor secreted by infected cells.

The model has several input parameters which were held constant during all runs (Table 3.1). Free parameters derived from comparing simulation runs to the experimental data include: 1) lifespan of an infected cell, 2) delay in release of interferon, 3) virion production rate, and 4) interferon production rate (Table 3.2 and Section 3.3.2).

### 3.2.3 Agent-Based Model: CyCells

To rule out implementation errors and other model dependencies, we corroborated the `ma_virions` model using an independently designed and implemented simulation called CyCells [2]. CyCells is a continuous-space cellular automata hybrid agent-based model, in contrast to the strictly grid-based cellular automata model `ma_virions`. CyCells was calibrated to match the setup of `ma_virions` as closely as possible to test for any inconsistencies between the two modeling implementations.

Free parameters included the incubation period ( $\tau_1$ ), cell decay rate ( $\delta$ ), viral infectivity ( $\beta$ ), viral secretion rate ( $p$ ), viral diffusion rate, and viral decay rate ( $c$ ). These parameters were set without any model calibration using values that matched as closely as possible those obtained by modeling the first aH5N1 experiment using the `ma_virions` model (Table 3.2). The CyCells model was simulated for 48 hours in 10 second time steps.

## 3.3 Model Parameterization

We fit our models to each influenza strain's empirical data using the techniques described in Section 2.2 to best represent the real biological system of influenza infection. Best fit values for unknown parameters in both models were found using the GA (Section 2.2.2). Each individual in the population can be thought of as a guess by the algorithm about a potentially good set of parameters. The guess is tested

by the fitness function, which runs the model (either the DDE or CA model) with the parameter values set to match those specified by the individual. The quality of the guess is evaluated by comparing the model behavior with the experimental data, using the inverse of least squares fits on a logarithmic scale. The GA was used instead of more standard non-linear regression (Section 2.2.1) or exhaustive search due to the size and complexity of the search space, and in most cases the GA found better fits than those found using regression techniques upon comparison. Algorithm details were tailored to meet the needs of each model.

### 3.3.1 DDE Model Fits

Parameters in the DDE model ( $\beta, p, \delta, e, \tau_2$ ) were fit using the **GA**lib package (<http://lancet.mit.edu/ga>) with a population size of 100 individuals, 250 generations per run, tournament selection, and mutation and crossover rates of 0.01 and 0.5 respectively. The search procedure was restricted to parameter values that were within biologically plausible ranges. In some cases the ranges spanned several orders of magnitude, and in these cases parameters were represented in logarithmic units. Fitness was evaluated by calculating the least-squares fit of the model output (viral population) to the experimental data using four iterations of a full GA optimization routine. Each iteration consisted of 50 independent runs of the GA. Results with scores over two standard deviations below the mean were removed. The search bounds were narrowed based on the remaining population, and these bounds were used to constrain the next phase. At the end of the fourth phase, the best individual of the 200 runs was retained. We also tried fitting parameters with the Berkeley Madonna regression package (<http://www.berkeleymadonna.com>). Of the six data sets, fits to the data were better in five out of six data sets when using the GA. In each case, however, the fits from both methods were similar, and only the GA results are reported.

Data	Expt	$\tau_2$ (h)	$p$ (PFU/h)	$e$	$\beta$ (PFU·h) <sup>-1</sup>	$R_0$	MOI
avian	1	3.3 2.0 — 24	0.18 0.10 — 0.41	6.2e-7 6.0e-7 — 2.6e-5	4.8e-7 3.7e-7 — 7.8e-7	2.9	.01
	2	21.5 2.1 — 24	0.20 0.16 — 1.33	6.0e-7 6.0e-7 — 4.2e-5	5.3e-7 4.4e-7 — 6.3e-7	3.3	.01
seasonal	1	23.9 6.7 — 23.9	0.71 0.60 — 1.69	5.7e-5 1.6e-5 — 6.7e-5	6.3e-7 6.3e-7 — 6.9e-7	11.7	.01
	2	23.6 5.9 — 24	1.36 0.99 — 5.4	9.7e-5 7.9e-6 — 1.8e-4	6.1e-7 5.0e-7 — 6.6e-7	22.7	.01
pandemic	1	22.8 2.1 — 23.6	19.0 13.6 — 19.0	1.3e-3 2.1e-6 — 2.1e-3	2.1e-7 7.7e-8 — 3.7e-7	316	.001
	2	21.0 17.8 — 23.9	18 10.6 — 19.0	0.21 1.8e-4 — 1.7e-2	2.7e-6 4.8e-8 — 1.7e-6	305	.01

Table 3.3: **Best Fit Parameter Values for DDE Model.**  $\delta$  - infected cell decay rate,  $\tau_2$  - antiviral molecule release delay,  $p$  - virion release rate,  $e$  - antiviral molecule effectiveness,  $\beta$  - infection rate,  $R_0$  - viral reproduction number, MOI - multiplicity of infection, 95% confidence intervals shown below.

### Chapter 3. Preliminary Models of Influenza Infection in Lung Tissue

In each independent fit the decay rate of the virus-secreting cell population,  $\delta$ , hit the minimum search boundary of  $.06h^{-1}$ , which corresponds to maximum lifespan of 1,000 minutes. The incubation time of the virus-incubating cells,  $\tau_1$ , was fixed at 10 hours. Viral decay (which we assume encompasses all aspects of viral clearance), was found to be negligible in the presence of secreted mucus [24], and was therefore not included in the model. The model was initialized with  $10^6$  target cells and  $10^4$  virions (the first pH1N1 strain had an initial virus population of  $10^3$ ).

Confidence intervals were calculated by the resampling residuals bootstrapping method [36], which reshuffles the residuals (error terms) of the primary fit onto the values predicted by the primary fits at time points when the experimental data was measured (Section 2.3.2). This creates a new set of data points, to which the model is refit. This process was repeated a minimum of 1,000 times for each dataset, producing new parameter values for each new fit. These values were sorted, and cutoffs for the inner 95% of the values are listed as the confidence interval (Table 3.3).

#### 3.3.2 CA Model Fits

In the CA model, free parameters were fit using a GA implemented by Hugh Mitchell (<http://www.adaptive.cs.unm.edu/mitchell/CA.tar>). For each fitness evaluation of the GA, candidate parameter values selected by the GA were used to parameterize `ma_virions`, and the simulation was run for 1,440 time steps of 2 minutes each. The behavior of the simulation was then compared to the experimental data at corresponding time points, and the error, computed identically as for the DDE, was interpreted as the fitness.

For all experiments except that of the low MOI pH1N1 influenza, a radius of 45 cells (7,333 total cells) was used to find candidate best fit parameters. For the low MOI Swine flu study, a radius of 60 cells (13,057 total cells) was used and



candidate parameter sets were tested 10 times to account for stochastic variations in model behavior. The low MOI study was treated differently because there is greater variation among runs when infecting with fewer viruses; more cells and more runs compensated for the higher variability. For each run of the GA, the output values from the simulation were scaled up to reflect a full size plate, so that the output of the simulation could be properly compared to the experimental results. After best-fit parameters were found with the GA, the winning parameter sets were simulated on a plate size representative of the actual plate size used in the experiments, which was a radius of 525 cells (999,853 total cells) to confirm that the error calculations of the simulation on a small scale were close to the score on a large scale.

CyCells was defined on a grid  $360 \times 360 \times 10 \mu m$  in size. The grid was filled with 1,448 cells in a hexagonal arrangement. Thirteen of the cells (approximating an MOI of 0.01) chosen randomly were infected at the start of the simulation.

## 3.4 Comparison of Influenza Strains

Our quantitative models were designed to identify the contribution of key steps in the viral replication process (Fig. 3.1). Our objective was to determine unmeasurable parameters of strain-specific replication capacity by calculating the basic reproductive number  $R_0$  (the number of infected cells that results from one initially infected cell) and virion production rate per cell,  $p$ .

### 3.4.1 DDE Results

We evaluate the individual virus strains in part by measuring their replication rate,  $R_0$ . The formula for  $R_0$  was inspired by Beauchemin (2008) [45] and is calculated as the amount of new virus a single infected cell can produce over its lifetime times the average number of cells a single virion will infect (a value less than 1).

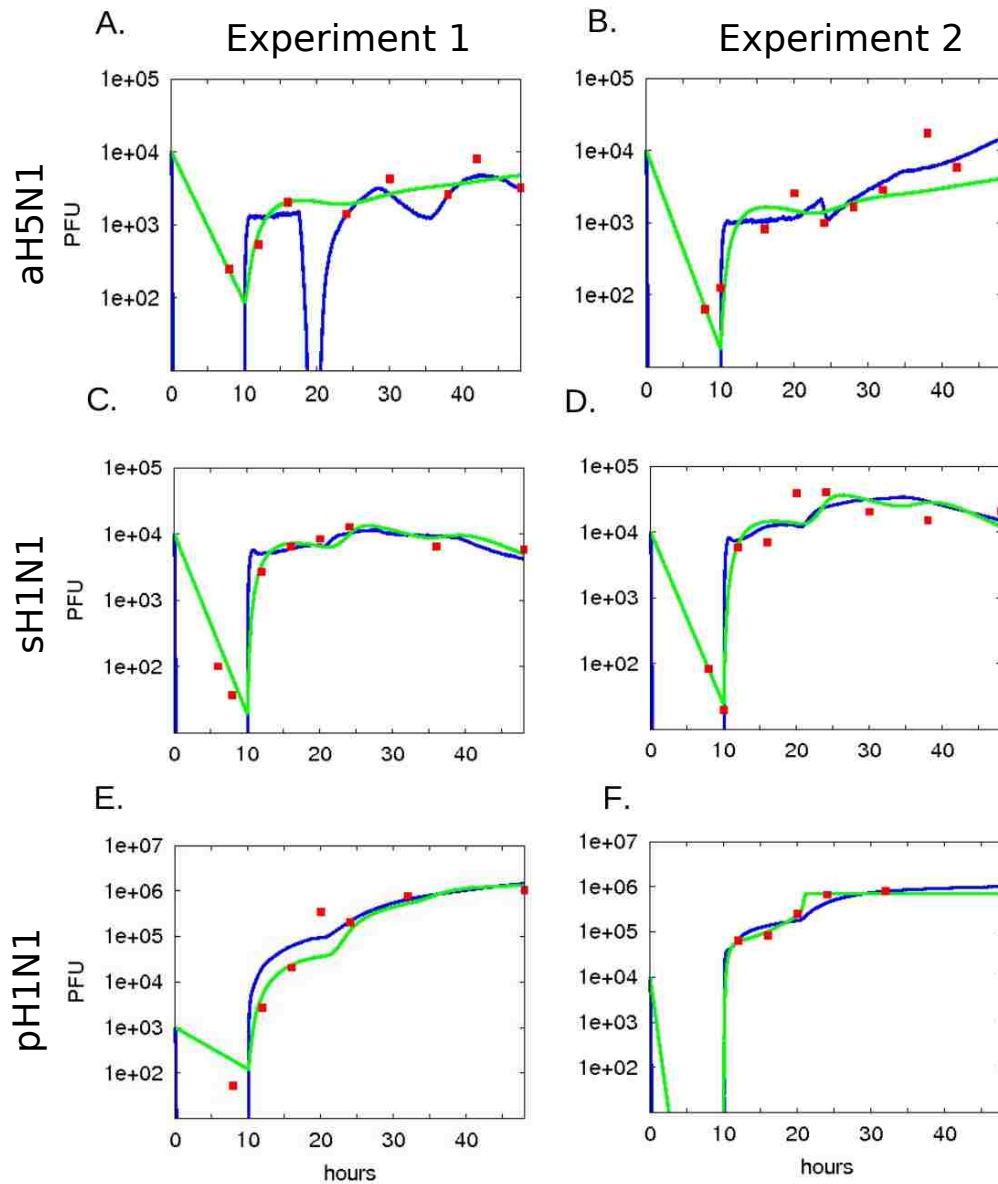


Figure 3.3: **Best fit results for differential equation and cellular automaton models.** Growth of each influenza strain was measured two times in isolation. A-B) aH5N1 influenza, C-D) sH1N1 seasonal influenza, E-F) pH1N1 2009 pandemic influenza. Model behavior using best-fit kinetic parameters in the DDE model from Eq. 3.1 (green lines) and CA model (blue lines). Experimental data are indicated with red squares.

$$R_0 = \frac{p\beta T_0}{\delta(c + \beta T_0)} \quad (3.2)$$

The DDE model results (Fig. 3.3, green lines) closely replicate the data points for each influenza strain, although the oscillatory pattern suggested by the data for the avian strain (Fig. 3.2) is not reflected in the model output. For both experiments,  $R_0$  was 2.9 and 3.3 for the avian strain; 12.0 and 23.0 for the seasonal strain; and 305 and 316 for the pandemic strain (Table 3.3). The rate of virion production per infected cell (plaque forming units per hour),  $p$ , was 0.18 and 0.20 PFU/hr for the avian strain, 0.71 and 1.36 PFU/hr for the seasonal strain, and 18 and 19 PFU/hr for the pandemic strain (Table 3.3). Thus, for each strain there was close inter-experimental agreement, and marked differences between the strains for the estimated values of both  $R_0$  and  $p$ . There were no significant differences in the infection rate  $\beta$  between the strains, in marked contrast to the differences in  $p$  and  $R_0$ , consistent with the hypothesis that strain differences in productivity are a function of intracellular replication efficiency, not cell entry efficiency or selectivity. The model accounted for the effectiveness of the antiviral cell defense ( $F$ ) with the coefficient  $e$ .  $e$  varies markedly among the strains ( $p < 0.001$ , F-test for each strain comparison), in rank order pandemic  $>$  seasonal  $>$  avian strains (Table 3.3), suggesting different abilities to inhibit or avoid the host antiviral response.

### 3.4.2 CA Results

In the CA model, fixed parameter values were taken either from the literature or laboratory results (Table 3.1), and the unknown parameters were determined using a GA (Table 3.2 and Section 3.3.2). Model estimates closely approximate experimental values at all time-points (Fig. 3.3, blue lines) and resemble the DDE model estimates. However, the CA model mimics the oscillatory secretion behavior seen in the avian strain data (Figure 3.3A). Rather than calculating  $R_0$  from the rate parameters (Eq.

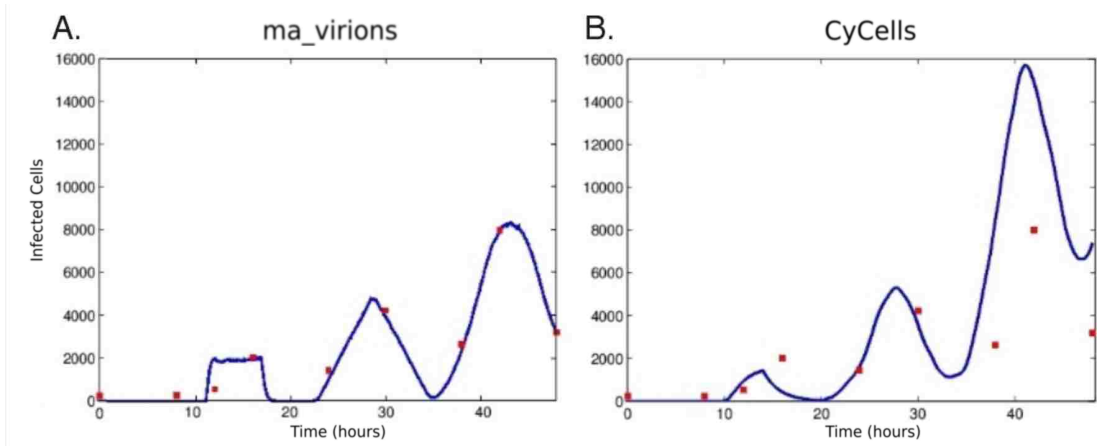


Figure 3.4: **Comparison of agent-based models.** The `ma_virions` model was used to obtain the best fit to the experimental virus data in Figure 3.2 (A). While direct conversions from `ma_virions` parameter values could not be made, approximately equivalent values were provided as input to the agent-based modeling system `CyCells`. The resulting output from `CyCells` is shown in (B). Blue lines are model output, red squares are data points. Note: To facilitate comparison of the agent-based models, model fitting, as well as graphing of the resulting output, were performed on a linear scale instead of a log scale as in Figure 3.3

3.2) as in the DDE model, we measured  $R_0$  directly in the CA simulation (Table 3.2). These values showed a pattern similar to the ODE model, namely that the avian strain had the lowest  $R_0$  (3.0 - 4.0),  $R_0$  for the seasonal strain was significantly higher, (12 - 15), and  $R_0$  for the pandemic strain was the highest (34 - 64). Infected cell virion production rates ( $p$ ) were 0.28 and 0.22 pfu/h for the avian strain; 3.8 and 1.4 for the seasonal strain; and 20.0 and 7.9 for the pandemic strain. The rank-order of  $p$  and  $R_0$  is the same for the three strains pandemic > seasonal > avian. Variation of each free parameter by 10% above and below the reported value resulted in negligible changes in the simulation results (results not shown).

To confirm that our CA results did not depend on subtle implementation details, we compared our simulation results using `CyCells`, which has been previously used for modeling tuberculosis infection [2]. By initializing `CyCells` similarly to `ma_virions`,

we demonstrated similar model output between the two models (Fig. 3.4). For the avian strain infection, the oscillatory pattern of infection was replicated by both models, although CyCells did not approximate the data points as well as the CA.

## 3.5 Extended Models to Parameterize T cell and Chemokine Secretion Rates

This chapter presents relevant models of immune system processes fit to experimental data to estimate values for parameterization of a more comprehensive agent-based model of chemokine-directed T cell search in the human lung. Sections 3.1 to 3.3 present models of viral infection and expansion, but do not include aspects of the T cell response. This section presents two new models examining the T cell search for infection. Virus-infected cells secrete signaling molecules called chemokines to recruit and aid the T cell immune response. There are no values for chemokine or T cell production rates in the literature, and both *in vivo* and *in vitro* rates cannot be observed directly. To obtain estimates of the production rates we fit two new differential equation models to empirical data and use the values of the resulting fits as parameters in the agent-based model described in the next chapter.

### 3.5.1 Estimating Chemokine Secretion Rates

To provide estimates of chemokine concentrations and secretion rates in lung tissue, chemokine levels were measured from the same *in vitro* experiments that provided virus counts (Fig. 3.2, Table 3.4) [24, 88]. Data revealed elevated levels of two chemokines: IP-10 (CXCL10) and RANTES (CCL5). IP-10 concentration increases were observed by 8h post-infection (p.i.), and RANTES by 16h p.i..

We estimate chemokine production rates,  $r$ , by extending the delay differential equation model of influenza infection described in Eq. 3.1 and adding one new equa-

aH5N1			sH1N1			pH1N1		
Time	CXCL10	CCL5	Time	CXCL10	CCL5	Time	CXCL10	CCL5
<i>hours</i>	<i>pg/mL</i>	<i>pg/mL</i>	<i>hours</i>	<i>pg/mL</i>	<i>pg/mL</i>	<i>hours</i>	<i>pg/mL</i>	<i>pg/mL</i>
0	440	1	0	94	62	0	179	45
2	391	0	2	24	45	12	1,835	55
8	1,723	1	8	87	50	16	1,349	46
12	3,462	1	10	296	57	20	23,458	150
16	4,618	2	16	770	87	24	12,073	93
20	6,807	6	20	3,048	134	32	30,700	380
24	8,164	16	24	3,901	151	48	19,814	1,224
30	> 8,500	68	28	8,261	150			
38	> 8,500	239	32	12,935	245			
48	> 8,500	655	38	26,970	780			
72	> 8,500	415	42	24,120	695			
			72	21,441	1,665			

Table 3.4: **Empirical cytokine titers for three strains of influenza.** Measured CXCL10 (IP-10) and CCL5 (RANTES) shown in pg/mL for three strains of influenza: Avian H5N1, Seasonal sH1N1, and Pandemic pH1N1. sH1N1 IP-10 secretion exceeded measurement accuracy above 8500 pg/mL. Human bronchial epithelial cells were infected at an MOI of 0.01 (10,000 virions) with one of the three strains of influenza. Basal media for chemokine secretion was collected at the given time intervals post infection.

tion ( $\dot{C} = rI_{1\tau_3} - dC$ ) to model chemokine production. Strain-specific values for  $r$  were found by fitting the equations to the experimental data in Table 3.4 using a GA to minimize the log squared error between the model and the data while holding the rest of the parameter values constant (Fig. 3.5, Tables 3.6 and 3.5). Next, 1,000 bootstrapping runs using resampled residuals (Section 2.3.3) were performed on each modeled strain to generate confidence intervals over  $r$ . The results of these fits are shown in Table 3.8.

$$\begin{aligned}
 \dot{T} &= -\beta TV \\
 \dot{I}_1 &= \beta TV - \beta T_{\tau_1} V_{\tau_1} \\
 \dot{I}_2 &= \beta T_{\tau_1} V_{\tau_1} - \delta I_2 \\
 \dot{V} &= \frac{p}{1 + e^F} I_2 - \beta TV \\
 \dot{F} &= I_{1\tau_2} \\
 \dot{C} &= rI_{1\tau_3} - dC
 \end{aligned} \tag{3.3}$$

$\tau$  subscript variables denote delay terms, signifying the value is the population quantity in existence at time  $t - \tau$ . Table 3.6 summarizes population and parameter values and descriptions.

### 3.5.2 Estimating the T Cell Production Rate

We calculate the rate of production,  $\sigma$ , of T cells using a differential equation model from Miao (2010) [47]. We extend the model system presented in Figure 3.1 with the addition of T cell production, navigation, and search (Fig. 3.1).

Equation 3.4 models T cell production and subsequent search over an area of infected lung tissue. T cells are produced in the lymph node where they enter the vascular network and travel to the lung. T cells that enter the lung near regions

Chapter 3. Preliminary Models of Influenza Infection in Lung Tissue

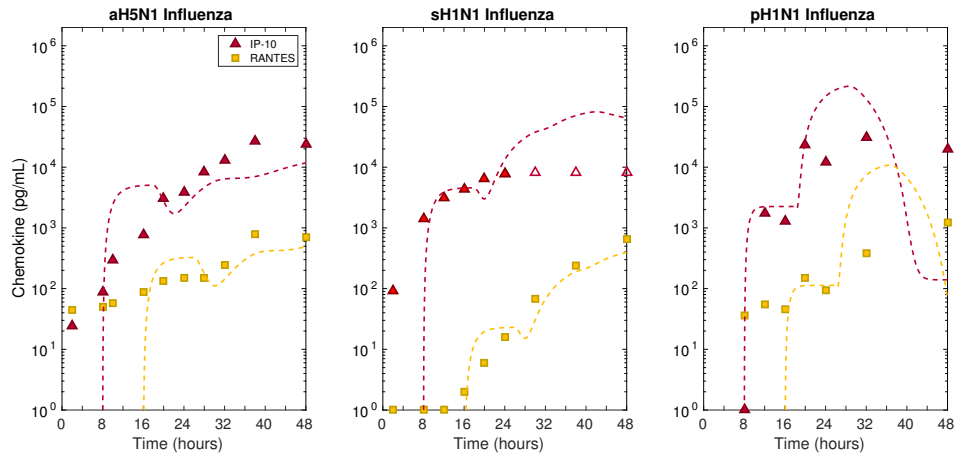


Figure 3.5: **Model 3.1 fits to data.** Model Eq. 3.1 was fit to experimental data (Table 3.4) using a genetic algorithm. sH1N1 IP-10 secretion exceeded measurement accuracy above 8500 pg/mL and these three values (empty red triangles) were not included in the model fitting. RMSE values of the fits can be seen in Table 3.5. IP-10 data and model fits shown in red (triangles), RANTES data and model fits shown in yellow (squares).

Strain	Chemokine	RMSE
aH5N1	RANTES	166
	IP-10	7,950
sH1N1	RANTES	90
	IP-10	2,732
pH1N1	RANTES	2,671
	IP-10	64,342

Table 3.5: **Root mean squared error of Preliminary Model 1 fits.** Preliminary Model 3.3 (Eq. 3.3) was fit to experimental data (Table 3.4) using a genetic algorithm as described in the main paper. The resulting model fits can be seen in Figure 3.5.

of infection exit the vascular network and enter the lung tissue to search for virus-infected cells. T cells that do not encounter infection recirculate through the lymph and vascular networks. T cells in the lung encounter infected cells and kill them. Cells infected by virus secrete more virus before eventually undergoing apoptosis,



Chapter 3. Preliminary Models of Influenza Infection in Lung Tissue

Population	Description	Initial Value
$T$	Healthy target epithelial cells	1,000,000
$I_1$	Virus-incubating cells	0
$I_2$	Virus-secreting cells	0
$V$	Free virus particles	10,000
$F$	Interferon quantity	0
$C$	Chemokine quantity (ng/mL)	0
Parameter	Description	Value
$\beta$	Viral infection rates	
	Avian	$5.3\text{e-}7 \text{ (PFU}\cdot\text{h)}^{-1}$
	Seasonal	$6.1\text{e-}7 \text{ (PFU}\cdot\text{h)}^{-1}$
	Pandemic	$2.7\text{e-}6 \text{ (PFU}\cdot\text{h)}^{-1}$
$p$	Viral production rates	
	Avian	0.20 (PFU/h)
	Seasonal	1.4 (PFU/h)
	Pandemic	18.3 (PFU/h)
$e$	Interferon strengths	
	Avian	$1.0\text{e-}8$
	Seasonal	$1.6\text{e-}6$
	Pandemic	$3.4\text{e-}3$
$\tau_2$	Interferon production delays	
	Avian	21.5 (h)
	Seasonal	23.6 (h)
	Pandemic	21.0 (h)
$\delta$	Virus-secreting cell decay rate	$0.6 \text{ (h}^{-1}\text{)}$
$d$	Chemokine decay rate	$1.386 \text{ (h}^{-1}\text{)}$
$\tau_1$	Viral incubation time	10 (h)
$\tau_3$	Chemokine production delays	
	IP-10	8 (h)
	RANTES	16 (h)

Table 3.6: **Parameters for Preliminary Model 3.3 (Equation 3.3)**. All parameters and populations are taken from [24] except for  $C$ ,  $d$ , and  $\tau_3$ . The value chosen for  $d$  corresponds to a 30 minute half-life and values for  $\tau_3$  were set from the observed chemokine data (Fig. 3.2). Interferon ( $F$ ) is an abstracted unitless quantity and thus  $e$  is a unitless multiplier.

Chapter 3. Preliminary Models of Influenza Infection in Lung Tissue

similar to Eqs. 3.1 and 3.3.

$$\begin{aligned}
 \dot{N}_c &= \sigma - \frac{r'^2 \cdot N_c}{R^2 \cdot t_{rc}} & \pi r'^2 &= a\sqrt{I} + b \\
 \dot{N}'_f &= \frac{(r'^2 - r^2) \cdot N_c}{R^2 \cdot t_{rc}} - \frac{v_{tcell}}{(r' - r)/2} & \pi r^2 &= \pi r_{cell}^2 \cdot I \\
 \dot{N}_f &= \frac{r^2 \cdot N_c}{R^2 \cdot t_{rc}} + \frac{v_{tcell}}{(r' - r)/2} \\
 \dot{T} &= \rho T - \beta TV & & (3.4) \\
 \dot{I} &= \beta TV - \delta I - k_e N_f I \\
 \dot{V} &= pI - \beta TV - \gamma(t)V \\
 \gamma(t) &= \begin{cases} 1/\text{day} & , \quad t < 5 \\ 3/\text{day} & , \quad t \geq 5 \end{cases}
 \end{aligned}$$

$N_c$  is the number of circulating T cells,  $N'_f$  is the number of circulating T cells that have found and exit into a region of lung tissue expressing chemokines,  $N_f$  is the number of circulating T cells that have found an infected region.  $T$  is the number of uninfected target cells,  $I$  is the number of productively infected cells, and  $V$  is the viral titer in serum. The infected region is assumed to be of radius  $r$  and is within a region expressing chemokines of radius  $r'$  ( $r < r'$ ). The lung is modeled as a circular region with radius  $R$ . The area of the infected region is equal to the area of an infected cell (of radius  $r_{cell}$ ) multiplied by the number of infected cells. See Fig.3.1 for more details about the model. The area of the region expressing chemokine was found to be related non-linearly to the number of infected cells ( $\pi r'^2 = a\sqrt{I} + b$ ) where  $a$  and  $b$  are constants that depend on the viral strain.  $a$  and  $b$  were fit to 5 experimental runs of a spatial model implemented in CyCells ( $R^2 = 0.997$ ,  $P < 0.01$ ).

Circulating CD8 T cells ( $N_c$ ) are assumed to be released from lymph nodes at a constant rate  $\sigma$  and circulate to the  $N'_f$  population over a time defined by  $t_{rc}$ . We assume these circulating cells transition into  $N'_f$  and  $N_f$  at rates proportional to the

Paramter	Value	Source
T(0)	1e6	[24]
V(0)	1e4	[24]
$r_{cell}$	5 $\mu$ m	[47]
$t_{rc}$	6s	[91]
$k_e$	6.4e - 5/cell/day	[47]
$\beta$	4.8e - 7/cell/PFU	[24]
$p$	0.18PFU/h	[24]
$\delta$	16.7hours	[24]
$a$	67.7	ABM fits
$b$	433.5	ABM fits

I	C
4	329
58	997
816	2550
3,264	4,410
10,050	8,069

Table 3.7: **Parameters for T cell production (Equation 3.4)**. Values are chosen to match the equations borrowed from both [24] and [47]. Values  $a$  and  $b$  were fit to multiple runs of a simplified CyCells ABM using the linear model  $C = a\sqrt{I} + b$  with a resulting  $R^2 = 0.997$ ,  $P < 0.01$ .

areas of the chemokine expressing and infected regions relative to the whole lung area.

The average time an exiting CD8 T cell takes to migrate to the infected region is equal to the difference in the radii between the two regions ( $r' - r$ ) times the T cell speed,  $v_{tcell}$ . Circulating cells that are in the chemokine expressing region ( $N'_f$ ) move into the infected region after performing chemotaxis for that time. Target cells ( $T$ ) become infected by virus at rate  $\beta TV$ , where  $\beta$  is the rate constant characterizing infection. Infected cells ( $I$ ) die at rate  $\delta$  in addition to being lysed by T cells ( $N_f$ ) at a rate  $k_e$ . Finally the viral titers ( $V$ ) increase due to production of virus at rate  $p$  by infected cells. Virus is also cleared due to uptake by infected cells (at a rate  $-\beta TV$ ) and due to antibody (at a rate  $\gamma(t)$  that changes after 5 days post infection). The initial viral titer was initialized to 10,000 PFUs and the initial number of target cells was one million. The initial number of infected cells is assumed to be zero. Parameter values are listed in Table 3.7. This ODE was fit to data taken from [47] using Matlab's `nlinfit` function in order to obtain a value for  $\sigma$ . The final value

was found to be 1,257 per hour.

## 3.6 Discussion

We used the initial DDE model (Eq. 3.1) to describe aggregate infection kinetics and to derive parameters for infected cell lifespan and the length of the *eclipse* phase, i.e. the time from infection of a single cell to the start of viral secretion in that cell. Differential equations are valuable for studying viral dynamics in a well-mixed system. However, spatial effects are not represented, yet they are likely important in the experimental system where influenza infection spreads through a monolayer of epithelial cells. Thus, we used a CA model that accounts for the spread of infection to nearby susceptible cells to better represent spatial influenza dynamics. We observed experimentally that pH1N1 influenza produced more virus after 24 hours in culture than the other strains, and the models calculated a higher  $R_0$  for pH1N1 than for the aH5N1 and sH1N1 strains.

CA modeling corroborated the results of the DDE model but in the pH1N1 infection,  $R_0$  values in the CA model were lower than those in the DDE model, more accurately reflecting the constraints on lateral spread of virus in the *in vitro* system. In contrast the many orders of magnitude variation in the absolute values of  $e$  (the effectiveness of the anti-viral activity in repressing viral production) between the two models is explained by the different way each model treated  $e$ , making comparison between the models not relevant. In the DDE model  $e$  refers to the inhibitory effect among all cells at once, whereas the CA model considers each cell at a given instant in time.

We use an extended DDE model fits to calculate previously unknown parameter values for strain-specific chemokine secretion rates (Eq. 3.3) and a generic T cell production rate (Eq. 3.4). Taken together, the resulting values for viral dynamics

Strain	IP-10 Production ( $pg/s \cdot cell$ )	RANTES Production ( $pg/s \cdot cell$ )	Viral Production ( $PFU/s \cdot cell$ )
Avian H5N1	2.0e-4	1.3e-5	5.4e-5
	8.4e-5 — 4.2e-4	7.9e-6 — 1.9e-5	4.4e-5 — 3.7e-4
Seasonal H1N1	1.8e-4	8.9e-7	3.8e-4
	1.2e-4 — 3.0e-4	4.8e-7 — 1.6e-6	2.8e-4 — 1.5e-3
Pandemic H1N1	8.7e-5	4.3e-6	5.1e-3
	1.7e-5 — 7.1e-4	5.0e-7 — 3.5e-5	2.8e-3 — 5.3e-3

Table 3.8: **Strain-specific parameters.** Small text values show 95% confidence intervals resulting from 1,000 bootstrapping runs for each parameter [36]. Bootstrapping for the chemokine values was performed using the original fit of Eq. 3.3 to the data in Fig. 3.2 to produce new data sets. Viral production values and confidence intervals are taken from [24].

(Tables 3.3 and 3.2), chemokine secretion rates (Table 3.8), and the T cell production rate (1,257 T cells per hour) inform the design and evaluation of a spatial model of the T cell search in the human lung in Chapter 4.

The results of this chapter demonstrate that *in vitro* infection of human epithelial cells can be used to derive numerical estimates for viral replication rates of three influenza strains. Evaluating the results of the DDE and CA model fits give new insight into the mechanisms that cause phenotypic differences between infection from different influenza strains. The DDE and agent-based models provide reproducible results for  $R_0$  and  $p$ . The higher  $R_0$  of the pH1N1 strain measured in this chapter may be one of several parameters predicting the higher epidemiological  $R_0$  observed in field studies in human populations [92]. Likewise, the low  $R_0$  calculated here for the avian aH5N1 strain may predict the rare human-to-human transmission of aH5N1 strains [93].

### 3.7 Summary

In this chapter I examined the effects of influenza infection in the lung resulting from three different strains of influenza. Because direct observation of *in vivo* infection is impossible, I made use of numerical and computational models calibrated with data from *in vitro* experiments. Both numerical and spatial models were able to replicate the infection profile as measured in the *in vitro* experiment. By examining the fit values of the model parameters, I showed that the three influenza strains are differentiated by their virion release rate ( $p$ ) and their response to antiviral defense ( $e$ ). I ranked the three strains in terms of their replication rate ( $R_0$ ): avian aH5N1 < seasonal sH1N1 < pandemic pH1N1. Finally, I fit two differential equation models to empirical data to obtain strain specific values for two chemokine secretion rates and a general value for the T cell production rate. In the next chapter, I use the strain-specific parameters described here to examine how the T cell response adapts to different infection profiles.

## Chapter 4

# Chemokine Directed T Cell Search in the Human Lung

This chapter focuses on the interactions between T cells, chemokines, and replicating influenza virus as T cells search for pathogens in the lung. Chemical signals called chemokines are known to enhance T cell recruitment and search in both acute infections and chronic inflammatory diseases [94–98]. Infected epithelial cells secrete chemokines, especially upon contact with CD8 T cells [99, 100]. The ability of T cells to follow the chemical signals of the chemokines and locate pathogens is an intriguing example of a distributed search strategy found in nature.

In this chapter I examine how the immune response uses chemical signaling to aid the T cell response. As measuring real-time T cell movement in *in vivo* lung tissue is a recent advance [101, 102], I use spatial modeling to study the localized processes of how T cells interact with the focus of infection (FOI) and the chemotactic environment. Here, I present a spatially explicit agent-based model (ABM) to describe T cell interactions with chemotactic signals and a dynamically growing plaque. Using the model, I investigate the pathogenic potential of the three influenza strains from

the previous chapter: Avian H5N1 (aH5N1), seasonal H1N1 (sH1N1), and pandemic H1N1 (pH1N1). These three strains were selected for their differences in *in vitro* replication rates [24] as well as differences in *in vivo* severity of human and animal pneumonia and mortality.

The model is parameterized with values taken from the literature when available. Because chemokine secretion rates are central to our model and appropriate data values are not available, I use the rate estimates from the previous chapter to parameterize the model.

The material described in this chapter was published in Levin (2016) [88].

## 4.1 Spatial Model of Chemokine Directed T Cell Search

This chapter describes the creation and evaluation of a spatial agent-based model of chemokine directed T cell search. The model focuses on activated CD8 T cells migrating through the vascular and lymph networks, including their movement over tissue after extravasation (Fig. 3.1). While the volume of the lung consists of small spherical air sacs called alveoli, the spread of influenza throughout the lung occurs through the interconnected bronchiole-alveoli complex, which can be viewed as a connected monolayer. Thus, we represent the lung as a two-dimensional sheet of healthy epithelial cells (Section 4.1.4). Because we assume that activated T cells descend at random into branches of the vascular network which have endpoints uniformly throughout the lung surface, T cells in the model are introduced uniformly at random across the modeled lung surface. If a chemokine signal is detected on the surface of the lung cell monolayer, the T cell remains in the lung tissue and follows the chemotactic gradient to the focus of infection (FOI). T cells that do not encounter chemokine recirculate to the lymph node where they reenter the bronchial



## Chapter 4. Chemokine Directed T Cell Search in the Human Lung

vascular network. Once inside the endothelium, T cells follow the chemokine gradient to localized areas of maximum chemokine concentration. When a T cell contacts an infected epithelial cell it induces apoptosis immediately (Section 4.1.1). Otherwise, infected cells continue to secrete virus for a fixed time and then die after a delay.

The model is initialized with a single infected cell. After the incubation period, the infected cell begins secreting virus and chemokine at rates determined by the virus strain and chemokine type as calculated in the previous chapter (Table 3.8). Virus from expressing cells diffuses locally, infecting neighboring cells. Chemokine diffuses from secreting cells, creating a spreading region of stimulation around the FOI. To represent the human immune system’s innate response to pathogens, an anti-viral molecule Immunoglobulin M (IgM) is introduced into the model four simulated days after initial infection, increasing the viral decay rate. After five simulated days, representing lymph node stimulation and T cell proliferation, activated T cells exit the lymph node at a constant rate and travel through the vasculature to the tissue as described above. These cell and molecular interactions and state transitions are depicted in Figure 3.1.

### 4.1.1 Model Implementation

The model is implemented using an updated version of the CyCells software [2] (Section 4.1.3, [github.com/drewlevin/cycells](https://github.com/drewlevin/cycells)), a modeling platform for two or three dimensional agent-based simulations of the immune response (Section 4.1.3). Due to computational constraints the model is scaled to the size of a mouse lung (approximately  $100\text{cm}^2$ ), a well-established practice in spatially explicit influenza models [103–105]. Parameters affected by this choice include the T cell production rate, T cell circulation time, and the total size of the lung. Parameters that are independent of the chosen influenza strain are shown in Table 4.1. Strain-specific values were obtained in the previous chapter and are shown in Table 3.8.

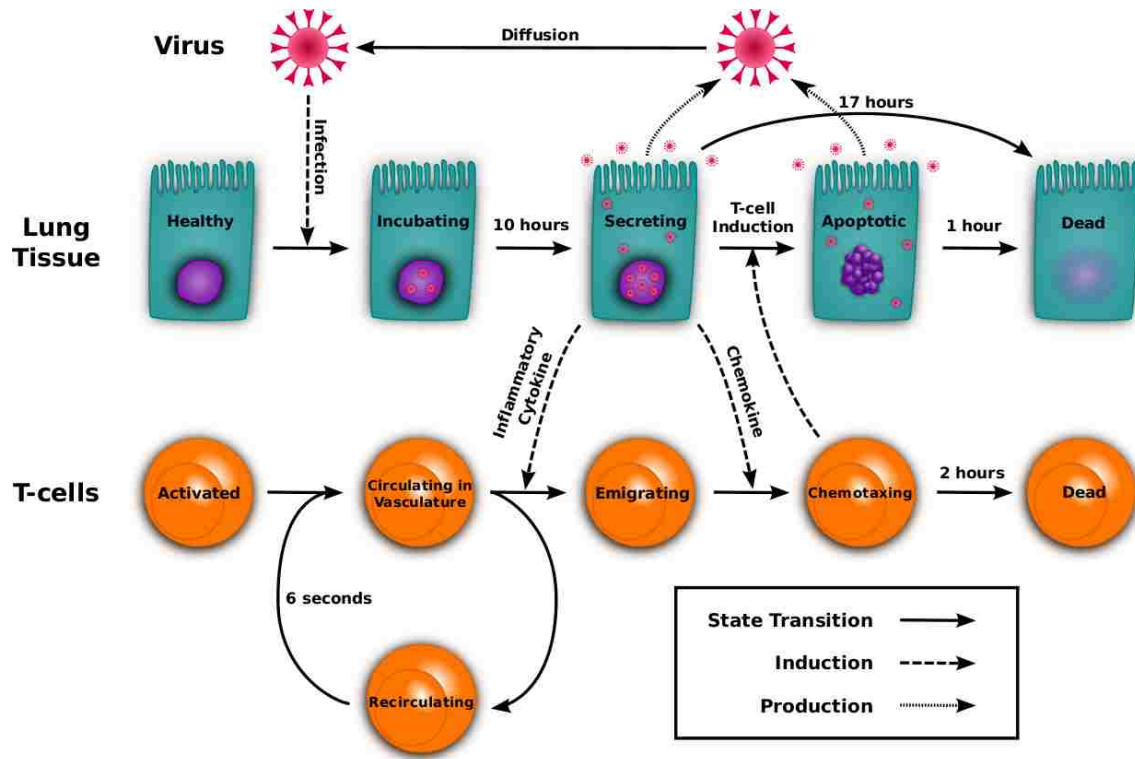


Figure 4.1: **Visual representation of the model.** CyCells model extended from Chapter 3 to include the T cell response. Healthy epithelial cells infected by virus begin secreting virus after the incubation delay. Activated T cells traverse the bronchial vascular network and may be recruited by inflammatory cytokine. Chemotaxing T cells climb the chemokine gradient and induce apoptosis in infected cells. Solid arrows represent a cell state transition from one behavior to another. Dashed arrows display the mechanism used to induce a transition. Dotted arrows indicate the production of new virus.

#### Chapter 4. Chemokine Directed T Cell Search in the Human Lung

In the model, epithelial cells are stationary and are described by one of five sequential states: **healthy**, **virus-incubating**, **virus-expressing**, **apoptotic**, and **dead** [24, 51, 106]. Cell states and transition criteria are illustrated in Figure 4.1.

**Healthy** cells transition to virus-incubating based on a probability that scales linearly with the local virus concentration. When a cell ‘becomes infected’ it removes the amount of viral concentration equal to a single virion.

**Virus-incubating** cells remain idle for 10 hours (with standard deviation  $\sigma = 1$  hour) and then transition to virus-secreting cells. Virus-incubating cells also begin to secrete both IP-10 and RANTES chemokines after 8 hours (except for aH5N1 which only secretes RANTES after a 16 hour delay).

**Virus-secreting** cells secrete both virus and chemokine (the RANTES portion of the chemokine production is added in 16 hours after the initial infection) and die after 1,000 minutes of production (16.7 hours). Virus-secreting cells also probabilistically transition to apoptizing cells in the presence of T cells (defined by the existence of a simulated T cell within  $2\mu m$ ). The number of T cells in the vicinity of the virus-secreting cell has no effect on the rate that apoptosis is induced in the model.

**Apoptizing** cells secrete virus and chemokine for one hour and then die.

T cells are described by two states: **circulating** and **chemotaxing**. T cells emerge from the lymph node at five days post infection (p.i.) at a rate of 1,257 cells per hour (Table 4.1). T cell travel time from the lymph node to a random location on the lung’s surface is six seconds. If chemokine is not encountered, the circulating T cell returns to a new location in the lung after another six seconds. If a circulating cell encounters chemokine, it changes state to chemotaxing and follows the chemotactic gradient to the FOI. Circulating T cells decay exponentially with an average lifespan of four days. Chemotaxing T cells follow the gradient through the two-dimensional

Referenced Parameters	Units	Min	Value	Max	Source
†Viral Diffusion in Airway	$\mu m^2/s$	3.18e-4	<b>3.18e-2</b>	3.18	[90]
†Viral Decay in Airway	$day^{-1}$	0.01	<b>1</b>	100	[50]
Chemokine Diffusion Rate	$\mu m^2/s$	3.18e-3	<b>.318</b>	318	[90]
Incubation Time	<i>hours</i>	5	<b>10</b>	20	[24]
Epithelial Cell Radius	$\mu m$	—	<b>5</b>	—	[107]
T Cell Radius	$\mu m$	—	<b>5</b>	—	[108]
T Cell Production Rate	<i>cells/h</i>	125	<b>1,257</b>	3,750	[47]
T Cell Speed	$\mu m/min$	6e-2	<b>6</b>	600	[109]
Blood Circulation Time	<i>seconds</i>	1	<b>6</b>	3,600	[110]
T Cell Sensitivity to Chemokine	<i>ng/mL</i>	—	<b>100</b>	—	[111]
Onset of T Cell Lymph Node Exit	<i>days</i>	—	<b>5</b>	—	[112]
†IgM Viral Decay Factor	—	1	<b>10</b>	1,000	[113]
IgM Onset	<i>days</i>	—	<b>4</b>	—	[113]
Estimated Parameters*	Units	Min	Value	Max	Footnote
Chemokine Decay Rate	<i>Hz</i>	3.8e-6	<b>3.8e-4</b>	3.8e-2	1
†Infectivity	<i>min/virion</i>	12	<b>120</b>	1,200	2
Expression Time	<i>min</i>	100	<b>1,000</b>	3,000	[24] <sup>3</sup>
T Cell Expected Kill Time	<i>min</i>	0	<b>10</b>	100	4
Apoptosis Time	<i>hours</i>	0	<b>1</b>	2	[114] <sup>5</sup>
T Cell Age (at FOI)	<i>min</i>	12	<b>120</b>	1,200	6
T Cell Age (in Blood)	<i>days</i>	0.04	<b>4</b>	400	6

Table 4.1: **CyCells model parameters.** Default values used for the model in bold. Min and max represent the extreme values tested in the sensitivity analysis (Table 4.2 and Figs. A.1-A.3). \* All estimated parameter values are examined in the sensitivity analysis (Table 4.2, Section A.1). † denotes parameters determined to be sensitive by the one-factor-at-a-time sensitivity analysis. Values were taken from experimental literature if possible and from earlier modeling papers if not. Parameters not found in literature were estimated as follows. 1) Corresponds to a 30 minute half-life. 2) Epithelial cells are infected at a probabilistic rate such that the expected time for infection in the presence of a single virion is 2 hours. This scales linearly with the number of virions in the cell’s vicinity. 3) Chosen as a plausible median time (1,000 minutes) between 6 hours and 24 hours. 4) T cells induce apoptosis in nearby virus-secreting epithelial cells at a probabilistic rate such that the expected time to induce apoptosis is 10 minutes. This rate does not scale with T cell numbers. 5) Calculated for low T cell densities. 6) Chosen to be at the lower end of biologically plausible values because increased T cell counts are shown not to affect the model behavior.

lung endothelium, inducing apoptosis when they encounter expressing epithelial cells. Chemotaxing T cells decay exponentially with an average lifespan of two hours.

In addition to T cells, the model contains two types of particles: virus and chemokine. Both are produced at constant rates by expressing epithelial cells. Virus diffuses through the lung tissue, infecting healthy cells at a rate proportional to the virus concentration at the location of the cell. Chemokine diffuses across the tissue but has no direct effect beyond recruiting T cells. Both particle types decay exponentially. IgM is modeled by increasing the viral decay rate by a factor of ten after the fourth day (Table 4.1).

### **4.1.2 Model Parameters**

We include only parameters that directly address the role of T cells and T cell migration because our study focuses on the role of T cells and the chemokine effects of T cells in influenza virus infection. Thus, we included parameters that affected virus, T cells, and chemokines. With regards to influenza, we chose parameters involving viral replication, infectivity in epithelial cells, and virus decay and diffusion rates based on our earlier study. The chemokine decay rate, diffusion rate, and secretion rate parameters are relevant to dynamic chemokine gradients required for T cell chemotaxis. Because we wanted to use our model to test the role of T cells and T cell migration in influenza clearance, we included relevant T cell parameters, such as T cell production rate, T cell death rate, T cell kill time, and T cell migration rate. We added IgM to represent innate antibody mediated virus clearance [113]. However, because our study does not address the role of antibody in clearance of influenza, we did not test the full range of B cell responses directly by including them in our parameters.

Model parameters are listed in Tables 3.8 and 4.1. These values were taken from the literature when available. Other parameters were determined by matching

Category	Parameter	Avian H5N1	Seasonal H1N1	Pandemic H1N1
Chemokine	Chemokine Decay Rate	bounded stable	bounded stable	stable
	Chemokine Diffusion Rate	bounded stable	stable	stable
	Chemokine Secretion Rate	bounded stable	stable	stable
T Cell	Circulation Time	bounded stable	bounded stable	bounded stable
	T Cell Kill Rate	stable	bounded stable	bounded stable
	T Cell Speed	stable	bounded stable	bounded stable
	T Cell Age in Blood	stable	bounded stable	bounded stable
	T Cell Age at FOI	stable	<b>bounded stable</b> <sup>†</sup>	bounded stable
	T Cell Production Rate	bounded stable	bounded stable	bounded stable
Delay	Apoptosis Time	stable	stable	stable
	Expression Time	peak change	<b>peak change</b> <sup>†</sup>	peak change
	Incubation Time	peak change	peak change	peak change
Virus	Viral Response to IgM	sensitive	<b>sensitive</b> <sup>†</sup>	sensitive
	Infectivity	sensitive	<b>sensitive</b> *	sensitive
	Viral Decay Rate	sensitive	<b>sensitive</b> *	sensitive
	Viral Diffusion Rate	sensitive	<b>sensitive</b> *	sensitive

Table 4.2: **Sensitivity Analysis Results:** The above parameters were varied over predetermined ranges in isolation, resulting in new model runs for every new value tested (Figs. A.1-A.3). The results of the sensitivity analysis were then qualitatively evaluated for each individual parameter. A model run’s behavior was determined by examining the height of the peak of the infection at day 5 post-infection and the number of infected cells at day 10 post-infection. Each combination of influenza strain and free parameter was classified as belonging to one of four categories. Parameters were classified as **stable** if all runs follow the same behavior, **bounded stable** if intermediate parameter adjustments did not affect the model’s behavior, even if the more extreme adjustments did, **peak change** if the peak of the infection differs but the result at day 10 is the same, and **sensitive** if any level of change in the parameter affects the resulting model behavior. PRCC analysis was also performed for the sH1N1 strain over these parameters (Figs. A.4-A.6). Bold text in the seasonal column denotes significant Spearman rank correlation ( $p < 0.01$ ) over the time period where the parameter was active. † indicates a maximum *absolute* Spearman’s  $\rho$  of less than 0.5. \* indicates a maximum *absolute* Spearman’s  $\rho$  of over 0.5.

empirical data to DDE and agent-based models (Chapter 3), and some parameters were estimated using biologically plausible ranges. All estimated parameters were studied in a sensitivity analysis to test their impact on model behavior (Table 4.2, Section 4.2.5 and Appendix A).

### **4.1.3 CyCells Implementation**

I use the CyCells modeling environment to simulate the T cell and chemokine search process [2]. CyCells splits its populations into two types: cells and molecules. Cells are considered to be unique agents and are modeled individually. Conversely, molecules are represented as compartmentalized concentrations. These compartments are arranged as a grid that covers the environment. Each square in the grid contains a unique concentration of the given molecule. Each molecule type is represented in an independent grid.

Molecular behavior in CyCells is limited to diffusion and decay. At each time step, CyCells decreases the concentration of each square in the grid according to the decay rate specified for the given molecule. CyCells then diffuses the molecules by applying a discrete diffusion equation to each grid square, taking into account the specified diffusion constant and the concentrations in neighboring squares.

New molecules can be secreted by cell objects, such as an infected cell secreting new virus. In this scenario CyCells calculates the amount of virus secreted per time step, based upon the cell's defined production rate, and then adds this quantity to the grid square that overlaps the secreting cell. If a cell ingests a molecule, it will subtract the appropriate concentration from the grid square it overlaps. If the concentration at that square is not enough to represent a full molecule, concentration will be removed from neighboring squares in an ever expanding diamond until the total concentration is equal to a single molecule.

Cells measure the local concentration at their position by calculating a linearly interpolated combination of the concentrations in the grid squares surrounding the cell in instances where cell behavior depends on the local molecular concentration.

#### 4.1.4 Modeling Decisions

As in the previous chapter, we design our model by simplifying components of the system being modeled. In designing the model for implementation in CyCells, we chose several explicit simplifications of the T cell response to infection. First, we represent the complex three dimensional structure of the lung as a two dimensional surface. Second, we assume blood flows through the branching bronchial vascular network at a uniform rate. Third, we chose a chemokine concentration threshold below which T cells cannot detect the present of the chemokine. Fourth, we assume chemokines combine additively such that local concentrations of IP-10 and RANTES can be summed together.

We model the lung as a two-dimensional system for the reasons previously described in Section 3.1.1.

Our model also assumes that blood flows uniformly through the lung vascular network. It is possible that blood flow is increased in the direction of an infected region of the lung by local inflammation. Thus our model may underestimate the recruitment efficiency of local inflammation and hyperemia. If this is indeed the case, our model underestimates the effect of the immune response.

T cell chemotaxis of the local chemokine gradient is only dependent on the relative chemokine concentration as compared to the surrounding environment. Absolute concentration values have no inherent impact on T cell movement. Thus, to place a limit on chemokine effectiveness, we choose a concentration threshold of 100 *ng/mL* below which T cells cannot detect chemokines. This value was chosen to best ap-



proximate *in vitro* distances of chemokine effectiveness. This assumption was tested using a preliminary version of the ABM (Section 4.2.1).

Our empirical results provide separate values and secretion rate estimates for chemokines IP-10 and RANTES. It is not known how multiple chemokines combine and interact. Therefore we assume the effective chemokine concentration is a simple summation of the two individual concentrations. The effects of combining chemokines is explored in Section 4.2.2.

## 4.2 Results

The evaluation of the spatial ABM provides many insights into the dynamics of the T cell search for infection. We first examine the effects of certain modeling assumptions in Sections 4.2.1 and 4.2.2. Section 4.2.3 examines the differences in the infections arising from the three different strains of influenza and the variance that arises from the stochasticity of the ABM. We use an ABM rather than a DDE model in this chapter to better represent spatial effects of the infection process. Section 4.2.4 describes the significance of these effects. Many of the chosen model parameter values are either estimated or obtained using model fits which result in wide confidence intervals. In Section 4.2.5 we examine the impact of each parameter value with two different sensitivity analyses. Finally, in Section 4.2.6, we seek to identify the maximum amount of time virus-secreting cells may produce virus while still allowing the infection to be controlled. This threshold determines the time window within T cells must complete their search process.

### 4.2.1 T cell sensitivity to chemokine

The model simulates a chemokine gradient surrounding the infected focus based on the calculated per-infected cell secretion rate (Table 3.8) and known chemical

parameters for a protein of weight 10 *kDa* (Table 4.1). T cell sensitivity depends on receptor density [115] and this was assumed to be constant.

Because this parameter is unknown, we simulated T cell sensitivity levels ranging over 10 orders of magnitude (Fig. 4.2). Concentration-dependent behavior of cells responding to chemokine was the same over a wide range of concentrations from 0.01  $\mu\text{g}/\text{mL}$  to 100  $\text{ng}/\text{mL}$  (model variance is discussed in the main text). Decreased recruitment was predicted only by supra-natural concentration sensitivity levels at and above 1  $\mu\text{g}/\text{mL}$ . We then set the sensitivity to 100  $\text{ng}/\text{mL}$  for all future runs (10 *nM* concentration assuming a chemokine molecular weight of 10 *kDa*) [116].

## 4.2.2 Chemokine combinations

Because avian aH5N1 influenza has been shown to suppress the production of interferon [24], we consider that IP-10 production is markedly reduced in comparison with that of RANTES and this may lead to elevated RANTES secretion measured in

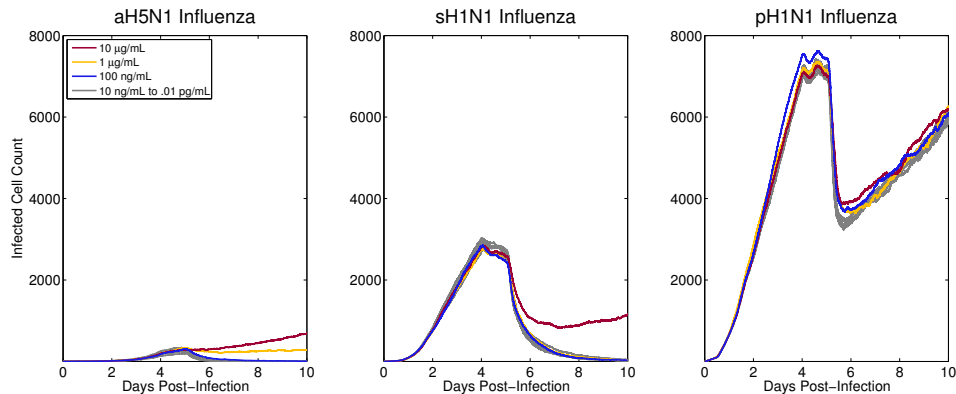


Figure 4.2: **Varying T cell sensitivity to chemokine.** H5N1 model results use RANTES only, and sH1N1 and pH1N1 use both IP-10 and RANTES. Total number of incubating, secreting and apoptotic cells are plotted for each infection. The sensitivity value specifies the minimum level of chemokine concentration required for T cells to detect it.

aH5N1 *in vitro* infection when compared to the other strains. Thus, IP-10 production was not included in aH5N1 simulations.

T cell sensitivity depends on receptor density [115] and this was assumed to be constant. Thus, the chemokines in combination work additively in our model.

Four models runs (both, IP-10, RANTES, and none) were performed for sH1N1 and pH1N1 strains and two runs (RANTES and none) for aH5N1. The runs show how the presence and/or absence of specific chemokines affect the simulated immune response (Fig. 4.3). The lack of both chemokines leads to runaway infections in all three strains. The presence of only RANTES is enough to contain the aH5N1 infection, but is weaker than IP-10 in both H1N1 strains. In sH1N1 and pH1N1 simulations, IP-10 alone proves to be as effective as the combination of IP-10 and RANTES. This suggests that RANTES does not play a significant role in infections that stimulate an IP-10 response due to the higher production rates of IP-10.

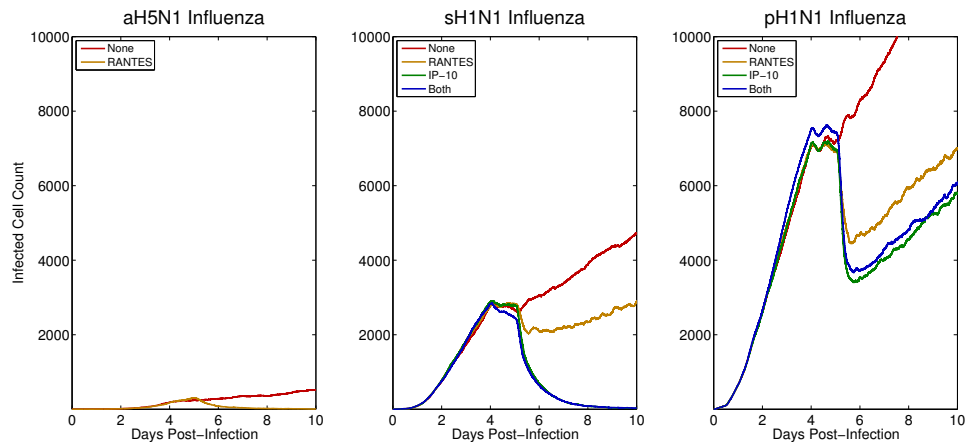


Figure 4.3: **Effects of different chemokine combinations.** A) aH5N1 does not stimulate an IP-10 response. B-C) sH1N1 and pH1N1 show no significant difference between IP-10 alone versus IP-10 and RANTES combined.

### 4.2.3 Primary Model Behavior and Variance

Because the CyCells model is stochastic, we conducted 100 runs of the model with the same parameter set for each of the three influenza strains, each run initialized with a unique random seed (Fig. 4.4). We count the number of infected cells in the simulated infection at every time point. For the avian and seasonal strains, all model runs cleared the infection by day 10. Conversely, each simulation of the 2009 pandemic influenza led to uncontrolled infection.

We estimated variance across runs by calculating the standard deviation at each time point. For aH5N1, the maximum standard deviation of the model output was 35, which occurred at day 4.8 where the mean number of infected cells was 257. For sH1N1, the maximum standard deviation was 212, which occurred at day 5.1 where the mean number of infected cells was 2,528. For pH1N1, the maximum standard deviation was 231, which occurred at day 5.1 where the mean number of infected cells was 6,988 (Fig. 4.4).

Infected cell counts over time are displayed in Figure 4.4. In all studied strains, virus population growth slows at four days post-infection (p.i.), due to IgM appearance. Subsequently, the simulated CD8 T cell response causes the number of infected cells to decline quickly after day five p.i. aH5N1 is cleared completely, sH1N1 is cleared more slowly. In contrast, at day six p.i., pH1N1 replication overwhelms the IgM and T cell response. These results are consistent with previously reported *in vitro* differences among the three different strains [24].

### 4.2.4 Spatial Effects

Infection growth in the CyCells ABM is illustrated in Figure 4.5. Because the virus particles are an order of magnitude larger than the chemokine molecules, they diffuse more slowly according to the Stokes-Einstein equation, which relates the diffusion

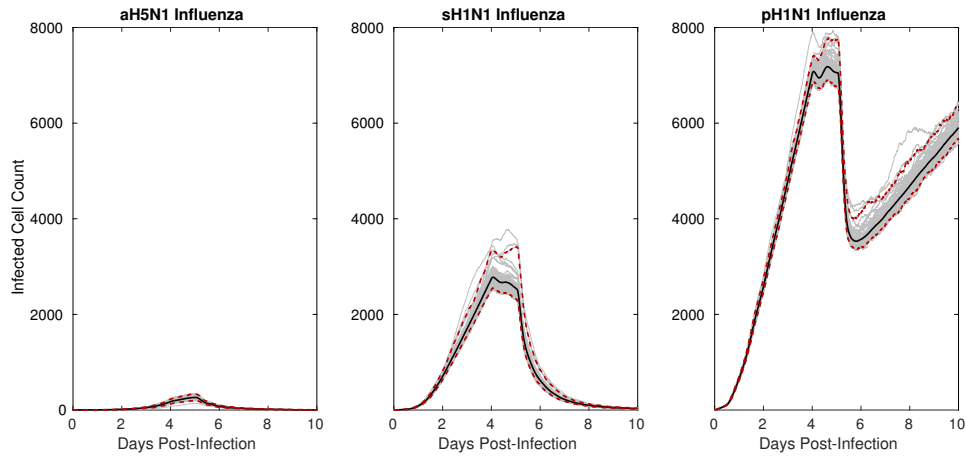


Figure 4.4: **Model results:** Time series plots of fifty runs of aH5N1 (A), sH1N1 (B), and pH1N1 (C) infections (gray). Each run took the calculated viral production and chemokine production rates for the three different strains of influenza as input (Table 3.8) and reported the total number of infected cells, including incubating, virus secreting and apoptotic, but not including dead cells. Therefore the figures approximate the rate of plaque growth over time. IP-10 and RANTES were simulated in each run, except for aH5N1, which produced only RANTES. Each run was initialized identically for each strain save for the random seed. The middle line shows the mean while the red dashed lines show the 96% credible confidence interval.

rate of a molecule to its size. However, chemokine molecules decay more quickly than the virus (Table 4.1). In the model these countervailing characteristics produce similar spatio-temporal profiles for the two particle types (Fig. 4.5). Before day 4 the plaque is populated primarily by virus-incubating and virus-secreting cells, with a small percentage of dead cells (top-left). Over time, cells in the plaque’s interior die, and active cells comprise a decreasing proportion of the plaque. T cells arrive at day 5 and begin killing the virus-secreting cells (top-center). By day 6 many expressing cells have been eliminated and the plaque is populated by dead cells (top-right).

In the aH5N1 infection, the plaque is densely populated with infected cells, allowing T cells to find secreting cells easily and clear the infection (Figs. 4.6A). However, secreting cells were not cleared in either the sH1N1 or pH1N1 simulations (Fig. 4.4).

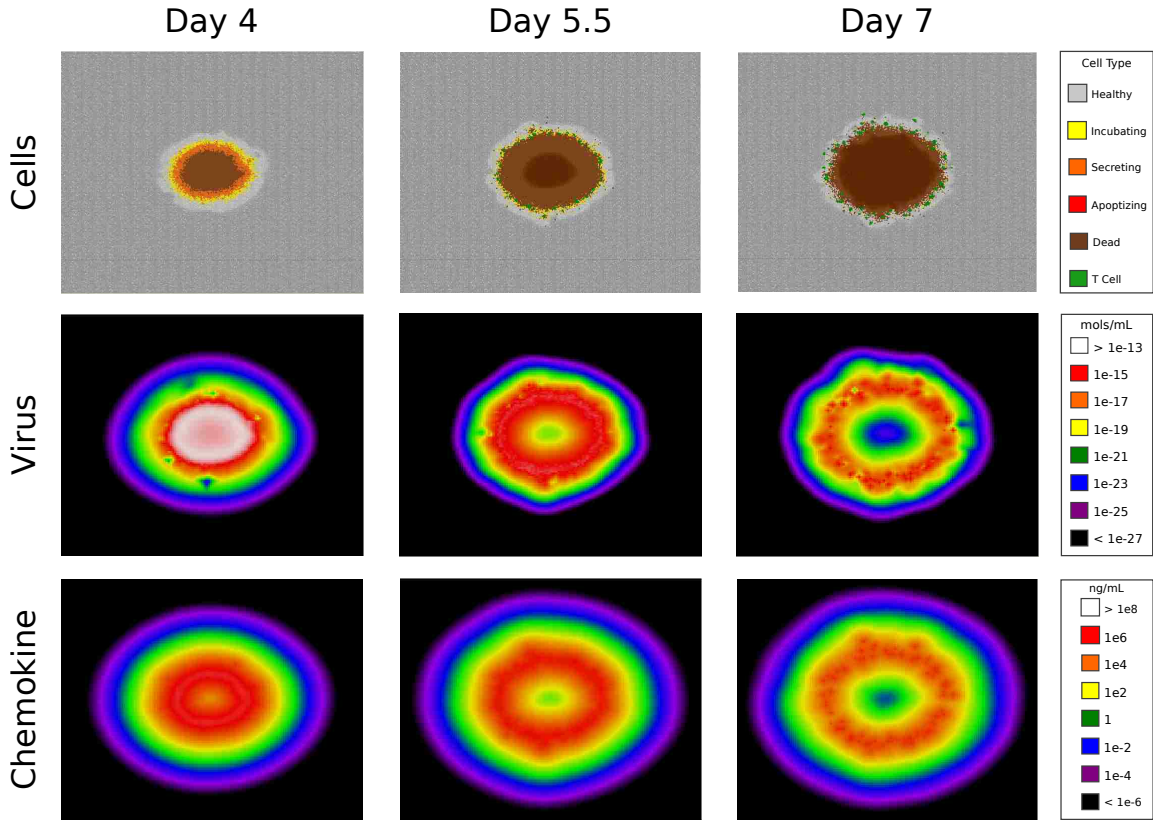


Figure 4.5: **Simulated sH1N1 infection.** Screenshots from day 4, day 5.5, and day 7. The top row shows the spreading focus of infection through the color coding of individual cells: healthy cells in uninfected tissue (gray), virus-incubating cells (yellow), virus-secreting cells (orange), apoptotic cells (red), dead cells (brown), and T cells arriving at day 5 (green). Free virus and chemokine particles are represented by compartmentalized concentrations of mols/mL and ng/mL. Chemokine shown is an aggregate of total IP-10 and RANTES concentrations.

In these runs, virus-secreting cells accounted for at most 10% of the active cell population and less than 1% of the total plaque at 6 days p.i. (Fig. 4.6). T cells continue to accumulate, but they arrive at a slower rate than the rate at which new cells become infected. Further, the newly arriving T cells are less efficient because the infected epithelial cells are more sparsely distributed in the plaque.

Finally, the spatial nature of the model reveals that the locations of maximal

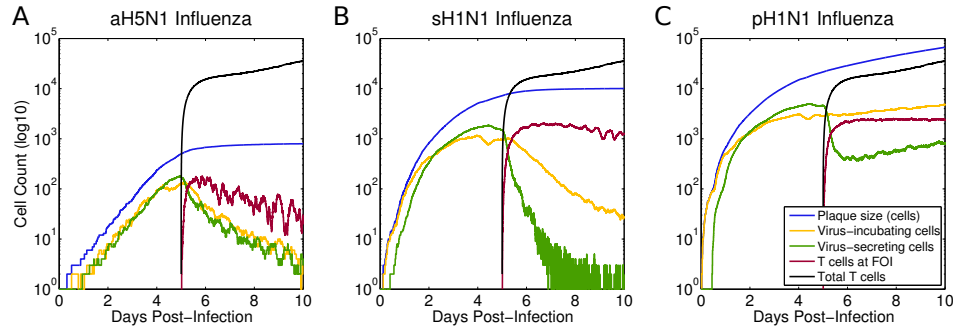


Figure 4.6: **Simulated infections of aH5N1, sH1N1, and pH1N1.** Plotted values: total plaque size (blue), number of virus incubating cells (yellow), number of virus secreting cells (green), total number of T cells (black), and T cells at the focus of infection (red). T cells clear secreting and incubating cells in aH5N1, fail to clear incubating cells in sH1N1, and fail to clear either type of infected cell in pH1N1. The number of incubating cells (yellow) after day 5 differs markedly among the three strains indicating that the T cells have differing success at controlling the infection.

chemokine concentration can move more slowly than the rate infected cells and virus expand, thus misdirecting the T cells. It takes time for infected cells to begin producing chemokine, while the preexisting areas of high chemokine density are slow to decay. Thus, T cells whose movements respond to the spatial layout of the chemokine gradient can become trapped, failing to locate new regions of infected cells in the growing plaque. The differences in the infection outcomes between the three strains illustrate the effect of the infection spreading faster than the chemokine maxima can move in the case of pH1N1. The velocity of the chemokine maxima positions depends on a combination of many factors, including the chemokine diffusion, decay, delay, and secretion rates (Table 4.1).

#### 4.2.5 Sensitivity Analysis

Because of the quantitative uncertainty of the values of many of our chosen parameters, we performed two sensitivity analyses of sixteen estimated parameters to identify

## Chapter 4. Chemokine Directed T Cell Search in the Human Lung

those that have an effect on determining model outcomes (Appendix A). We studied all but five of the parameters listed in Tables 3.8 and 4.1, excluding T cell radius, T cell sensitivity to chemokine, T cell onset time, IgM onset time, and the viral production rate. T cell radius is known [108] and does not factor strongly into the spatial model's implementation. T cell sensitivity was set to mimic a realistic threshold of response to chemokine and was analyzed separately in Section 4.2.1. The two onset parameters are known [113]. Viral production is an independent variable, as discussed above, and determined by each of the three strains [24].

Each of the remaining sixteen parameters was varied independently (over multiple orders of magnitude when appropriate) for each viral strain using a one-factor-at-a-time (OFAT) approach, and the model was rerun once for each parameter setting (Section A.1). Results were compared for the quantitative size of the peak infection and the number of uncleared infected cells at day 10 p.i. Based on these runs, we classified each parameter for each strain as follows (Table 4.2): *stable* parameters are those that do not affect the model behavior over any of the different values tested; *bounded stable* parameters are stable across a biologically plausible range of values but model behavior diverges at some threshold; *peak change* parameters affect the peak size of the infection, but do not change the ultimate number of infected cells at the end of the run; and *sensitive* parameters substantively change the outcome of the simulation within biologically plausible ranges.

Table 4.2 shows that four parameters are sensitive: viral response to IgM, infectivity, viral decay rate, and viral diffusion. In addition, Figure 4.4 shows that viral production rate is a sensitive parameter, varying among the three panels. The four parameters (infectivity, viral decay rate, viral diffusion rate, and viral production rate) are specific viral characteristics. Further, viral response to IgM, as coded in the model, directly affects viral decay rate and can be considered a viral parameter. None of the T cell parameters nor the chemokine parameters has a strong effect on



the model’s prediction of control or lack of control of viral replication. This result supports the hypothesis that virus dynamics dominate the course of an influenza infection.

Because OFAT analysis examines each parameter in isolation, we also performed a partial rank correlation coefficient (PRCC) analysis on the sH1N1 strain as described in Marino (2008) [38] (Section 2.3.3). PRCC was performed on the same 16 parameters plus the viral secretion rate to determine the strength and significance of the parameters’ partial correlation with the model output (Section A.2). The PRCC analysis used Latin hypercube sampling (LHS) to select parameter sets that provide coverage of the full parameter space (Section 2.3.3). The five viral parameters are confirmed by the PRCC analysis to be significantly correlated with model output (Table 4.2). Consistent results between the two approaches suggests the less computationally intensive OFAT analysis is also accurate for aH5N1 and pH1N1.

#### 4.2.6 Window of Control

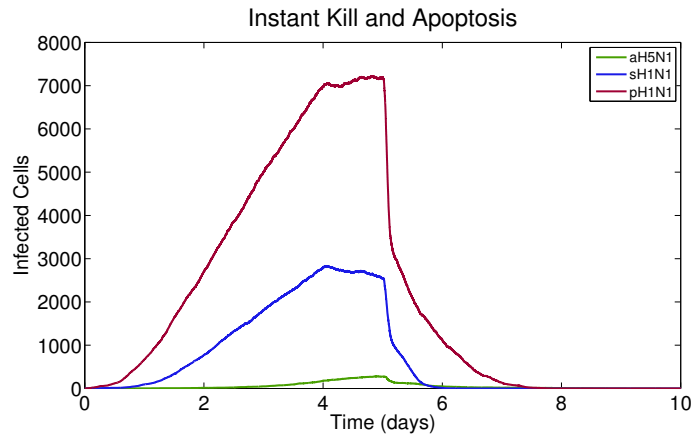


Figure 4.7: **Simulated infections of aH5N1, sH1N1, and pH1N1 with instant cell death.** The model results show that the combined delay of the T cell kill time and apoptosis time form a barrier to infection clearance. Removing both delays results in infection clearance for all strains.

## Chapter 4. Chemokine Directed T Cell Search in the Human Lung

The sensitivity analysis suggests that viral parameters are the most sensitive and have the largest effect on the course of the infection (Table 4.2). Here we examine the role of T cells by assessing multiple factors that control the overall T cell response. We seek to identify the maximum amount of time that virus-secreting cells can produce virus before the infection becomes uncontrollable and can't be cleared. We name this threshold the 'window of control'. This window can be defined as the combination of the time of T cell arrival at a virus-secreting cell ( $T_{arr}$ ), the time it takes for the T cell to induce apoptosis ( $T_{kill}$ ) and the time for the infected cell to apoptose ( $T_{apop}$ ). We examine how much virus a virus-secreting cell can produce inside this window. Values above one suggest a growing infection, while values below one suggest infection control.

An epithelial cell infected with pH1N1 produces new virus at the rate of  $5.1e-3$  PFU/s (Table 3.8). Assuming instantaneous T cell arrival at a new virus-secreting cell, T cell-induced apoptosis occurs within 10 minutes ( $T_{kill}$ ) and viral secretion continues for 60 minutes after that ( $T_{apop}$ ) according to our model (Table 4.1). Under these circumstances the cell produces 21 new viral particles that can infect new cells in this 70 minute window that occurs after T cell arrival. In contrast, with the sH1N1 strain, a single infected cell produces 1.6 viral particles, and for the aH5N1 strain this number is 0.2.

Even if T cell arrival and apoptosis time were instantaneous, a cell would still secrete virus for 10 minutes ( $T_{kill}$ ), producing 3 new virions in the case of pH1N1. Only in the case where both  $T_{kill}$  and  $T_{apop}$  are simultaneously reduced can the pandemic infection be limited to less than one new virion per infected cell. We confirm this effect in the model by first setting both relevant parameters (apoptosis time and T cell kill time) to zero, and as expected all three strains were cleared (Fig. 4.7). We now ask how much T cell arrival delay the model can tolerate and still clear a given viral strain. To answer this, we formulate the basic equations of

#### Chapter 4. Chemokine Directed T Cell Search in the Human Lung

the model in terms of  $R_0$ , the viral replication rate, and solve for the case where  $R_0 < 1$ .

$$R_0 = p \times (T_{arr} + T_{kill} + T_{apop}) \times E \quad (4.1)$$

where  $p$  is the secretion rate of the virus and  $E$  is viral effectiveness, i.e. the proportion of virus that infects cells. Equation 4.1 assumes each free virus particle can infect no more than one healthy epithelial cell.

$p$ ,  $T_{kill}$ , and  $T_{apop}$  are model parameters (Table 4.1).  $E$  incorporates both spatial effects of viral diffusion and temporal effects of viral decay. Calculating the true value for  $E$  is more difficult as it depends on many factors, including the area of the plaque, the depth of the ring of infected (but not dead cells) around the plaque, the local viral density, and the arrival rate and effectiveness of T cells. Each of these factors changes during the infection and thus we examine  $E$  as a function of time:  $E(t)$ .

To find values of  $E(t)$  for each of the three strains we reran the model in the absence of T cells and calculated the amount of virus that infects cells,  $I(t)$ , and the amount that decays  $D(t)$ , at each time step. The viral effectiveness rate is then calculated to be

$$E(t) = \frac{I(t)}{I(t) + D(t)} \quad (4.2)$$

This approach accounts for the factors listed above and provides a unique measure of  $E(t)$  at each point in time. We focus on results in the model after day 5 p.i. corresponding to the arrival of T cells in the main model. In each simulation  $I(t)$  remains relatively constant, while  $D(t)$  increases linearly with time. We perform

Chapter 4. Chemokine Directed T Cell Search in the Human Lung

least-squared regressions on  $I(t)$  and  $D(t)$  for each strain and combine the results to give an estimate for  $E(t)$  (Table 4.3).

We use the estimated values of  $E(t)$  to calculate  $R_0$  values for each strain in the absence of T cell mediated apoptosis. Because all three  $E(t)$  values change by less than 10% over the five day window, we estimate  $R_0$  values using the average  $E(t)$  value for each strain (Table 4.3). The three strains show similar values for  $R_0$ , suggesting a saturation of viral effectiveness as the plaque grows. This is not the case early in the infection as pH1N1 growth clearly outpaces sH1N1 which in turn expands more rapidly than aH5N1. In the main model, the T cell response limits viral production time, which in turn increases the effectiveness of the virus that is produced. Thus, it is reasonable to assume viral effectiveness remains high in the presence of the T cell response.

Strain		Slope ( $PFU/s$ )	Intercept ( $PFU$ )
Avian H5N1	I(t)	5.86e-8	-0.0133
	D(t)	1.03e-7	-0.0223
Seasonal H1N1	I(t)	3.25e-7	0.0225
	D(t)	6.93e-6	0.0439
Pandemic H1N1	I(t)	8.79e-7	0.093
	D(t)	2.71e-4	5.28
Strain		Avg. $E(t)$	Estimated $R_0$
Avian H5N1		0.355	1.16
Seasonal H1N1		0.049	1.11
Pandemic H1N1		0.0037	1.12

Table 4.3: **Linear fits to model runs for  $I(t)$  and  $D(t)$  across all three strains.** Every parameter was significant at the  $p < 0.05$  level.  $E(t)$  is estimated to be  $I(t)/[I(t) + D(t)]$ .  $E(t)$  does not vary by more than 10% from day 5 to day 10, thus the average of each is calculated and used to estimate  $R_0$  for all three strains in the absence of a T cell response.

As the infected plaque expands, some secreted virus diffuses to cells that are already infected or dead (the existing plaque) and do not contribute to new infections.

Chapter 4. Chemokine Directed T Cell Search in the Human Lung

This effect is minimal in the beginning when the edge of the plaque is small and tightly curved. As the plaque expands, its radius grows and the edge approaches the limiting case of a straight line when 50% of the virus could be expected to diffuse over healthy cells and 50% over infected or dead cells. In the presence of healthy cells, virions infect the cells at a rate of  $0.5/h$  and decay at a rate of  $0.42/h$  in the presence of IgM after day 4 p.i. (Table 4.1). This suggests that 50% of free virus contacts healthy cells and 54% of that virus succeeds in infecting a healthy cell, implying that  $E \approx 0.27$ , a value similar to that calculated for aH5N1 (Table 4.3). By accounting for basic spatial and temporal effects, we approximate  $E$  as 0.27 over the course of the infection.

A controlled infection will satisfy  $R_0 < 1$  (Section 3.1). Solving for  $T_{arr}$  where  $R_0 < 1$  gives:

$$\lim_{t \rightarrow \infty} T_{arr} < \frac{1}{pE} - (T_{kill} + T_{apop}) \quad (4.3)$$

Using the model parameters, this equation suggests that  $T_{arr} < 18h$  for aH5N1 and  $T_{arr} < 93m$  for sH1N1. pH1N1, however, cannot be cleared for any value of  $T_{arr}$  in this scenario.

The above calculation suggests a combined transition point for sH1N1 at 163 minutes ( $R_0 = 1$  for  $T_{arr} + T_{kill} + T_{apop} = 163m$ ), which is consistent with the sensitivity analysis (Fig. A.2: Apoptosis Time: 90m - 120m) and suggests that the modeled  $T_{arr}$  is between 33 and 63 minutes for sH1N1. These data show that the balance between viral production and T cell response is a key factor in clearance of infection: pH1N1 can theoretically be cleared with instantaneous T cell effectiveness (Fig. 4.7), but in actuality viral production exceeds T cell response time and this leads to uncontrolled infection.

### 4.3 Discussion

We used a spatial model to study how T cell search in the lung affects the host's ability to clear virus. CD8 T cells conduct two 'searches' in two different tissue environments, first to encounter antigen-loaded dendritic cells in the lymph node, and second to encounter local inflammatory signals in the infected lung. Search problems in the lymph node have been simulated using live cell imaging data to provide reliable parameters of cell movement [54, 117]. Our model focused on the second search, i.e. the process of recruiting activated CD8 effector cells to infected sites in the lung, which is not as well understood. In contrast to the dense 3D lymph node volume, our model approximates a mouse lung as a 2D surface where the epithelial cell monolayer is  $100 \text{ cm}^2$  (only the area containing the FOI is modeled explicitly). T cells must somehow rapidly locate the infection, which at day 5 p.i. is approximately  $0.5 \text{ mm}$  in diameter. Our model revealed four main insights regarding chemokine aided T cell search in the lung.

First, confirming earlier results [24] and observed phenotypic differences, the model consistently clears the aH5N1 strain by day 10 p.i., contains sH1N1 by day 10 p.i., and fails to clear the pH1N1 infection. Second, the model revealed spatial constraints on T cell search when the infection spreads more quickly than the chemokine gradient can diffuse. In these cases, T cells become trapped in areas of high chemokine concentration which lag behind the expanding infection. High concentrations of chemokine also attract and trap arriving T cells, thus limiting the direct benefit of increasing numbers of T cells. Third, the sensitivity analysis tested the individual effects of each parameter on the model's behavior. The analysis theorizes that infection outcomes are highly sensitive to the properties of the specific virus. Finally, we examined the window of control which suggests that pandemic influenza could not be cleared even if T cell arrival was instantaneous.

#### Chapter 4. Chemokine Directed T Cell Search in the Human Lung

There is clear evidence from multiple models that chemokines and chemokine receptors are required for effector T cell localization to infected tissues [118–133]. However, the actual effect of chemokines on effector T cells in tissues is still largely unknown as *in vivo* quantitative chemokine parameters in the infected lung are difficult to estimate. Dynamic chemokine concentrations secreted by bronchial epithelial cells *in vitro* depend on infection intensity and cell maturation state [24, 100, 134, 135] and are reasonable approximations of chemokine levels in tissue, probably better than blood levels documented in virulent influenza infections [136]. Our experimental collaborators have determined the level of chemokine released by infected bronchial epithelial cells (Table 3.8), and we used these as the best available approximation for chemokine levels in tissues. Our model ignores the potential contributions from other chemokines that did not show different levels in the physical experiment. This includes chemokines such as CXCL8/IL-8 detected in bronchial cell cultures [137, 138] as well as chemokines secreted by immigrant macrophages [139], neutrophils [140], and amplification of epithelial cell secretion by CD8 T cells [99]. Because the sensitivity analysis classified the chemokine secretion rate as stable, ignoring these other factors is reasonable.

Although some immune cells exhibit directional behavior to chemokines [141, 142], T cells have not yet been shown to follow chemokine gradients. While T cell chemotaxis may not be proven, the previous chapter suggests severe limitations to viral clearance in the absence of this ability [112]. Our sensitivity analysis shows that slowing T cell speed (and consequentially, the ability of T cells to climb gradients) reduces clearance of sH1N1 and pH1N1 strains. These results are consistent with the hypothesis that T cells are guided by chemokine in the lung epithelium.

## 4.4 Summary

In this study, I tested how chemokine directed T cell search contributes to infection clearance. Explicitly modeling T cell chemotaxis revealed unique challenges to viral clearance brought about by infections that spread more rapidly than the chemokine gradient could decay and diffuse. Furthermore, the model was also able to account for spatial constraints of viral diffusion and infection. Figure 4.5 illustrates how new virus diffuses over cells that are no longer able to be infected. Differential equation models can account for this discrepancy by adjusting parameters relating to virus-cell infection by a constant amount, but this assumes that the proportion of unsuitable virions will always be the same. My spatial ABM visualizes how the early infection has a higher proportion of virus overlapping healthy cells (Figure 4.5).

In the spatial model, T cells climb a chemokine gradient to find infected epithelial cells and cluster at local maxima of chemokine concentration. Because T cells are clustered, they cannot cover the expanding plaque effectively, where infected cells on the periphery become more highly dispersed as the plaque grows. Thus, T cells in the model become redundant at a relatively low threshold, beyond which additional T cells do not improve clearance rates. These observations provide an explanation for pH1N1 dynamics that would be obscured without the visualizations provided by spatial modeling.

The window of control describes the limit on the amount of time an infected cell may secrete virus while still allowing the infection to be cleared. Studying the time of T cell arrival at virus-secreting cells in the model revealed one reason why T cells fail to clear pH1N1 infections. I formalized this effect in an equation that calculates an upper bound on the search time for T cells to find new virus-secreting cells, and I showed that this value is consistent with the results of our sensitivity analysis.

The model captures important quantitative and spatial aspects of T cell response



*Chapter 4. Chemokine Directed T Cell Search in the Human Lung*

to influenza infection which have not been addressed by earlier T cell models. Further, the inability of T cells to efficiently handle with the rapidly spreading dense infection of the pH1N1 strain suggests unique challenges to distributed search processes in spatially disperse and volatile environments. In the next chapter I examine how ant foraging strategies are affected by similar environmental conditions.

## Chapter 5

# Environmental Determination of Ant Foraging Strategies

In this chapter I explore the hypothesis that the best ant foraging strategy is determined in large part by the environment in which the ants live. This hypothesis is supported by the observation that animal species use different strategies in different environments [143]. Chapter 4 demonstrated that agents using chemical recruitment in environments of rapid change can be led astray if the chemical signal moves slower than the search target. Based on these findings, I examine the effects on ant foraging efficiency caused by how food resources are distributed in space and their temporal variability, which I call *volatility*.

Here, I evaluate distributed foraging strategies based on the environment in which their search is conducted, using an agent-based model of ant foraging. I model three different ant-based foraging strategies, which subsume most strategies observed in nature: solitary random walk, nest recruitment, and pheromone recruitment. I quantify the efficacy of each strategy across two environmental dimensions, the spatial layout and the volatility of food resources, and I evaluate the ability of each strategy

to adapt to different environments.

I find that both pheromone and nest recruitment perform best in clustered stable (non-volatile) environments, and their efficiency declines as food dispersion and volatility increase. To be effective, each recruitment strategy requires that the ants complete at least two round trips to a location before the food disappears (volatility) it takes one round trip to obtain information and at least one more to exploit it. Because random search has no memory, volatility does not strongly affect its performance, and it performs well in environments where food is highly dispersed. The model's predicted strategy is consistent with over 90% of ant species whose foraging strategy can be classified. Because implementation of synthetic physical pheromone trails is difficult, nest recruitment is a promising alternative search strategy for robot swarms.

The material described in this chapter is published in Levin (2015) [144].

## 5.1 Spatial Model of Ant Foraging

We developed an agent-based model to study the effectiveness of various foraging strategies in different environments (Fig. 5.1), focusing on the spatial and temporal distribution of food resources on a flat surface. Our model extends the central-place foraging algorithm (CPFA) and swarm robotics platform studied in Hecker (2015) [1], by adding volatile resources and new foraging strategies. The model consists of a two-dimensional grid with discrete food units (analogous to seeds) placed on the grid before the run begins. There is a single nest, where ants congregate, leave to search for food, return with food if successful, and possibly recruit other ants to follow them to a food source. Food sources can be arranged in different distributions (e.g., grouped together in a small number of clumps or dispersed randomly across the environment). Volatility is modeled as movement: food moves at a rate param-

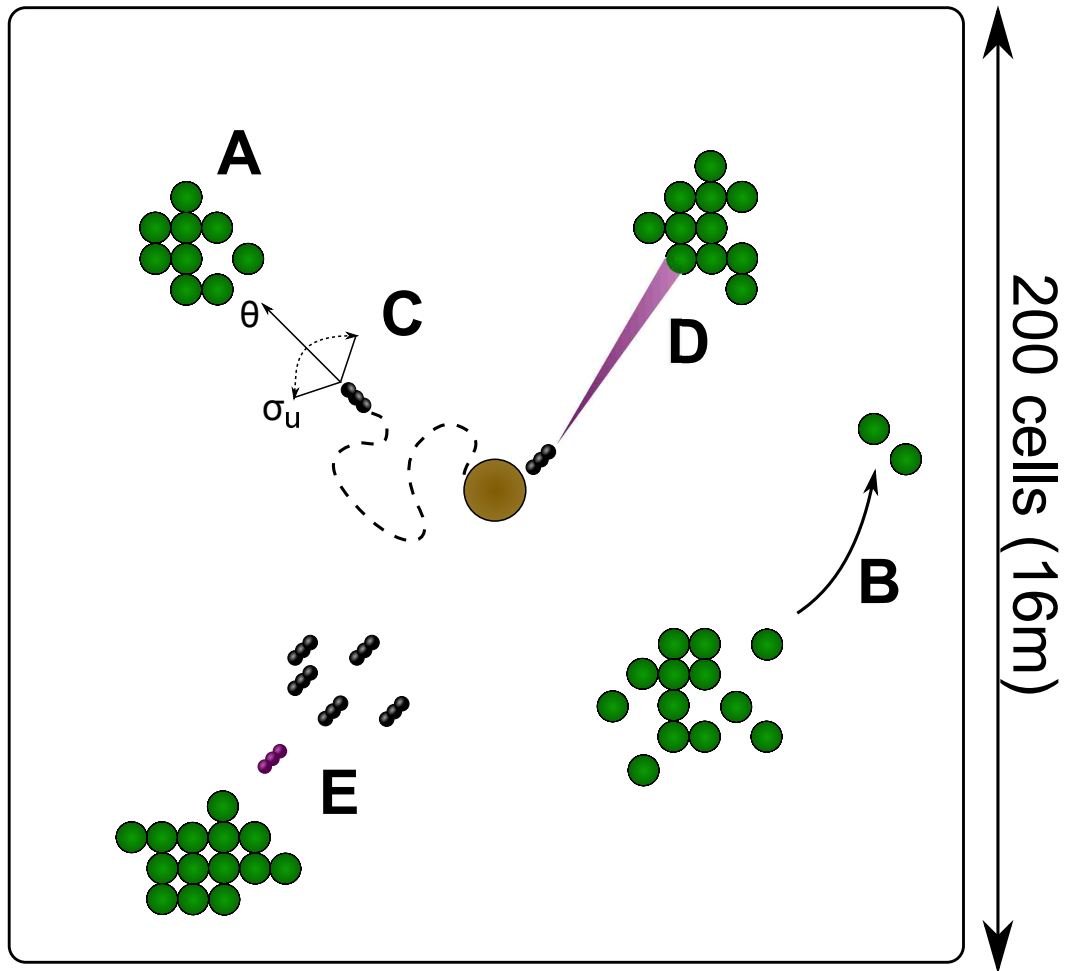


Figure 5.1: **Ant Foraging Model Conceptualization.** The model is initialized on a square grid with the nest in the center. **A)** Food is randomly placed in patches a minimum distance from the nest. **B)** Food from a single patch moves at a constant rate to a new patch (volatility). **C)** Ants perform a random walk from the nest. An ant's angular trajectory  $\theta$  is varied at each time step by choosing from a normal distribution:  $N(\theta, \sigma_u)$ . **D)** An ant lays a pheromone trail (decays exponentially) to a specific location if it detects enough food in its vicinity. **E)** Ants returning from a successful foraging search recruit ants resting at the nest.

eterized by the model. In the experiments for this paper, we varied the number of clustered food piles and the rate at which piles move, while holding the total amount

of food constant (Fig. 5.1 A and B).

### 5.1.1 Foraging Strategies

A recent paper by Lanan (2014) [23] catalogs certain ant foraging strategies used by monodomous (single nest) ant species: random search, site fidelity, tandem running, pheromone recruitment, nest recruitment, and trunk trails. Two of these strategies, tandem running and trunk trails, closely resemble other strategies: nest recruitment is similar to tandem running and pheromone recruitment is similar to trunk trails. Therefore, we model three distinct foraging strategies: *random search*, *pheromone recruitment*, and *nest recruitment* (Fig 5.1 C-E), where each strategy is designed to mimic foraging strategies used by real ants [56, 58, 65, 68, 69]. We enable site fidelity in both recruitment strategies based on field observations.

- **Random Search:** Individual ants leave the nest all at once and perform a correlated random walk through the two-dimensional space. Ants continue searching until they encounter food. Ants that encounter food pick it up, return to the nest, and begin a new search. The correlated random walk chooses a new trajectory on each time step for each ant from a normal distribution  $N(\theta, \sigma_u)$ , centered on the ant's current trajectory  $\theta$ . Ants using the random search strategy have no memory and perform no recruitment.
- **Pheromone Recruitment** is implemented following the central-place foraging algorithm (CPFA) detailed in Hecker (2015) [1]. Ants leave the nest and search randomly as described above. Ants may give up searching at any time with a probability  $\rho_q$  and return to the nest. However, if an ant finds food at any point it picks it up and immediately checks the neighboring 8 grid cells for more. Next, it decides to reinforce the location with probability  $P(k; \lambda_p)$ , where  $P$  represents the cumulative Poisson distribution and  $k$  is the amount of

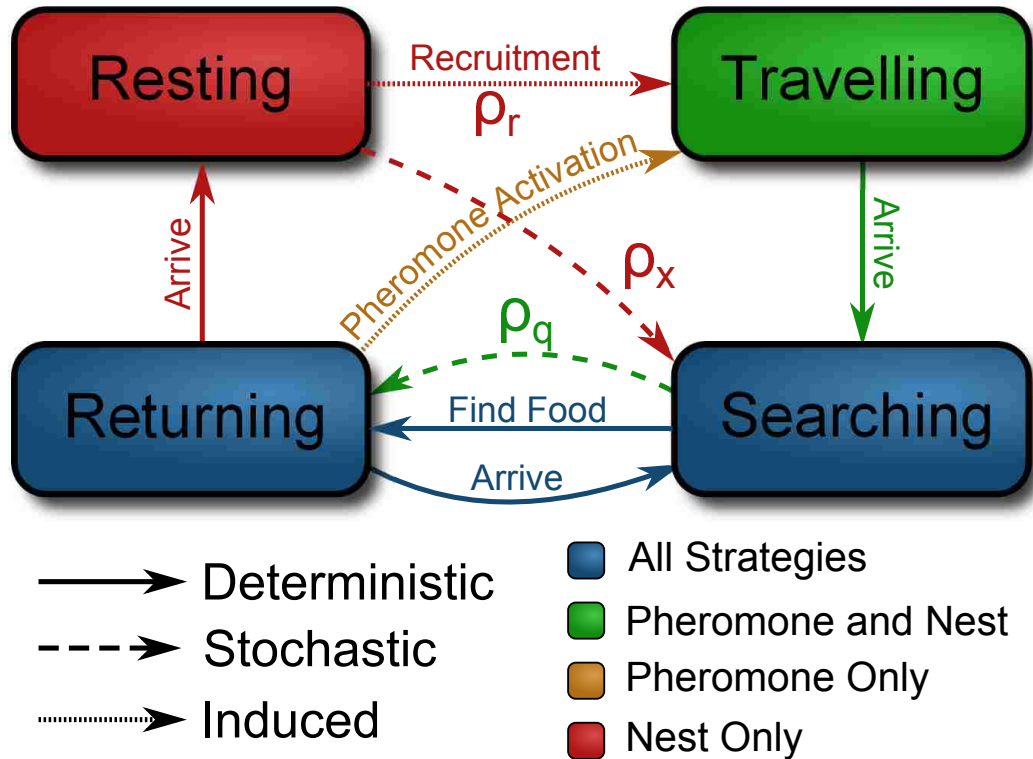


Figure 5.2: **Ant Foraging Model States.** Ants are initialized at the nest in either the **resting** (with nest recruitment) or **searching** state. Ants transition between behaviors based upon cues from the environment, random chance, and stimulation by other ants or pheromone.

food found in the immediate neighborhood. The ant may also use site fidelity to return to the previously visited location with probability  $P(k; \lambda_f)$ . It then returns to the nest, creating the trail upon its arrival if  $P(k; \lambda_p) > U(0, 1)$ ; the trail decays at a rate of  $\sigma_p$ . Subsequent ants may follow this trail to the same location before the trail evaporates. Recruited ants perform an informed correlated random walk upon arrival and may also lay a pheromone trail back to the nest.

- **Nest Recruitment:** Ants probabilistically leave the nest and use random

Abbr.	Name	Distribution
$\sigma_u$	Uninformed Search Correlation	$U(0, 2\pi)$
$\delta_i$	Informed Search Decay Rate	$E(5)$
$\rho_q$	Search Quit Probability	$E(30)$
$\lambda_f$	Site Fidelity Rate	$U(0, 20)$
$\lambda_p$	Reinforcement Rate	$U(0, 20)$
$\delta_p$	Pheromone Decay Rate	$E(10)$
$\rho_r$	Recruit Probability	$E(10)$
$\rho_x$	Leave Nest Probability	$E(100)$

Table 5.1: **Ant parameters tuned by the GA.** Parameters were initialized randomly using either a uniform distribution ( $U$ ) or an exponential distribution ( $E$ ).  $\sigma_u$  and  $\delta_i$  define turning parameters.  $\rho_q$ ,  $\rho_x$ , and  $\rho_r$  are probability rates.  $\lambda_f$  and  $\lambda_p$  are Poisson probability parameters.  $\delta_p$  is the pheromone decay rate. Parameters extend the model described in Hecker (2015) [1]. As in Figure 5.2: blue,  $\sigma_u$ , is used in by all three strategies, green parameters are used by the two recruitment strategies, yellow,  $\delta_p$ , is used by pheromone only, and red parameters are used by nest recruitment only.

search to look for food. Ants that are not actively searching remain in the nest. As in pheromone recruitment, ants give up and return to the nest with probability  $\rho_q$ , and ants that find food pick up the food, survey the area, and if they find food above the threshold, they return directly to the nest and recruit more ants (otherwise they return to the nest and being a new solitary search). The number of ants recruited on a single return to the nest is a fraction of the ants currently in the nest, probabilistic determined by  $\rho_r$ . The original ant and the newly recruited ants then return directly to the previous location and perform an informed correlated random walk.

An informed random walk behaves as the uninformed random walk, but with a turning parameter of  $4\pi$  radians that decays at the informed search decay rate,  $\delta_i$ , until it reaches the uninformed turning rate,  $\sigma_u$ .

These three strategies are well known in the ant literature, but the details vary among individual species, and in some cases the exact parameters are simply un-

known or difficult to measure accurately, such as the nest recruitment rate and the pheromone decay rate. Therefore we use a genetic algorithm (GA) to select each parameter for each environment. The GA-evolved parameters can significantly alter the outcome of a specific strategy. For example, in the pheromone recruitment strategy, if the decay rate of the pheromone is very high, it will dissipate before it can be used by other ants. This effectively reduces the pheromone recruitment strategy to random search.

Ants begin each simulation in the nest at the center of the grid. Ant behavior is governed by the eight parameters listed in Table 5.1. Behavior transitions among four possible states: **resting**, **traveling**, **searching**, and **returning** (Fig. 5.2) at rates determined by the evolved parameters.

- **Resting:** If using the nest recruitment strategy, ants begin by resting in the nest and transition to **traveling** if they are recruited by a returning ant. At each time step each ant may exit the nest and switch to the **searching** state at a probability  $\rho_x$ . Upon a successful return by another ant, ants in the nest may be recruited with a probability  $\rho_r$  and switch to the traveling state.
- **Traveling:** An ant in this state travels to a specified location deterministically. After arriving at the destination, the ant switches to the **searching** state.
- **Searching:** An ant in the searching state performs a correlated random walk, as described above.
- **Returning:** A returning ant travels directly back to the nest. After arriving, it returns to its previous location (*site fidelity*) with probability  $Poisson(k; \lambda_f)$ . If pheromones are enabled and the ant decides to reinforce the path, it lays a trail of strength 1.0, which decays at a rate of  $\delta_p$  each time step. Pheromones disappear completely (are undetectable) below a strength threshold of 0.001. If nest recruitment is enabled, and the ant has decided to reinforce the path,



it recruits with probability of  $\rho_r$  for each ant in the nest. The ant transitions to **traveling** if returning to a location using site fidelity or if it is following a pheromone trail. It transitions to the **resting** state if recruitment is enabled. Otherwise it transitions to the **searching** state.

### 5.1.2 Foraging Environments

To study the effects of different environments on optimal foraging strategies, we model two environmental dimensions: the spatial distribution of food and the volatility of food, where food is a discrete unit, analogous to a seed.

- **Spatial Distribution:** Food is placed randomly in space at one of one, four, or 16 piles, or distributed uniformly (Fig. 5.1A). These values were chosen to correspond to food distributions in known ant habitats [56, 65, 68]. Piles were never placed within 20 grid cells of the nest.
- **Food Volatility:** Food piles are moved at a specified rate to new locations in the grid to simulate growth and decay of resources (Fig. 5.1B). The volatility rate corresponds to the number of times an ant can make a round trip from the nest to a food pile before the pile has moved. We use round trips as the measurement of volatility because it reduces the values for ant speed and the distance to food to a single quantity. Food volatility was set to be either stable (no volatility), or it was moved at a rate of ten, five or 2.5 round trips. Rates of less than 2.5 round trips eliminated the value of recruitment, while rates above 10 round trips did not show behavioral difference from the stable scenario.

### 5.1.3 Model Design and Implementation

We use the model to assess the performance of three different strategies across 16 different environments, using a GA to find good parameter values for each strat-

Chapter 5. *Environmental Determination of Ant Foraging Strategies*

egy/environment pair (Section 2.2.2). This process mimics natural selection, which occurred over evolutionary time scales as each ant species evolved. Each run of the simulation evaluates a single colony of 64 ants foraging on a  $200 \times 200$  two-dimensional grid over one simulated hour. Each cell in the grid represents an  $8 \times 8$  cm patch, so the model simulates a  $16 \times 16$  m area of flat land. The number of ants, spatial extent of the search, and its duration were each based on small desert seed-harvester ant colonies [145]. Selected runs using 320 ants showed results consistent with the main model (data not shown). The simulated ants move through the grid one cell per time-step (Moore neighborhood), foraging for 1,280 food resources (*seeds*) where each cell can contain at most one seed.

We use a generational GA with population size of 25, runs of 50 generations, tournament selection (tournament size of 2), uniform crossover, 10% Gaussian mutation, and elitism, where the single best individual in each generation is retained unchanged. Full details of the algorithm are given in Hecker (2015) [1]. Each individual represents a single ant colony, and individuals are initialized using parameter values drawn from the distribution functions shown in Table 5.1, column 3. Because the strategies are non-deterministic, each individual's fitness is determined by summing up the number of seeds collected over 8 independent runs of the simulation, where each run lasts for 7,200 time steps. At the end of each GA run, the genomes of the final population of 25 are combined by averaging each gene's values to avoid over-fitting to a specific configuration of the environment. This combined genome constitutes the resulting colony of the evolutionary run. Finally, 1,000 additional simulations are run with these strategy parameters to assess variance of foraging performance for the strategy/environment pair.

Parameter.	Value
Search Area	$16m \times 16m$
Search Time	1 hour
Colony Size	64 ants
Food Quantity	1,280
Food Distribution	[1, 4, 16, 1,280] piles
Food Volatility	[Stable, 10, 5, 2.5] R.T.

Table 5.2: **Environment Parameters.** The model exists as a 2D grid on which ants forage for a simulated hour. 1,280 food items are placed in 1, 4, 16, or 1,280 distinct piles to examine the effect of the food’s spatial distribution on foraging. Similarly, food volatility is varied to be either stable, moved after 10 ant average round trips to the nest, 5 round trips, or 2.5 round trips.

## 5.2 Experimental Results

An optimal parameter set was evolved for each combination of strategy, food distribution, and food volatility. Because the model is stochastic, 1,000 runs of the model were then performed with these fixed parameters to generate 95% credible intervals<sup>1</sup> for each experiment (Fig. 5.3). Pheromone and nest recruitment perform well when food is stable and arranged in large piles. Random search performs the best in environments with highly dispersed food. This is to be expected as pheromone and nest recruitment leverage information about their environment to improve performance; increasing dispersal and volatility decreases the amount of information gained by finding food. In a majority of cases, nest recruitment performs as well as, or only slightly worse than pheromone recruitment.

The performance of all strategies equalizes when food is uniformly distributed in space (1,280 ‘piles’ of single seeds). In this scenario, the collection of one food item gives no information about the location of any other, and the recruitment strategies cannot out-perform random search (as predicted by Flanagan (2011) [25]). Because

---

<sup>1</sup>Contains inner 95% of model outcomes.

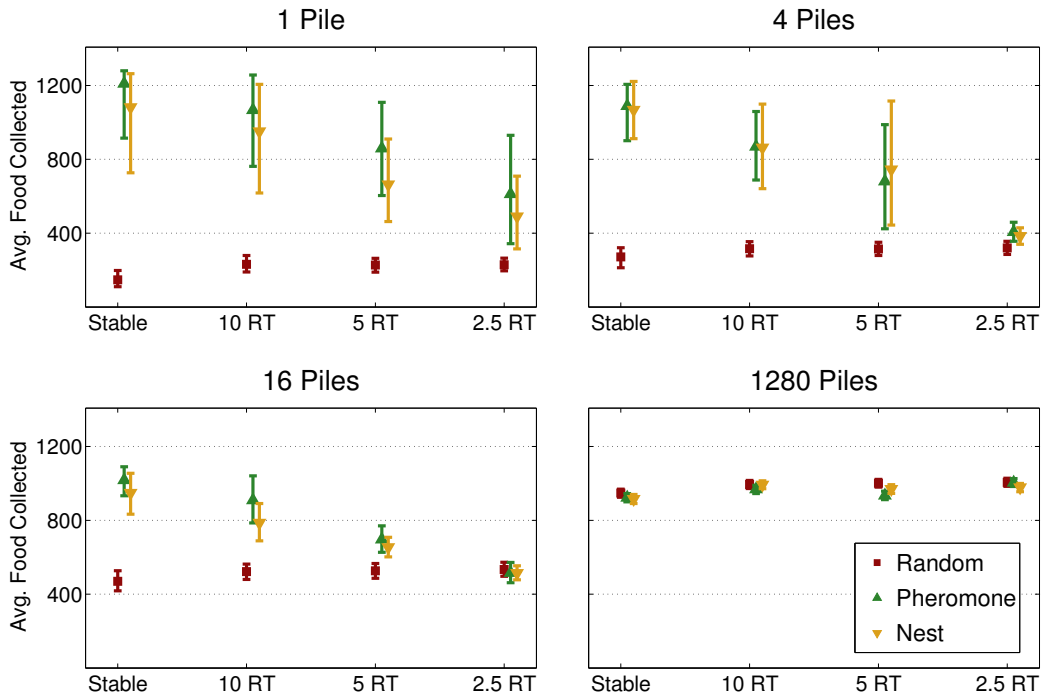


Figure 5.3: **Search Performance vs. Volatility and Spatial Distribution.** Random, pheromone, and nest recruitment evaluated over 16 environments (four pile counts by four volatility rates). Each bar represents the 95% credible interval resulting from 1,000 runs of the model after an optimal parameter set has been evolved for the specific environment-strategy pair. The two recruitment strategies show decreased performance both as the number of food piles increase and as the food volatility rate increases. The decrease in variance as pile count and volatility increases represents the dominant effect of finding and exploiting clustered piles quickly in the low pile, low volatility environments.

each strategy uses an optimized parameter set for each experiment, and because random search outperforms both recruitment strategies for 1,280 piles, the recruitment strategies evolve parameters that eliminate information exchange among ants. This explains the similarity of the results in the 1,280 pile environment (Fig. 5.3). Similarly, information becomes less valuable in highly volatile environments. At volatility rates of 2.5 round trips, recruitment is only effective in the single pile scenario. In

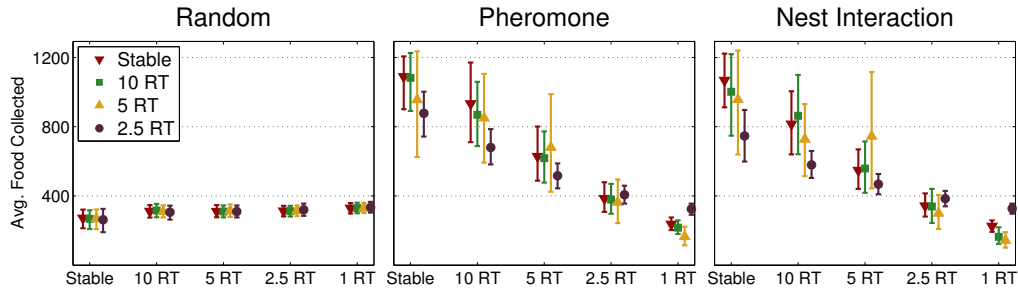


Figure 5.4: **Fixed Parameters Applied to Other Volatility Rates.** Four pile parameter sets evolved for a specific volatility rate applied to four pile environments with different volatility rates. The consistency of random search illustrates its independence from volatility rate. Pheromone strategy shows little difference between parameter sets evolved for different volatility rates, suggesting a level of robustness. Conversely, nest interaction shows increased performance by strategies evolved for the specific environment, suggesting environment specialization.

these cases of both maximum food piles and high volatility, recruitment strategies evolve away the use of any form of recruitment and behave as random search.

Because random search does not use any memory, food volatility has a minimal effect on its efficiency. Ant colonies using a random walk use similar parameter sets independent of volatility. Conversely, pile distribution does have an effect on random search efficiency. Tight clustering of food resources hinders random search even though total food quantity is held constant. There is also a positive relationship between pile size and overall variance (Fig. 5.3). Food spread evenly through space results in very consistent searches for all ants. Conversely, the distance from the nest to large piles of food will have a strong effect on the result of any given run. This effect is minimized as volatility increases because high volatility leads to multiple random pile sites.

Evolved values for the correlation in the random walk tend to be small: on the order of 0.1 radians. These values lead to relatively straight search vectors. Because ants return to the nest after finding food, and because their search paths

are relatively straight, food hidden behind nearer patches will be found last. The number of straight trajectories from the nest that find food is proportional to the sum of the diameters of the food piles, which scale as the square root of the size of the pile, not counting overlap. This means there are fewer straight trajectories from the nest that intersect food in clustered environments, and may explain why the random search strategy does worse there. Evolved values for the uninformed turning coefficient for the two recruitment strategies are generally higher than those of the random search: on the order of 0.15 radians. A larger turning coefficient corresponds to a random walk closer to the nest. This difference may indicate that the recruitment strategies more thoroughly exploit food resources close to the nest.

### 5.2.1 Fixed Parameters

Pheromone and nest recruitment are efficient strategies in the 1,280 pile environments and the 2.5 round trip volatility environments only because the optimal parameter sets for each reduce them to random search. We used fixed parameter sets that enforce the use of informed search to better evaluate the recruitment strategies in these environments.

Each strategy evolved unique parameter sets for each level of volatility. Here, parameter sets from specific volatilities were applied to each of the other volatility rates as well as the extremely volatile one round trip case for the four pile food spatial distribution (Fig. 5.4).

The results show a consistent decrease in performance by the two recruitment strategies as volatility is increased. Recruitment of ants to a localized pile works until the pile moves, at which point ants are recruited to an area that no longer contains food. The fixed parameter strategies are not outperformed by the random strategy in the 2.5 round trip environment, suggesting 2.5 round trips may be an approximate threshold beyond which recruitment does not work. This may be because while

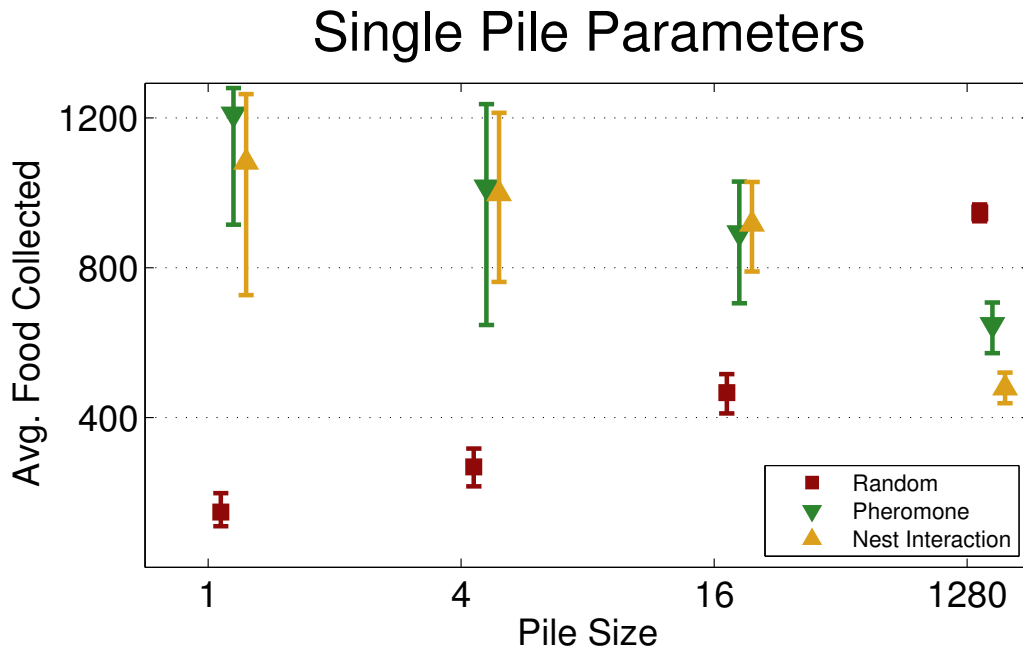


Figure 5.5: **Fixed Parameters Applied to Other Pile Sizes.** Parameter sets evolved for the stable single pile environment applied to stable environments of all pile distributions. Recruitment strategies designed for a highly clustered environment do poorly as food becomes more distributed. Randomized search strategies evolved for a clustered environment perform well in distributed environments.

highly volatile, persistence of 2.5 round trips still offers ant colonies enough time to leverage information before the pile disappears. Recruitment strategies evolved for more stable environments perform worse than the effectively random search strategy in the extreme volatility case, showing that recruitment strategies can actively hinder search when used in the wrong environment.

Further analysis reveals that pheromone success is not dependent on whether the pheromone parameters were evolved for the proper volatility rate, except in the extreme case: specifically the parameters evolved for the 10 round trip volatility case perform *worse* for that environment than the parameters evolved for the stable case. Conversely, nest interaction performance is always highest by the parameter

set evolved for that specific volatility rate. The best example of this in the five round trip environment, where the five round trip strategy improves on the others by over 33% (Fig. 5.5).

As explained in the previous section, the effectiveness of the random strategy is not strongly correlated with the volatility rate. The results of Figure 5.4 confirm this and also highlight the increase in performance as volatility increases.

Because each parameter set is evolved for a specific volatility rate, we expect that it should out-perform any parameter set evolved for a different volatility rate. This holds true except for the pheromone 10 round trip data point, where the parameters evolved for the stable environment perform the best. This is likely due to the similarity between the stable and 10 round trip environment, and the inability of the GA to evolve a globally optimal set of parameters for each environment due to computational limitations.

Similar to the fixed volatility parameter runs, results show that recruitment strategies evolved for clustered environments perform poorly in a spatially distributed environment (Fig. 5.5). In each case, recruitment of other ants to the location of a previously collected food source leads to a now empty area, actively hindering search. Conversely, a random strategy evolved for a clustered environment is able to perform well in more distributed environments as it does not get stuck looking for more food in the same location.

## 5.2.2 Variable Volatility

A consistent environment is not a reasonable assumption in the real world. To test the ability of each search strategy to cope with an uncertain level of volatility, we evolved new parameter sets in the 1, 4, and 16 pile cases where volatility was chosen uniformly at random to be between 1.25 round trips and stable for each iteration. Assuming 2.5



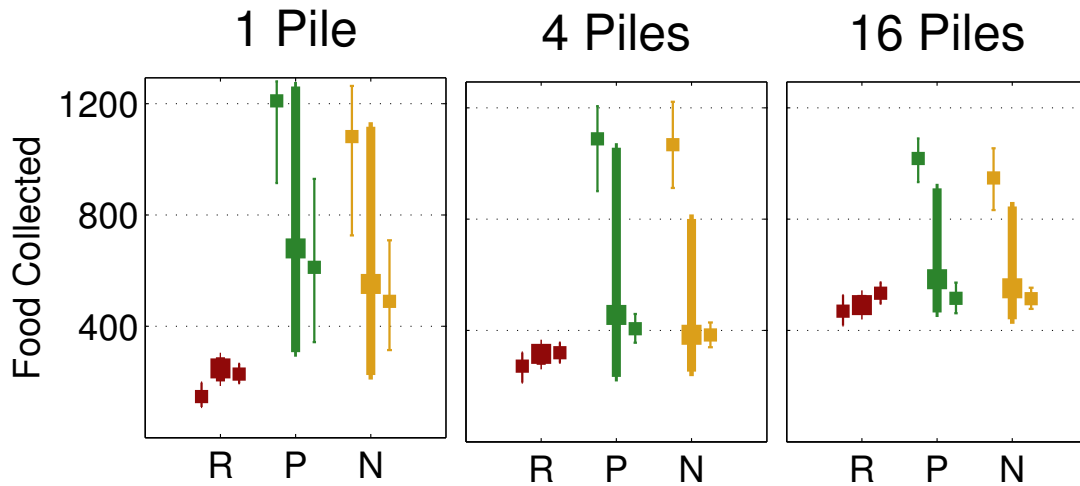


Figure 5.6: **Search Performance in Random Environments.** Parameter sets were evolved for environments with variable volatility. Results are plotted as mean values inside their 95% credible intervals. Each result (thick middle) is plotted between results for the stable environment (left) and results for the 2.5 round trip environment (right). The results show a slight performance increase versus the 2.5 round trip scenario for the two recruitment strategies. The variance of the results for the recruitment strategies are large, suggesting the use of recruitment in random environments is helpful.

round trips is a reasonable threshold between environments where recruitment may be used beneficially and not, this produced a distribution of environments where half would benefit from recruitment and half would not (the functional volatility parameter is inversely proportional to the round trip unit). Similar to the original experiment, 1,000 runs were performed (each with a random volatility rate) once the parameters were set by the GA to generate credible intervals (Fig. 5.6).

Similar to previous results, the success of the random strategy is not strongly affected by volatility. A slight increase in performance versus the stable environment is consistent with results shown in Figures 5.3 and 5.4. Pheromone and nest recruitment strategies show slight improvements over the 2.5 round trip environments. Because random search results in low variance, the large variance of the recruitment

strategies, as well as the general improvement in performance, shows that the recruitment strategies evolve parameter sets that make use of information, even when that information is short-lived.

## **5.3 Discussion**

Ant colonies are known for their ability to efficiently exploit food sources in their environment through decentralized strategies involving many simple individuals. The fact that ants use different foraging strategies in different environments [23] suggests that each strategy has been selected and tailored through evolution to perform well in that environment. We used a spatial computational model to study this hypothesis, simulating three general and customizable strategies that subsume most known biological ant behavior. The results show that information-based strategies, such as pheromones and nest recruitment, perform worse as food becomes more spatially distributed and volatile. Success of random search, which does not rely on information, is not affected by food volatility, and performs better when food is widely dispersed. In extreme cases of dispersion and volatility, information-based searches perform worse than random search. Ants foraging in environments with unpredictable volatility are able to improve their performance only slightly using recruitment.

### **5.3.1 Ant Foraging Dynamics**

Foraging strategies that are optimized for one environment can be detrimental in other environments. For example, when we optimized the (nest and pheromone) foraging strategies in stable environments, they performed poorly when resource volatility increased, in the extreme performing worse than random search. A similar pattern was observed when strategies were optimized for and studied on clustered environments and applied to environments with high food dispersal. Both recruitment

strategies perform worse than uninformed random search. However, in environments with unpredictable volatility, nest and pheromone recruitment strategies outperform random search, suggesting that recruitment is a powerful mechanism even in highly randomized environments. Finally, in environments where resource locations are sufficiently predictable, pheromone-based strategies are efficient, similar to Flanagan (2011) [25]. In most cases, however, nest recruitment performs similarly to pheromones and requires a simpler mechanism (local interaction).

In addition to specific insights about ant foraging strategies, our results suggest that new approaches could be adapted into ACO algorithms. Since pheromone-based recruitment is nontrivial to implement in a fully distributed artificial system, nest recruitment could be an attractive alternative. Moreover, the modeling approach used here could be used to classify more generally which distributed search strategies perform best in which environments.

### 5.3.2 Implications For Robot Swarms

Chemical pheromones provide foraging ants with a stigmergic, mass recruitment method that is highly scalable, fully decentralized, and generally tolerant of environments with little or no volatility. Robot swarms that mimic ant pheromones, on the other hand, are restricted to foraging in tightly controlled environments that require complex, monolithic infrastructure. Examples include stigmergic mechanisms using an always-on ink pen and white paper flooring [81]; a tightly-coupled video camera, video projector, and vision processing system [82]; and a phosphorescent-painted floor combined with ultraviolet light emitters [83].

The results of this study demonstrate that nest recruitment strategies are at least as efficient as pheromone recruitment strategies for many environments. Nest recruitment is relatively simple to implement in robot swarms, while pheromone recruitment requires robot- and environment-specific infrastructure to track the global state of

the pheromone trails and communicate the information to each robot as needed. Further, the foraging success of nest recruiters depends only on local, agent-to-agent communication, while pheromone recruiters often depend on global coordination with a single point of failure. We therefore suggest that research in swarm robotics should focus less on mimicking ant stigmergy, and more on designing and evaluating new decentralized information-sharing protocols that are more scalable and easier to implement in natural environments as foraging strategies for real robots.

### **5.3.3 Comparison To Biological Ants**

Ants are known to use different foraging strategies in different environments, and there are many classifications of strategy environment pairs [23, 56]. Here, we provide a computational rationale for which strategies have evolved for which environments, focusing on spatial distribution and volatility of food. Our results support the hypothesis that ant foraging strategies are tailored carefully for their environment through the processes of natural selection. As we mentioned earlier, although ant colony optimization methods focus on pheromone recruitment. The results suggest that other known ant foraging strategies may also prove useful to computational problems of distributed search and that these other approaches may outperform pheromone recruitment in the right environments [58, 65, 67–69, 143].

Our results show that pheromone and nest recruitment are successful in stable clustered environments and that random search works best in environments of high dispersal. We evaluated these statements by comparing them to a comprehensive review of physical ants and their habitats [23]. Lanan categorizes the use of ant search strategies over four environmental dimensions, one being spatial distribution of food and another being frequency of food occurrence, which is similar to volatility. Of the 402 species of ants examined in Lanan (2014) [23], 58 were able to be classified completely into non-overlapping categories.

## Chapter 5. Environmental Determination of Ant Foraging Strategies

Of these 58, 13 forage in environments of high food dispersal: seven use random search to forage, three use long term trails. Of the three remaining species, two use a form of nest recruitment in what can be considered moderately volatile, which agrees with our model. The three species that use long term trails forage in a space of high food abundance, such that a trail to a specific location will not exhaust the resources located there. Our model did not explore the effects of high food abundance.

Of the rest of the 58 categorized ant species, 39 forage in environments of high spatial clustering of food. All but four of these use long term pheromone trails, as predicted by our model. Of those, one uses site fidelity in a resource rich area, one is listed as random although the author notes they visit the same location repeatedly, one harvests insects in a highly volatile environment, and one forages randomly and seems to be an exception worthy of future study.

Of those, the *camponotus sericeus* uses site fidelity in a resource rich area [57], *formica pallidefulva* are listed as random although the author notes they visit the same location repeatedly, *ectatomma brunneum* harvest insects near quickly decomposing carcasses which creates a highly volatile environment, and *ectatomma opaciventre* forage randomly and seem to be an exception worthy of future study.

Thus, of the physical ants able to be classified into categories defined by our model, our model immediately agrees with 80.7% of the observations, with an additional 9.6% consistent with the addition of site fidelity and food volatility.

### 5.3.4 Caveats

Our model attempts to answer the question: how do different search strategies perform in different environments? To answer this question, we limited the model to three strategies and two environmental dimensions. To avoid over-parameterization and excessive computational complexity, these strategies and environments were im-

## *Chapter 5. Environmental Determination of Ant Foraging Strategies*

plemented in a simpler form than their biological analogues. Ant species vary widely in size and speed, are known to use multiple strategies in concert, and often maintain multiple nests, all factors not included in our model. By keeping our model simple, we are able to answer direct questions regarding the direct relationship between specific components of ant recruitment and environmental conditions.

Parameters such as the size of the search area, the number of ants in the colony and the quantity of food were fixed and not explored. These parameters are set on the small end of plausible biological ranges for computational tractability. Other possible environmental characteristics such as ant predation, food quality were not included at all. It is reasonable to assume that variations in any of these properties would affect search efficiency and should be explored in future work.

Finally, ant colonies in our model are given the benefit of the doubt and are allowed to use optimal parameters for their specific environment. We assume that these parameter values approximate those used by ants in the real world, but have not done a direct comparison. Certain parameters are difficult to measure in nature and the values of these evolved parameters may be informative.

Ants are known to use many other types of foraging strategies beyond the three we studied here, including site fidelity [58, 65], tandem running [67], raids [69], trunk trails [68, 146], and fans [147, 148]. Ants are also known to use hybrid strategies. Further, ants inhabit a wider range of environment classifications, including areas of different food density [149], food size [150], and danger [151]. Finally, some ant species are known to be nomadic [152] while many others are polydomus (multiple nests) [72, 153]. Computational modeling is an important tool to supplement field observations and can help elucidate why ants use which strategies in which circumstances, how they manage their interaction pathways.

## 5.4 Summary

The phrase ‘ant foraging’ in the context of computer science is nearly synonymous with the use of pheromone trails. However, field studies have shown that numerous ant species do not use pheromone recruitment. This suggests that there are environments for which alternative foraging strategies are at least as efficient as the use of pheromone, or that pheromone based search can be detrimental to the nest.

Lanan (2014) [23] cataloged hundreds of species of ants to create a classification of ant foraging strategies given their environment. Here we analyze three of these foraging strategies across two environmental parameterizations: spatial distribution and volatility of food. We find that nest recruitment performs nearly as well as pheromone recruitment in all environments, and that simple random search is more efficient than either when resources are highly dispersed or volatile. Our results, coupled with observations by Lanan (2014) [23], suggest ant species have evolved the use of optimal foraging strategies for their environment.

Understanding how and why ants use different strategies in different environments is critical for biology-inspired algorithmic design. In many cases, an algorithm ill-suited to its environment will perform worse than a simpler naive strategy. Knowing when and how to use these simpler strategies may improve distributed search and swarm robotics.

## Chapter 6

# Analytical Model of Ant Foraging Behavior

In this chapter I present a conceptual analytical model of expected ant foraging rates as determined by the environment. Chapter 5 uses an agent-based CPFA model to demonstrate that the spatial distribution and temporal volatility of resources constrain possible foraging strategies. In this chapter I attempt to classify the specific scaling relationships between a foraging strategy's expected collection rate and the environmental conditions in which it operates. The agent-based model of Chapter 5 is not ideal for calculating mathematical scaling relationships. Therefore, in this chapter I derive and evaluate analytical models of both uninformed foraging and foraging using nest recruitment.

To develop a simplified analytical model of ant search, I examine ballistic and auto-correlated random movement and conclude simple straight-line exploration is a reasonable assumption in finite space. In developing an equation to describe foraging rates of ant colonies using recruitment, I also derive an equation to determine an optimal number of scouts in a colony and determine the threshold where volatility



## *Chapter 6. Analytical Model of Ant Foraging Behavior*

affects foraging, both as determined by known parameters of the colony and the environment.

The CPFA model from Chapter 5 shows both pheromone and nest recruitment perform best in clustered stable environments, that their efficiency declines as food dispersion and volatility increases, and that recruitment strategies optimized for one environment can be detrimental in other environments. I use the analytical model to show how foraging rates scale versus the spatial distribution of food, the volatility rate of food, the quantity of food, the search area, and the size of the colony. I also determine an optimal scout population and identify the threshold beyond which volatility effects take place. Finally, I confirm the results of Chapter 5 that recruitment works well in environments of spatially clustered stable resources, and that proper recruitment allows a large proportion of the ant colony to collect food at optimal rates. Further, I identify when recruitment strategies falter and how random search improves in environments of increasing dispersion. Given knowledge of a foraging population and its environment, the analytical model is able predict foraging rates using either uninformed search or recruitment.

### **6.1 Analytical Model of Ant Foraging**

The agent-based model of Chapter 5 captures the behavior of ant foraging in different environments and illustrates the effects of food distribution and volatility on the three modeled foraging strategies. Due to the black-box nature of a GA-optimized spatial model, it is challenging to classify the exact relationship between the environment and the various foraging strategies. Further, it is difficult to reason about the parameter values chosen by the GA. Therefore, we developed a simplified analytical model of instantaneous ant foraging rates to complement our agent-based CPFA model (Fig. 6.1).

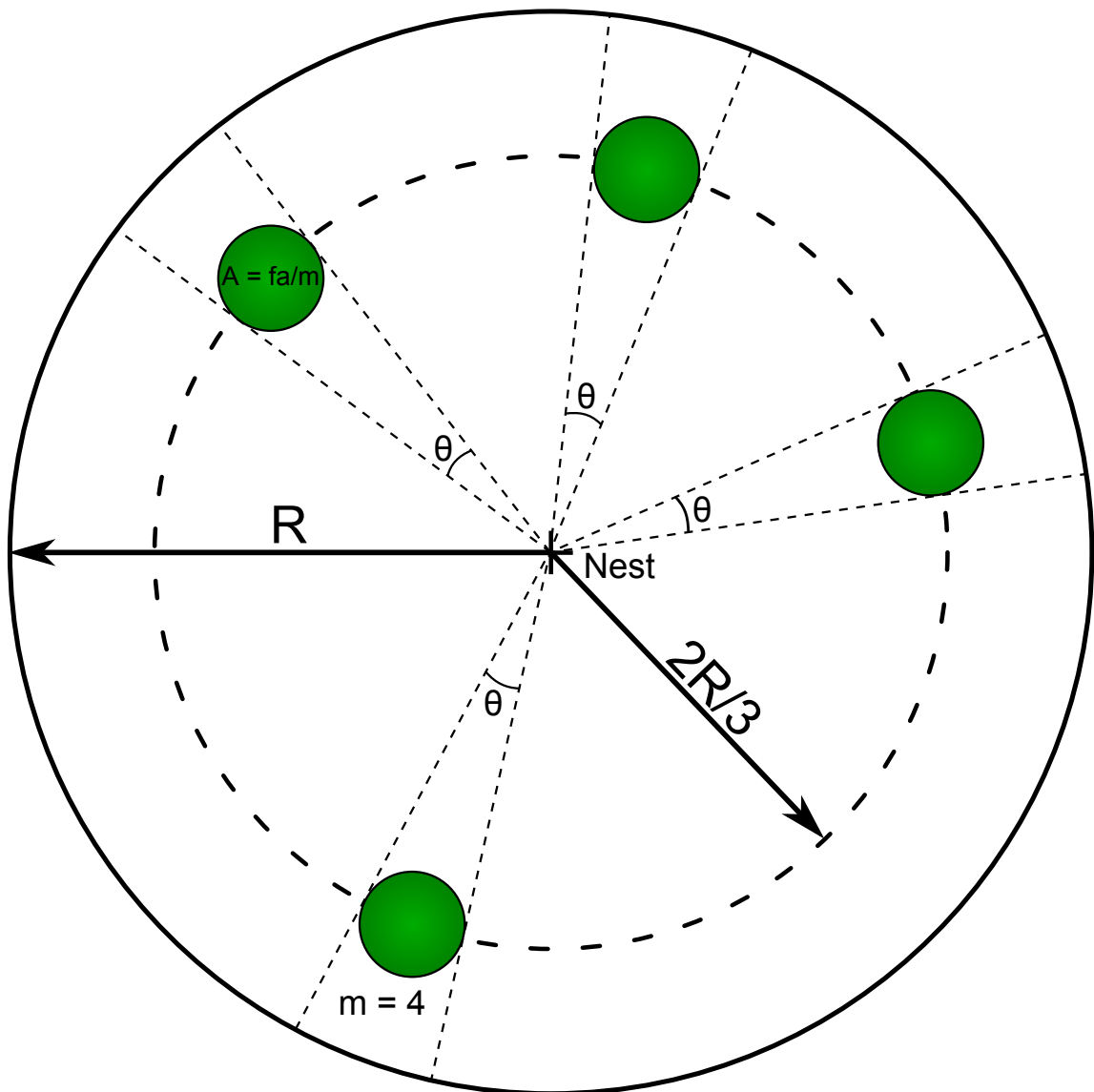


Figure 6.1: **Analytical Model Design.** Ants leave the nest traveling in a random direction. On trajectories that miss food, ants travel a distance  $R$  before giving up and returning to the nest. On trajectories that find food, ants travel a distance  $2R/3$  before encountering the food pile (the expected distance between the center and a random point inside the circle). The area of each pile is calculated as the total area of food ( $fa$ ) divided by the number of piles ( $m$ ). In this figure, there are four piles ( $m = 4$ ).

While temporal effects of foraging such as ramp-up and depletion are important, we focus on instantaneous rates to avoid the use of intractable non-linear differential equations. First, we derive a simple equation describing the maximum possible foraging rate as a way to test the optimality of different strategies. Next, we develop equations to describe expected foraging rates for uninformed search and nest recruitment. Pheromone search is not modeled explicitly because the ABM showed that nest recruitment has similar behavior at a lower bound and because pheromone recruitment is difficult to describe using simple equations. In generating a simple analytical model, we examine the spectrum of Brownian versus ballistic motion to justify the use of ballistic motion in our equations. Finally, in creating an equation for the expected nest recruitment foraging rate, we generate an equation describing the optimal proportion of scouts versus followers in terms of the known parameters.

### 6.1.1 How Well Can Ants Do?

An optimal foraging rate is equal to the expected rate of food gathered, assuming each individual ant has perfect information about its environment. Because we are interested in instantaneous rates without depletion effects, we focus on the rate of foraging food at the average expected distance from the nest. For food distributed uniformly at random in a circular area of radius  $R$  with a nest at the center, the expected distance of food from the nest is  $2R/3$ . Assuming ants always find food on their first attempt, a reasonable approximation of the instantaneous optimal foraging rate ( $df/dt$ ) of food is the inverse of the time it takes an ant to make one trip to the food and back multiplied by the number of ants ( $n$ ) in the colony, where  $s$  is the speed of an ant:

$$F_{opt} = n \cdot \frac{df}{dt} = \frac{3ns}{4R} \tag{6.1}$$

### 6.1.2 Random Walks

Random search processes for the purposes of foraging must balance the trade-off between prioritizing discovery of resources near the nest against the drawback of wasting time spent searching area that has already been seen. A random walk with no directional autocorrelation (e.g. Brownian motion) excels at exploring the nearby vicinity of the nest, but does poorly at avoiding previously visited locations. At the other extreme, straight-line ballistic motion never visits the same location twice, but rapidly leaves the local area. A hybrid strategy, such as a Levy random walk [154] or the autocorrelated random walk used in the CPFA model, may increase foraging rates, but rates using these strategies can be difficult to quantify analytically.

Discrete time Brownian motion describes the process of taking a fixed-length step in a randomly chosen direction each time step. Brownian motion has been studied in depth, e.g. [155], and is known to behave as a diffusion process [156]. Although it is possible to calculate the spatial probability density function of an agent performing Brownian motion. given the walker's step size and rate, it is a much more complicated process to calculate first-passage properties of such a walk. The term first-passage, or hitting time, refers to the expected time it takes a walker to first enter a specific sub-region of the walker's domain for the first time. Since we are interested in random walks as a search process, we must focus on these first-passage relations. In the case of ant foraging, we are interested in the time it takes an ant to first encounter food. Further, because we are interested in colony-wide search by many ants at once, we must account for the first discovery of food by one of many possible foragers. This requires knowledge of the full cumulative probability distribution of hitting times.

The canonical Redner (2001) [157] contains exact solutions for single Brownian random walkers in one dimension, and approximations for single walkers in multi-dimensional systems that exhibit symmetry. Evans and Majumdar (2011, 2014)

give expected first-passage times for single walkers in the vicinity of a single target [158, 159], but do not provide solutions for multiple searchers or multiple targets. While the dynamics of one-dimensional random walks are well understood, there are no known approaches for an analytical solution to first-passage times of multiple walkers in the presence of multiple targets in multi-dimensional systems.

Even were we able to derive an analytical solution of searching as a Brownian random walk, it would prove unnecessary. Brownian motion is a poor search strategy and is not used in nature. The highly auto-correlated nature of Brownian motion traps the walker near its origin. Rather, organisms from birds [154], to mammals [160] forage in patterns classified as Levy walks or flights [161] or a close approximation [162]. Levy walks are classified as random walks where the length of each step is drawn from a power-law distribution, resulting in a significant proportion of large displacements. These large hops ensure that an organism will not remain trapped near its origin while still allowing it to initially prioritize regions near its starting point.

After comparison of straight-line ballistic motion, Levy-like auto-correlated random motion, and Brownian random walks, we choose to model uninformed ant foraging as a simple ballistic straight-line walk with periodic resetting. Ballistic motion has been shown to be optimal in simple systems of destructive foraging [163] and we corroborate this result in Section 6.2.1. We present this model as a tractable analytical model of uninformed search and use it to estimate a lower bound on ant foraging efficiency given knowledge of the spatial distribution and volatility of food resources.

### **6.1.3 Ballistic Model of Uninformed Search**

We model the expected foraging rate of an uninformed ant traveling in a straight line with resetting after travel of a fixed distance by examining the expected number of

Chapter 6. Analytical Model of Ant Foraging Behavior

attempts required to find a patch of food (Fig. 6.1). Ants leave the nest in a direction chosen uniformly at random and continue in a straight line until they either find food or travel a fixed distance  $R$ , the radius of the search area. For simplicity, we assume food patches are located at the average expected distance from the nest:  $2R/3$  given that they are distributed uniformly at random inside the circular area of radius  $R$ . We calculate the expected instantaneous foraging rate for  $n$  ants as defined by the parameters listed in Table 5.2.

Assuming  $m$  food patches, each of equal size, with  $f$  collectible food objects, each with individual area  $a$ , the diameter  $d$  of a single pile of food is given by:

$$d = 2\sqrt{\frac{fa}{m\pi}} \quad (6.2)$$

Assuming patches are located at a distance  $2R/3$  from the nest, the angle in radians ( $\theta$ ) of a pile as seen from the nest is:

$$\theta = 2\sin^{-1}\left(\frac{3d}{4R}\right) \quad (6.3)$$

The probability of hitting any pile is equal to one minus the probability of missing every pile. If every pile has an angle of  $\theta$  and piles are independently distributed in a circle around the nest, the probability of hitting at least one pile,  $p$ , is:

$$p = 1 - \left(\frac{2\pi - \theta}{2\pi}\right)^m \quad (6.4)$$

If we assume that an ant that misses food will travel distance  $R$  and then return to the nest to try again, the expected time taken to discover food can be represented as a geometric probability series where  $s$  is the speed of an ant:

Chapter 6. Analytical Model of Ant Foraging Behavior

$$\frac{dt}{df} = \sum_{i=0}^{\infty} (1-p)^i p \left[ \frac{(6i+4)R}{3s} \right] \quad (6.5)$$

The series converges to the following equation, which gives the expected time taken for one ant to discover food.

$$\frac{dt}{df} = \frac{2R(3-p)}{3sp} \quad (6.6)$$

Finally, the expected rate of resource collection is simply the inverse of the expected time taken to collect one food resource:

$$\frac{df}{dt} = \frac{3sp}{2R(3-p)} \quad (6.7)$$

Because uninformed ants act independently with no exchange of information, it can be assumed that the rate of foraging of a colony is simply the rate of foraging of a single ant times the number of ants ( $n$ ) in the colony:

$$F_r = n \cdot \frac{df}{dt} = \frac{3nsp}{2R(3-p)} \quad (6.8)$$

This equation is of a similar form to Eq. 6.1. Combining the two produces a simple equation giving the efficiency of the random search when compared to the optimal foraging rate in terms of the probability of an ant finding food during a single trip:

$$\frac{F_r}{F_{opt}} = \frac{2p}{3-p} \quad (6.9)$$

### 6.1.4 Analytical Model of Nest Recruitment

We next develop an equation to describe the expected foraging rate of an ant colony using nest recruitment. Nest recruitment requires task allocation between scouts performing uninformed search and reserve ants waiting at the nest to be recruited by successful scouts. First we give an equation that predicts the optimal proportion of scouts versus reserve foragers in a colony given the food distribution and volatility of the environment. We assume that an optimal proportion of scouts versus foragers will balance the time it takes scouts to find a new pile of food with the time it takes the remaining population to collect the pile. Were scouts to find piles faster than the followers could collect them, foraging would be improved by moving some scouts to the foraging population. Were foragers able to collect a pile before scouts could find a new one, foraging would be improved by moving some foragers to the scout population.

First, we solve for the expected time for  $x$  scouts to discover a single pile and return. We use the ballistic model of random search from the previous section to describe the search process. The expected time to discover a pile is related to the probability of discovering a pile for the first time after  $\tau$  trips. The probability of at least one ant succeeding for the first time on trip  $\tau$  is the probability that no ants were successful before trip  $\tau$  times the probability that at least one ant found a pile on the last trip:

$$P(\tau = i) = q^{x(i-1)}(1 - q^x) \tag{6.10}$$

Where  $q = 1 - p$ , as determined by Eq. 6.4 and  $x$  is the number of scouts.

For a given number of trips  $\tau$ , the total search time  $T(\tau)$  is:



$$T(\tau) = \frac{2R[3(\tau - 1) + 2]}{3s} \quad (6.11)$$

The expected value of  $T$  is given by the summation from  $\tau = 1$  to  $\infty$  of  $P(\tau = i)$  multiplied by  $T(\tau)$ :

$$E[T] = \sum_{i=1}^{\infty} [q^{x(i-1)}(1 - q^x)] \left[ \frac{2R[3(i - 1) + 2]}{3s} \right] \quad (6.12)$$

which simplifies to

$$E[T] = \left[ \frac{2R(1 - q^x)}{3s} \right] \sum_{i=0}^{\infty} q^{xi}(3i + 2) \quad (6.13)$$

The summation is a geometric series that reduces to  $(2 + q^x)(1 - q^x)^{-2}$ . Combining this term with the part of the equation before the summation of Eq. 6.13 gives a formula for the expected time for the first of  $x$  ants to find food:

$$E[T] = \frac{2R(2 + q^x)}{3s(1 - q^x)} \quad (6.14)$$

In the optimal case, this time is equal to the time it takes the remaining ants to collect a discovered pile, where scouts optimally discover food at the same rate that the remaining foragers can collect it. To quantify this time, we quantify the discovery of a pile in terms of how much food the pile is worth to the foraging population. There are two scenarios where the pile collection is complete. First, the ants could collect all the food. Second, the pile could move as defined by the volatility rate  $v$ . For the first case, the discovery of a pile is worth the amount of food in the pile:  $f/m$

Chapter 6. Analytical Model of Ant Foraging Behavior

minus the one food resource that the scout collects. For the second case, the pile is worth the number of round trips the  $(n - x)$  reserve foragers can make to the pile before it moves. Given the pile is in existence for  $v$  round trips, the scout discovers the pile on average at time  $v/2$ . It then takes one round trip to the nest and back before the foragers can start collecting food from the pile. Thus, the expected total number of trips foragers may take to a discovered pile is  $(n - x)(v - 1)/2$ . Therefore, we quantify the amount of food that the discovery of a pile is worth ( $k$ ) with the following equation:

$$k = \min \left( f/m - 1, \frac{(v - 1)(n - x)}{2} \right) \quad (6.15)$$

Given a pile worth  $k$  food items, the expected time taken for the remaining  $n - x$  ants moving at speed  $s$  to collect the contents of a single pile located at a distance  $2R/3$  from the nest is:

$$E[T] = \frac{4Rk}{3s(n - x)} \quad (6.16)$$

Setting this equal to the expected time for the scouts to find a food pile (Eq. 6.14) and simplifying gives us the following equality. Solving for  $x$  gives an optimal proportion of scouts versus foragers.

$$\frac{2k}{n - x} = \frac{2 + q^x}{1 - q^x} \quad (6.17)$$

This equation is irreducible in terms of  $x$ . Fortunately, numerical approaches to equation solving are well understood [164]. We use the Python `scipy.optimize.brentq` function to find values for  $x$  given Eq. 6.17. Once a solution is obtained

for  $x$ , it is a simple matter to calculate the expected foraging rate of the colony when using nest recruitment. We assume the  $(n - x)$  recruited ants forage with perfect information at a rate equivalent to the optimal predicted rate from Eq. 6.1. The  $x$  scouts forage at a rate related to their uninformed search, as described in Eq. 6.8. Once simplified, this gives an expected foraging rate for a colony using nest recruitment of:

$$F_n = n \cdot \frac{df}{dt} = \frac{3s[(n - x)(3 - p) + 2xp]}{4R(3 - p)} \quad (6.18)$$

## 6.2 Model Results

We have generated equations describing the expected foraging rate of ants using both random search and nest recruitment given food distributed in  $m$  piles and a volatility rate,  $v$ . To do so, we made the assumption that straight-line ballistic motion was a reasonable approximation of ant movement. We also derived an equation to describe the best possible foraging rate and an equation to determine the optimal number of scouts versus reserve foragers for a colony performing nest recruitment. In this section we use these equations to examine the predicted scaling relationships between the placement of food in the environment and the expected foraging rates and the implications of these dependencies.

### 6.2.1 Auto-Correlated Random Walk Versus Ballistic Motion

We assume simple straight-line ballistic motion in our analytical model of ant foraging. To test the validity of this assumption, we evaluated a basic form of the CPFA simulation to evaluate search efficiency as a function of the random walk uninformed search coefficient,  $\sigma_u$ . At each time step, an uninformed forager will alter its trajectory by a random value chosen from a normal distribution with mean 0 radians

and standard deviation  $\sigma_u$ . The value  $\sigma_u = 0$  corresponds to ballistic motion and the limiting case of  $\sigma_u = \infty$  corresponds to discrete time Brownian motion. The simplified CPFA simulation implements uninformed search using parameters from Table 5.2, with the search area,  $A$ , varied over three grid sizes ( $200 \times 200$ ,  $280 \times 280$ ,  $400 \times 400$ ).

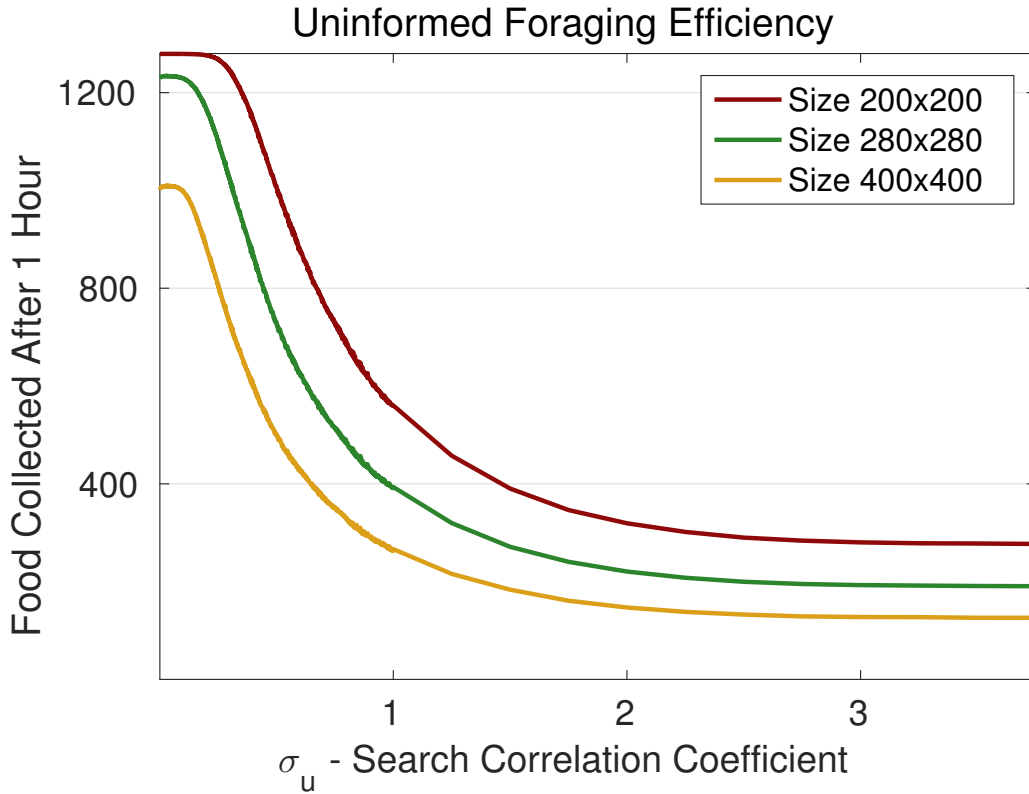


Figure 6.2: **Success of correlated random walks.** The CPFA model of uninformed search was evaluated over a wide range of values of  $\sigma_u$ , the uninformed search correlation coefficient. Low values of  $\sigma_u$  imply less variance in a walk’s trajectory, corresponding to straighter movement. High values of  $\sigma_u$  correspond to high variance in a walk’s trajectory, corresponding to a Brownian motion random walk. The model was evaluated 1,000 times at each sampled value of  $\sigma_u$ ; 95% confidence intervals are narrower than the line width. Our results suggest ballistic movement performs better than random walks in bounded space as predicted by [163] and justifies our use of ballistic motion in the analytical model.

The results confirm our hypothesis that ballistic motion is a reasonable, and apparently optimal, search strategy for uninformed ants in confined space (Fig. 6.2). Further, it demonstrates the predicted poor performance for discrete time Brownian motion.

### 6.2.2 Scaling of the Uninformed Foraging Rate

The CPFA model shows an increase in performance of random search foraging efficiency as the number of separate piles increases and no difference in performance as the volatility rate changes (Fig. 5.3). We classify the relationship between the foraging rate and our model parameters using the predictions of Eq. 6.8 to investigate more precise scaling relationships. In the specific case of pile size, we plot the results of the analytical model on a log-log axis (Fig. 6.3) and observe a slope near one half, suggesting a square root relationship between the number of distinct piles of food and the foraging rate of uninformed ants. Intuitively, ants performing a search on a two-dimensional surface stop searching once they encounter food. Therefore, any food in the middle of a pile is not available to be discovered. As food is broken up into smaller piles, the perimeter to area ratio of food resources increases as the square root of the number of piles.

Here we simplify our analytical model to demonstrate this relationship and others. Our calculation relies on two critical assumptions. First, this relation holds if food is sparse enough that the chance of two piles sharing a trajectory is relatively small. Second, this relation assumes the center of food piles are located farther from the nest than the length of their diameter. Foraging becomes trivial due to a necessarily high density of resources if either of these assumptions are invalid. These effects can be seen in the eventual curvature of the lines in Figure 6.3 and the divergence of the lines in Figure 6.4.

The rate of uninformed foraging is proportional to the probability of finding food

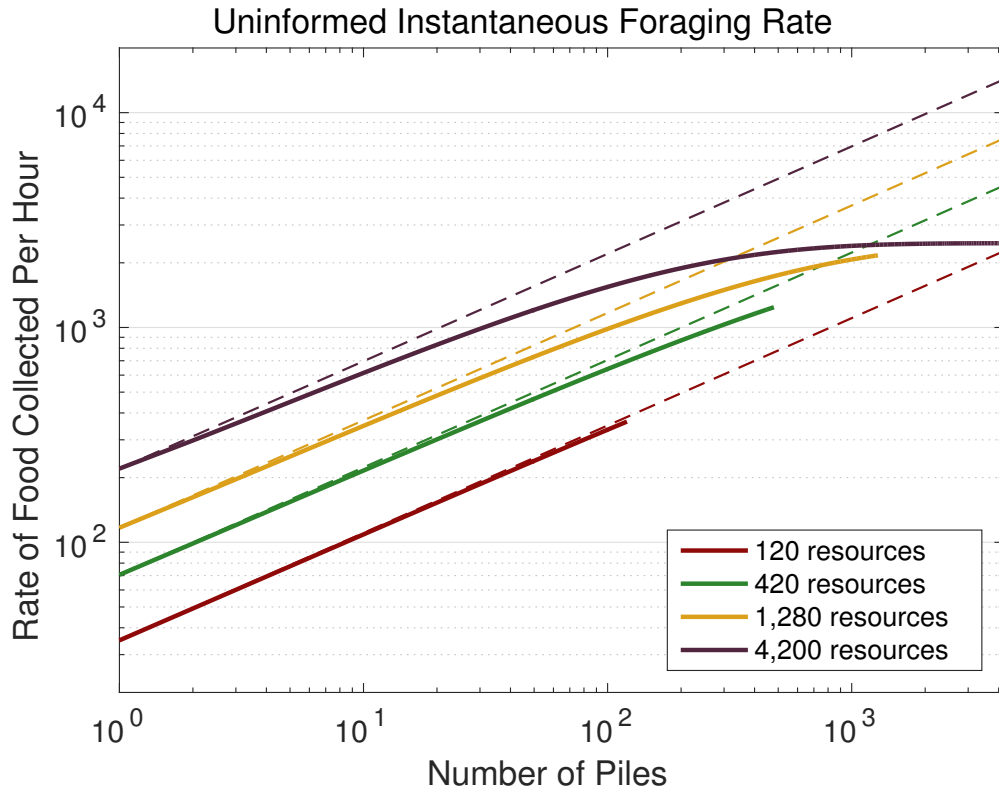


Figure 6.3: **Expected instantaneous foraging rate of uninformed ants.** Eq. 6.8 predicts foraging rates for uninformed ants in terms of the number of distinct food piles. Uninformed search is unaffected by volatility. Expected foraging rates are shown on a log-log plot to demonstrate the predicted square root relationship between the number of distinct food piles and the expected foraging rate (dashed lines).

(Eq. 6.7). Since we are interested in the scaling relationship, we focus directly on Eq. 6.4 where  $m$  is the number of food piles and  $\theta$  is the angle of the pile as seen from the nest (Fig. 6.1). Expanding  $\theta$  from Eq. 6.3 gives the following equation where  $d$  is the diameter of a single pile and  $2R/3$  is the expected distance from the nest to the pile:

$$p = 1 - \left[ \frac{2\pi - 2\sin^{-1}\left(\frac{3d}{4R}\right)}{2\pi} \right]^m \quad (6.19)$$

Chapter 6. Analytical Model of Ant Foraging Behavior

Assuming piles are located farther away from the nest than their diameter implies that  $\theta$  is relatively small. We use the small angle assumption to assume  $\sin^{-1}(x) \approx x$  and simplify:

$$p \approx 1 - \left[ 1 - \frac{3d}{4\pi R} \right]^m \quad (6.20)$$

If food is relatively sparse to the point that piles are unlikely to share a trajectory, we can say that the  $\frac{3d}{4\pi r}$  term is significantly less than one. For small  $x$ , the term  $(1 - x)^n$  approximately equals  $1 - nx$ . Substituting and simplifying gives:

$$p \approx \frac{3dm}{4\pi R} \quad (6.21)$$

The final step is to note that the  $d$  term is defined in Eq. 6.2 and is proportional to  $m^{-1/2}$ . Substituting and simplifying results in a final equation proportional to  $\sqrt{m}$ :

$$p \approx \frac{3}{2\pi R} \sqrt{\frac{fam}{\pi}} \quad (6.22)$$

Where  $f$  is the total amount of food available and  $a$  is the area per unit of food.

To test the accuracy of Eq. 6.22, we plot the relationship between the probability of finding food on a ballistic trajectory and the number of piles the food is divided into (Fig. 6.4). As predicted, the approximate relation of Eq. 6.22 breaks down when the sparse assumption is violated. Further, as food becomes increasingly dispersed, the probability of multiple piles sharing a trajectory increases, causing the approximation to overestimate the probability. In sparse environments, the relationship does hold, especially in environments without a wide dispersal of food (Fig. 6.4B Inset).

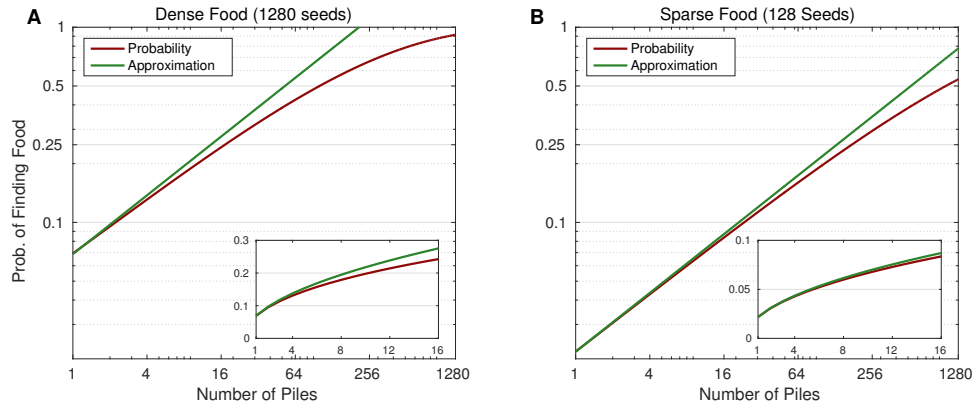


Figure 6.4: **Random Search Improves as the square root of the number of piles.** A simplified version of the model (Eq. 6.22) predicts a square root scaling relationship between the probability a trajectory intersects food and the the number of piles food is divided into. This relationship holds as long as food is sparse enough that overlapping effects do not come into play. Inset plots are versions of the same plot zoomed in to show 1 to 16 piles. The approximation of Eq. 6.22 holds in sparse and/or clustered environments and begins to diverge when the probability of a trajectory intersecting food passes 0.2.

Beyond the square root relationship with  $m$ , Eq. 6.22 also suggests a square root relationship with the density of food,  $f$ ; an inverse linear relationship with the length of the search area’s radius,  $R$ ; no relationship with the volatility rate,  $v$ ; and a linear relationship with the number of foragers,  $n$ , as predicted by Dornhaus (2006) [73] (Table 6.1).

### 6.2.3 Analytical Predictions of Ants Using Nest Recruitment

We use Eq. 6.18 of the analytical model to predict expected foraging rates of ants using nest recruitment. The predicted instantaneous rates (Fig. 6.5A) mirror the results of the CPFA model (Fig. 5.3). Further, the analytical model predicts the optimal ratio of scouts to the reserve population given the environment (Fig. 6.5B) and describes the threshold under which volatility affects foraging (Eq 6.15). In environments with high food clustering, recruitment works well: the discovery of a



Parameter	Foraging Scales As	
	Random	Recruitment
$m$	$O(\sqrt{m})$	$x \sim \text{Eq. 6.24}$
$v$	$O(1)$	$x \sim \text{Eq. 6.25}$
$n$	$O(n)$	$O(n)$
$f$	$O(\sqrt{f})$	$O(1)$
$R$	$O(R^{-1})$	$O(R^{-1})$

Table 6.1: **Scaling relationships of ant foraging.** The analytical model predicts how the foraging rate scales with respect to various parameters of the model. Random search scaling relationships are taken from the simplified model described in Eq. 6.22. Recruitment scaling analysis is limited to environments where recruitment has been shown to perform better than random search. In these cases, recruitment foraging rates are largely determined by Eq. 6.1. Recruitment does not scale directly in terms of  $m$  and  $v$ , instead  $m$  and  $v$  determine the optimal number of scouts which in turn defines the proportion of ants in a colony that are able to be recruited and forage optimally.

patch of food is valuable, a large proportion of the colony remains at the nest to be recruited, and the overall foraging rate is high.

As food becomes moderately dispersed, the value of pile discovery and subsequent recruitment decreases, leading to lower foraging rates. As food becomes uniformly distributed in space the value of recruitment disappears, but the efficiency of random search and thus overall search performance increases as more ants leave the nest to search for food on their own. Similarly, in environments of high food stability, the value of pile discovery is high and a large proportion of ants remain at the nest in reserve. As volatility increases, the value of discovery and recruitment disappears and the optimal scout population approaches zero.

Recruitment gives the advantage of allowing a significant proportion of foragers the ability to forage at optimal rates, as described by Eq. 6.1. To this end, recruitment foraging rates do not suffer from the same scaling relationships as uninformed foraging rates, as long as the environment is amenable to recruitment. There are

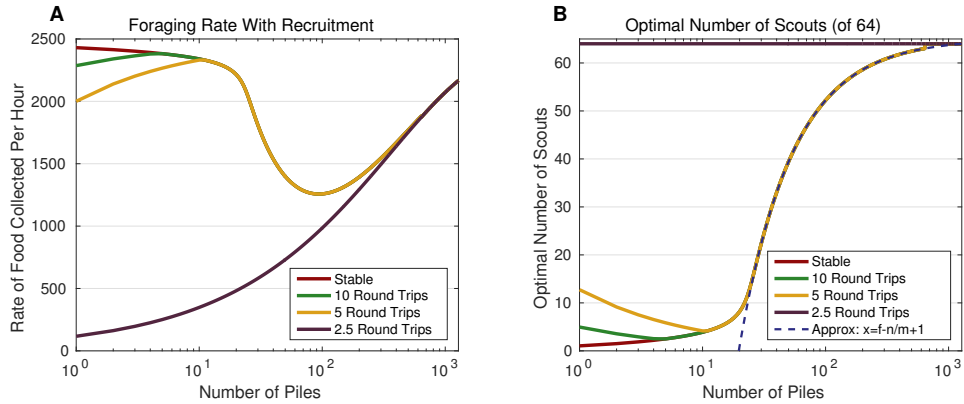


Figure 6.5: **Expected instantaneous foraging rate of ants using nest recruitment.** **A)** Eq. 6.18 predicts the foraging rate of an ant colony using nest recruitment. Foraging is improved by recruitment in environments of high food clustering and stability. As food becomes moderately dispersed and volatile, the value of recruitment decreases. Once food becomes uniformly dispersed and/or highly volatile, the value or recruitment disappears, but the expected foraging rate of uninformed search increases. **B)** the optimal number of scouts (uninformed foragers) is low in regions of high clustering and stability. Recruitment becomes less valuable in environments of higher food dispersion and volatility and thus the optimal proportion of independent uninformed foragers increases. For non-small values of  $m$ , the optimal number of scouts approximately equals  $f - n/m$  (dashed line). The x-axes are shown in log scale to better illustrate the effects of low food dispersion.

two ways recruitment can break down: food can be too volatile, and food can be too dispersed. Our model is able to identify these thresholds. As described by Eq. 6.15, volatility only affects foraging below a certain threshold. Solving Eq. 6.15 for  $v$ , the volatility rate results in an equation that describes this threshold explicitly:

$$v < \frac{2(f/m - 1)}{n - x} + 1 \quad (6.23)$$

Foragers are limited by food volatility only if this inequality holds. Otherwise, foragers are expected to exhaust a pile of food before it disappears. As resource volatility increases (and  $v$  decreases), the optimal number of scouts tends to zero

and recruitment is replaced by random search as predicted by Eq. 6.17. If volatility does not affect foraging, the optimal scout population depends on the dispersion of food, as described by  $m$ , the number of unique piles. As with volatility, as dispersion increases, the number of scouts tends toward zero and recruitment is once again replaced by random search (Fig. 6.5B). Thus, the foraging rate of colonies using nest recruitment does not scale in a traditional sense directly with the volatility and distribution of food resources, but as a secondary effect of the transition from optimal to uninformed foraging. Although Eq. 6.17 is irreducible, it contains multiple  $q^x$  terms that tend to zero as  $x$  increases. Simplifying and substituting  $f/m - 1$  for  $k$  (assuming no volatility effects) gives a simple approximation of how the optimal number of scouts scales with the number of distinct resource piles:

$$x \approx n - f/m + 1 \tag{6.24}$$

This relationship breaks down for small  $x$  where the  $q^x$  term is non-zero, but is highly accurate otherwise (Fig. 6.5B, dashed line). Similarly, it is also possible to approximate how  $x$  scales in terms of the volatility rate. Substituting the volatility term from Eq. 6.15 for  $k$ , and assuming  $2 + q^x \approx 2$  results in the following approximation:

$$x \approx \log_q \left( 1 - \frac{2}{v-1} \right) \tag{6.25}$$

The logarithmic nature of this relationship means  $x$  tends to  $n$  (scouts become dominant) very quickly as the volatility rate increases (and  $v$  decreases).

#### 6.2.4 Robustness of the Scout Population

Our analytical model of nest recruitment suggests an optimal proportion of scouts in a foraging population. This value holds under optimal circumstances but can break

down quickly. In comparison to the predicted scout population of the analytical model (Fig. 6.5B), the evolved parameters of CPFA model (data not shown) use a much higher proportion of scouts. A large reserve population is a costly proposition: if food is not discovered, a large portion of the colony is idle. Conversely, an excess of scouts is not costly: the reserve population will stay active and excess scouts will continue to find and gather food.

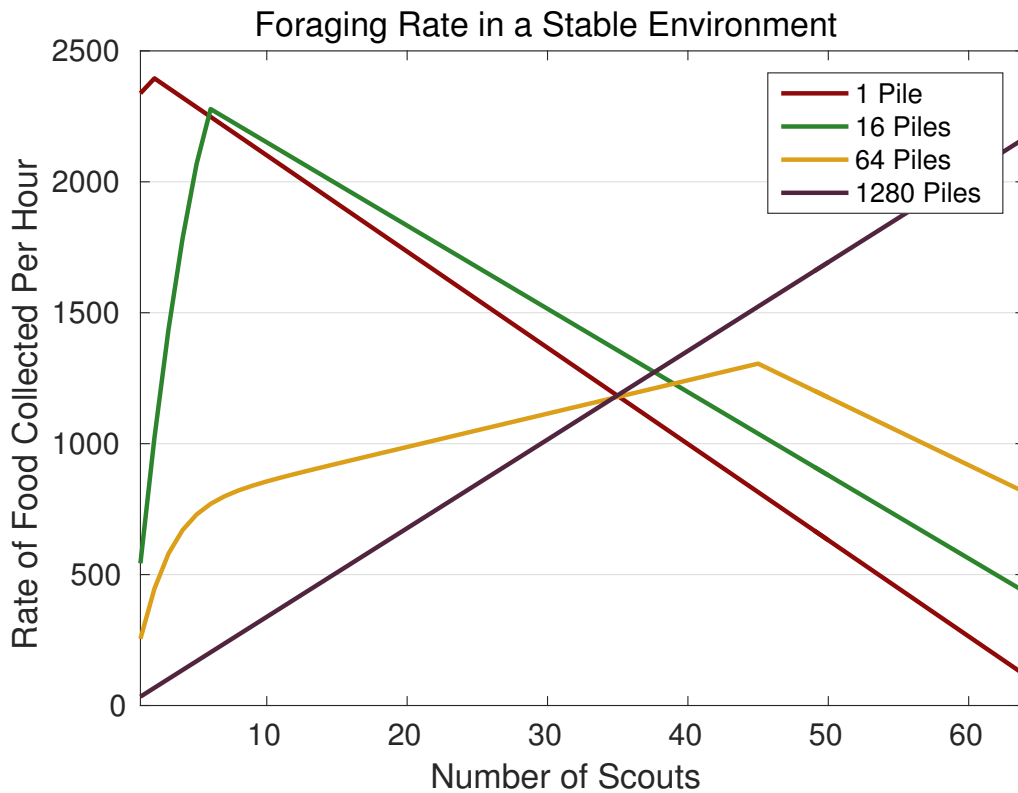


Figure 6.6: **Foraging rate in a stable environment.** The expected instantaneous foraging rate of ants using nest recruitment as a function of the number of scouts used as predicted by the analytical model. In environments of high food clustering, the drawbacks of excessive scouts is minor compared to the drawbacks of too few. In more dispersed environments, the drawbacks are still minor, if present at all (64 and 1,280 pile environments). This effect is even more pronounced in volatile environments (not shown). Using more scouts than the analytical model suggests is a more robust option for real ant colonies.

We examine the expected foraging rate of stable environments as a function of the number of scouts in the colony (Fig. 6.6). In clustered environments, a lack of scouts leads to a dramatic drop in performance, while an excess results in only a mild decrease. In environments of medium and high food dispersion, the penalty of excess scouts remains low, and may even be nonexistent. Given that environments are dynamic and may be changing, and given that ants do not know the exact conditions of their environment *a priori*, it makes sense that real ant colonies and our evolved CPFA colonies would use a higher proportion of scouts than predicted by our analytical model.

### 6.3 Assumptions and Conclusions of the Analytical Model

For simplicity, we limited the analytical model to two foraging strategies: uninformed search and nest recruitment. The similarity of performance between nest recruitment and pheromone recruitment seen in Chapter 5 suggests the predicted performance of the analytical model of nest recruitment is also a reasonable predictor of pheromone recruitment performance. When designing the analytical model, several simplifying assumptions were made beyond those already listed in Section 5.3.4. We primarily use expected values in our calculations of the analytical models. A deeper understanding of the effect of random variation would likely be informative and is a candidate for future work. Specifically, the calculation of the optimal scout population assumes that scout discovery and reserve collection events are synchronized in time, which is not realistic. Nest recruitment with random discovery events would benefit from a larger proportion of scouts, an effect demonstrated by both the evolved CPFA parameters (data not shown) and our evaluation of scout population robustness (Fig. 6.6). Therefore, foraging rate estimates of the analytical model are likely high.

In developing a conceptual analytical model of ant foraging, we evaluate the efficiency of ballistic versus random movement. Our results corroborate that of Viswanathan (2000) [163] by demonstrating that ballistic motion is optimal in the constrained environment (Fig. 6.2). We then use the assumption of ballistic motion to describe both an analytical models of uninformed ant foraging and foraging using nest recruitment. These models describe basic scaling relationships between expected foraging performance and modeled environmental factors (Table 6.1). Further, the analytical model of foraging using nest recruitment reveals the threshold for volatility effectiveness and identifies the optimal ratio of scout and reserve populations (Fig. 6.5B). Finally, we demonstrate that the GA optimized CPFA model uses higher scout populations than predicted by the analytical model by examining the trade-offs of having too many versus too few scouts (Fig. 6.6).

## **6.4 Summary**

The model results are consistent with previous findings [1, 62, 64, 73, 75], which considered resource distribution or volatility in isolation. The results confirm that colonies foraging for clustered resources use recruitment-based strategies to exploit information, while colonies foraging for randomly distributed resources avoid recruiting and instead focus on efficient correlated random search. Further, colonies are most efficient when foraging on the distribution for which they are evolved, although some foraging strategies are sufficiently flexible to function well on different distributions.

In this chapter I extend this work by developing a conceptual analytical model of both uniformed and recruitment-based foraging. The results formalize key scaling relationships between the search strategies and different environmental factors, identify a threshold on volatility effects, and suggest an optimal ratio of scout versus reserve population when performing nest recruitment foraging. Knowledge of these

*Chapter 6. Analytical Model of Ant Foraging Behavior*

relationships will prove useful in the design and evaluation of distributed computational algorithms.

# Chapter 7

## Conclusion

This dissertation examines how the environment affects natural examples of distributed search mechanisms. I focused on two specific examples of distributed search in nature: T cells searching for pathogens in the human lung and ants foraging for food near their colony. In both instances, I use numerical and agent-based models fit to empirical data to examine difficult to measure aspects of the search process. My spatial model of the T cell search reveals challenges to clearance caused by a rapidly spreading infection. The issue arises because the highly virulent 2009 pandemic influenza infection moves faster than the chemokine gradient used to direct T cells to the focus of infection. This result suggests drawbacks to chemical-directed stigmergic search in volatile and dispersed environments.

I further investigate this theory with a spatial model of ant foraging behavior. The model examines how different foraging strategies perform in different environmental conditions. Similar to T cell search, ant foraging strategies based on stigmergic pheromone trails suffer in environments of high dispersion and volatility. I find that while strategies that use various forms of communication such as chemical and local recruitment perform well in stable and clustered environments, these strategies run



the risk of leading naive ants to areas of little value in the wrong environment. The spatial model suggests that in these environments, uninformed random search performs better than recruitment approaches. These results are then confirmed by my analytical model of both uninformed and recruitment-based ant foraging.

## **7.1 Summary of Results**

The main results of this dissertation are presented in chapters 3 through 6.

An agent-based model of T cell search in the human lung requires knowledge of many parameters that are difficult or even impossible to measure directly. In chapter 3, I make use of several preliminary numerical and agent-based models fit to empirical data to estimate values for previously unknown parameters affecting the immune system process of infection and response. Specifically, I examine three separate strains of influenza chosen for their differences in pathogenicity: H5N1 avian flu, H1N1 seasonal flu, and H1N1 2009 pandemic flu. The models are consistent in predicting unique viral replication rates for each strain of influenza, and demonstrate how avian influenza may proliferate by suppressing the innate immune response.

I then use the knowledge gained from Chapter 3 to parameterize and evaluate a spatial agent-based model of the chemokine-directed T cell response to influenza infection in the human lung. The model predicts uncontrolled infection only for the 2009 pandemic strain, corroborating real-world evidence of heightened mortality rates for pandemic flu. The spatial nature of the model reveals specific challenges to chemokine-directed search that arise from the rapidly spreading frontier of infection caused by pandemic influenza infection. Two sensitivity analyses of the model show that viral kinetics determine the course of the infection.

The inability of chemokine-directed T cell search to clear the highly virulent 2009 pandemic flu strain suggests general challenges to stigmergic-based search in volatile

## *Chapter 7. Conclusion*

environments. I examine this situation in the context of ant foraging in Chapter 5. I extend a prior agent-based model of ant foraging to examine the effect of food distribution and volatility on three different foraging strategies. The model shows that recruitment strategies suffer in environments of high resource dispersion and volatility. Conversely, the performance of uninformed random search increases as resources become more spread.

While able to demonstrate the effects of food distribution and volatility, the spatial model of ant foraging used in Chapter 5 is not able to classify the direct relationship between strategies and environments. In Chapter 6 I derive and evaluate conceptual analytical models of ant foraging using both uninformed random search and nest recruitment. The model of uninformed random search clarifies the scaling relationship between the expected foraging rate and various factors of the search environment. In deriving the analytical model of nest recruitment foraging, I develop an equation to identify the threshold beyond which volatility effects dominate and an equation to predict the optimal ratio of scouts to reserve ants for given environments. The resulting equation predicts expected foraging rates using nest recruitment given parameters of the environment. Finally, analysis suggests nest recruitment performance is more robust if the scout population is over-estimated.

## **7.2 Future Work**

This dissertation suggests several possibilities for future work. First, there are logical extensions to both the agent-based and analytical ant models from Chapters 5 and 6. These lead naturally to ideas for improving swarm robotics algorithms, and finally they provide interesting new hypotheses about T cell search. While the models were able to examine the relationship between foraging efficiency and the environmental factors of resource spatial distribution and volatility, there are a number of other factors that were not evaluated.

## *Chapter 7. Conclusion*

Both of my ant models assume that the ants live and search on a continuous unobstructed two-dimensional surface. However, ant species are known to live in environments with complex topologies, such as in trees or foliage. Examining the effects of recruitment in these environments would be informative. The models could also be extended to represent costs of active foraging, such as energy expenditure, water evaporation, and predation effects. Currently there are no physical advantages for an ant to remain in the nest, other than to be eventually recruited by a scout with knowledge of a pile location.

The analytical model of ant foraging presented in Chapter 6 does not represent pheromone recruitment. While the CPFA model of Chapter 5 demonstrates that the performance of nest recruitment and pheromone recruitment is similar across environments, it would be of interest to develop an analytical model of pheromone recruitment to support better quantitative comparisons between pheromone and nest recruitment strategies.

A third area of future extension of the analytical model involves the variance of the expected foraging rate. The analytical model of ant recruitment solves for the expected foraging rate using the assumption of perfect expectation, which is unrealistic. This assumption likely causes the model to overestimate the advantages of nest recruitment. Examining the effects of variance on the predictions of the analytical model would better inform predictions regarding the optimal scout population and the subsequent expected foraging rate of a colony performing nest recruitment.

This dissertation examines the environmental effects on various ant foraging strategies, but stops short of examining computational applications of these strategies. Pheromone-based ACO methods have been applied successfully to solve a number of computational problems [77, 80, 165, 166], yet these algorithms require explicit global representation of a digital pheromone trail. Because nest recruitment requires only local interactions between agents, it would be worthwhile to design and examine

## *Chapter 7. Conclusion*

ACO algorithms based on nest recruitment, and to compare these implementation to similar pheromone-based ones.

The models of T cell search for influenza demonstrate both the benefits and drawbacks of the diffusion of chemical-based signals to create a chemical gradient that searching agents can follow. Currently, the majority of ACO algorithms implement only decay properties of pheromones [76] (as physical ant pheromones are generally considered to be static in space). Adding diffusion effects to pheromone-based ACO algorithms may allow for a signaling process that covers a wider range and is therefore better able to recruit and direct agents.

The design and implementation of robotic swarms to perform collective tasks is an emerging field. Teams of cheap robots can travel through and monitor areas that may be inaccessible or dangerous to humans [167, 168]. Data sets generated by large swarms of robots may be too large for direct human supervision, suggesting a need for automated control mechanisms [169]. The work in this dissertation can extend previous work in natural inspired robotic search mechanisms [170].

Finally, the division of an ant colony into scout and reserve populations when performing nest recruitment search suggests benefits to heterogeneous task allocation in distributed search algorithms [171]. The model of influenza infection demonstrates challenges to T cell search caused by chemotaxis, but modeled T cells as a homogenous population. It is reasonable to assume that the immune system provides mechanisms to differentiate leukocyte behavior in response to challenging infections. For example, our experimental data showed elevated levels of IP-10 and RANTES chemokines, but also measured a host of other chemokines that were not expressed during the infection. It is worth studying whether chemokines may be utilized by the immune system to direct unique responses to other types of infection. Further, our model included only antigen-specific CD8+ T cells, ignoring a number of other known leukocyte types. It would be informative to examine under which conditions

these other chemokines were expressed and how they interact with the variety of adaptive immune agents.

### **7.3 Final Remarks**

The recent rise of resource-constrained distributed systems in both computation and robotics demands new methods of algorithmic design. Traditional approaches to designing control mechanisms are not guaranteed to generalize to these emerging fields. Naturally occurring biological systems have evolved solutions to problems of distributed coordination over millions of years. Here, I study known examples of natural distributed systems to develop insights about distributed system design.

Systems studied in this dissertation have search processes whose performance is highly sensitive to its environment. Specifically, traditional recruitment-based search strategies perform worse than the simpler uninformed random search in environments of high resource dispersion and volatility. My work identifies these trade-offs and presents methods to choose the best distributed autonomous search strategy given the environment in which the search will be performed.

# Appendices

# Appendix A

## CyCells Sensitivity Analysis

This section details the methods and results of the two sensitivity analyses performed on the model presented in Chapter 4.

### A.1 One-factor-at-a-time sensitivity analysis

A general sensitivity analysis serves two purposes. First, it allows us to observe how the model behavior changes as a single parameter is varied. Second, it allows us to examine which parameters affect the model strongly and which do not. Here we describe the sensitivity analysis of 16 of our parameters in detail.

The model parameters were chosen from literature when available and estimated within plausible ranges otherwise (Tables 3.8 and 4.1). One-factor-at-a-time (OFAT) analysis was performed by varying Individual parameters over ranges of plausible (and sometimes even implausible) values while the rest of the parameter set was held constant. Each parameter was varied over all three influenza strains, creating three sets of sensitivity plots (Figures A.1-A.3).

We then categorized the model parameters into one of three qualitative groups:

## Appendix A. CyCells Sensitivity Analysis

parameters that do not affect the model’s qualitative behavior (stable parameters), parameters that affect the model’s quantitative peak infection size but do not affect final clearance, (peak change parameters), and those parameters that affect the final clearance of the infection (sensitive parameters) (Table 4.2). Parameters are presented along with their range using the following format:

**Parameter Name:** [ *min* / *default* / *max* ].

### A.1.1 Stable Parameters

Stable parameters do not affect the outcome of the infection unless they are adjusted to values outside the realm of possibility. Specifically, the plausible parameter values in this group are within a range that does not significantly affect the asymptotic behavior of the model. We further split this group into two categories in Table 4.2, *stable* parameters and *bounded stable* parameters. *Bounded stable* parameters are only stable within a specific range of values, while *stable* parameters seem to be stable over all the values that were tested.

Of interest, this group can be split into two types of parameters: those governing chemokine behavior (chemokine decay, chemokine diffusion, and chemokine secretion) and those governing T cell behavior (T cell circulation time, T cell kill rate, T cell velocity, and the two T cell decay rates). Only the apoptosis time parameter does not fit into one of these two groups, and its inclusion as a stable parameter may be suspect due to the limited range of the values tested. The importance of the apoptosis time parameter is discussed in more detail in the main results section: Temporal Effects.

The stability of these parameters makes sense in the context of our model. Chemokine exists in our model to provide a chemical gradient that T cells may



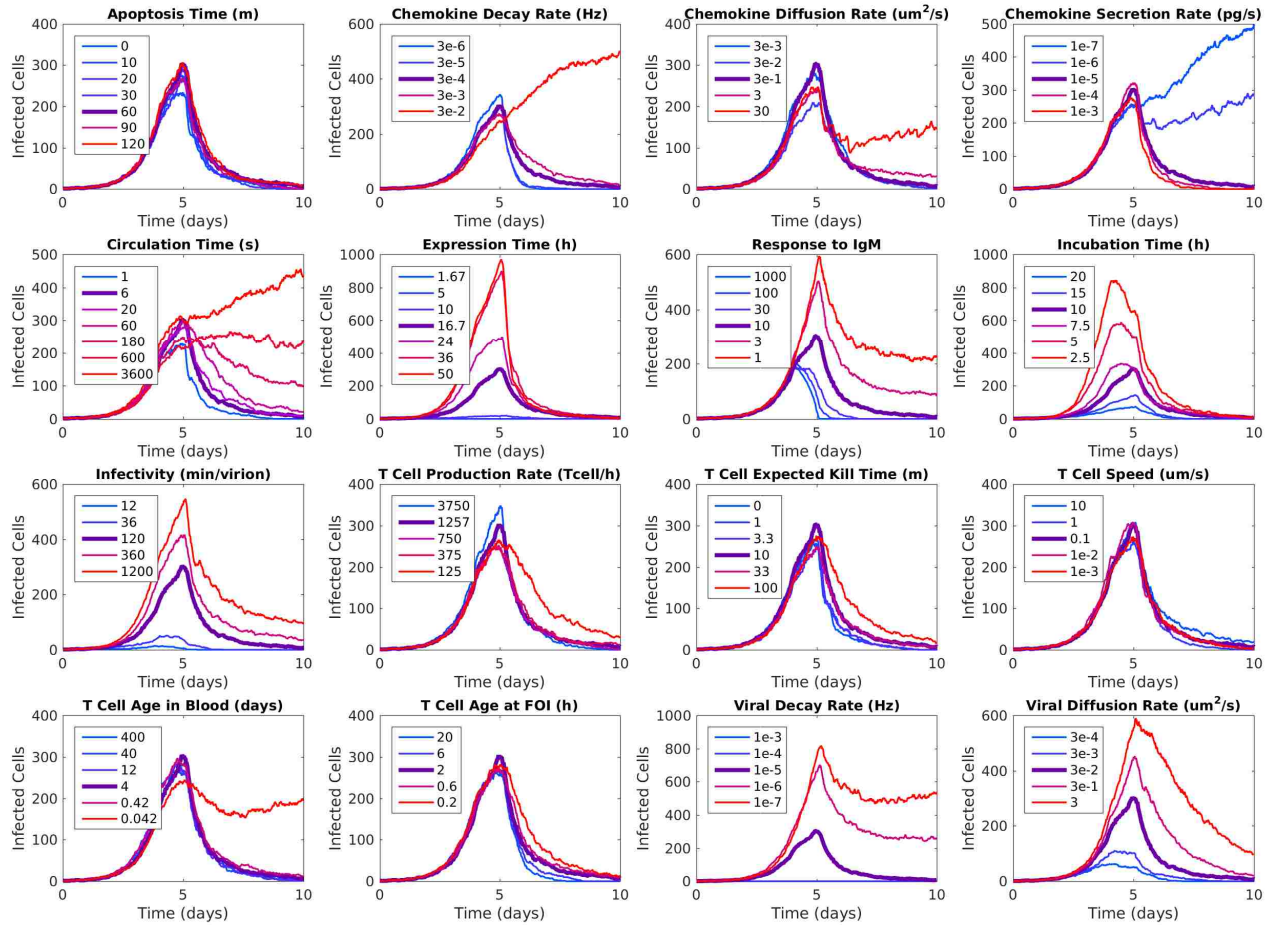


Figure A.1: **aH5N1 sensitivity analysis.** The sensitivity analysis results for 16 parameters applied to the aH5N1 model simulation. In each subgraph, the thicker purple line shows the results of the model using the default parameter value.

## *Appendix A. CyCells Sensitivity Analysis*

follow to the focus of infection. The total quantity of the chemokine in the lung does not have a strong effect on the location and size of the gradient. Thus, the model will be stable for any values of chemokine secretion, diffusion, and decay that provide a gradient that T cells are able to follow.

Similarly, T cells affect the model by clearing cells expressing virus. As discussed in the main paper, the chemokine gradient creates areas of maximal concentration that attract all the T cells inside its basin of attraction. Thus, most T cells are attracted to the same areas of the infection and overlap considerably. Increasing T cell numbers and efficiency will not help clear the infection beyond a certain point.

While we have claimed that these parameters are qualitatively stable within a certain range, many of the parameters were set to values that did lead to a divergent model behavior. Testing these extreme values provides bracketing information regarding the range of stability of the parameter in question. We have deemed these deviations acceptable on an individual basis as described below.

Compartmental modeling design suggests the removal of such stable parameters. Due to the nature of the ABM, inclusion of these parameters is mechanistically necessary. Thus, we include and parameterize them using the methods described in the main paper and Table 4.1.

**Apoptosis Time:** [ 0 / 1h / 2h ] - Apoptosis time describes how long it takes for a cell to complete apoptosis and transition to an inert dead state.

We chose to limit the increase of the apoptosis to a factor of two because larger values would not be biologically relevant. We considered values as low as zero.

Adjusting the apoptosis time causes the model results to diverge slightly in the sH1N1 strain (Fig. A.2). Also of interest, reducing the apoptosis time to zero still did not allow the immune response to clear the pandemic infection. As stated earlier,

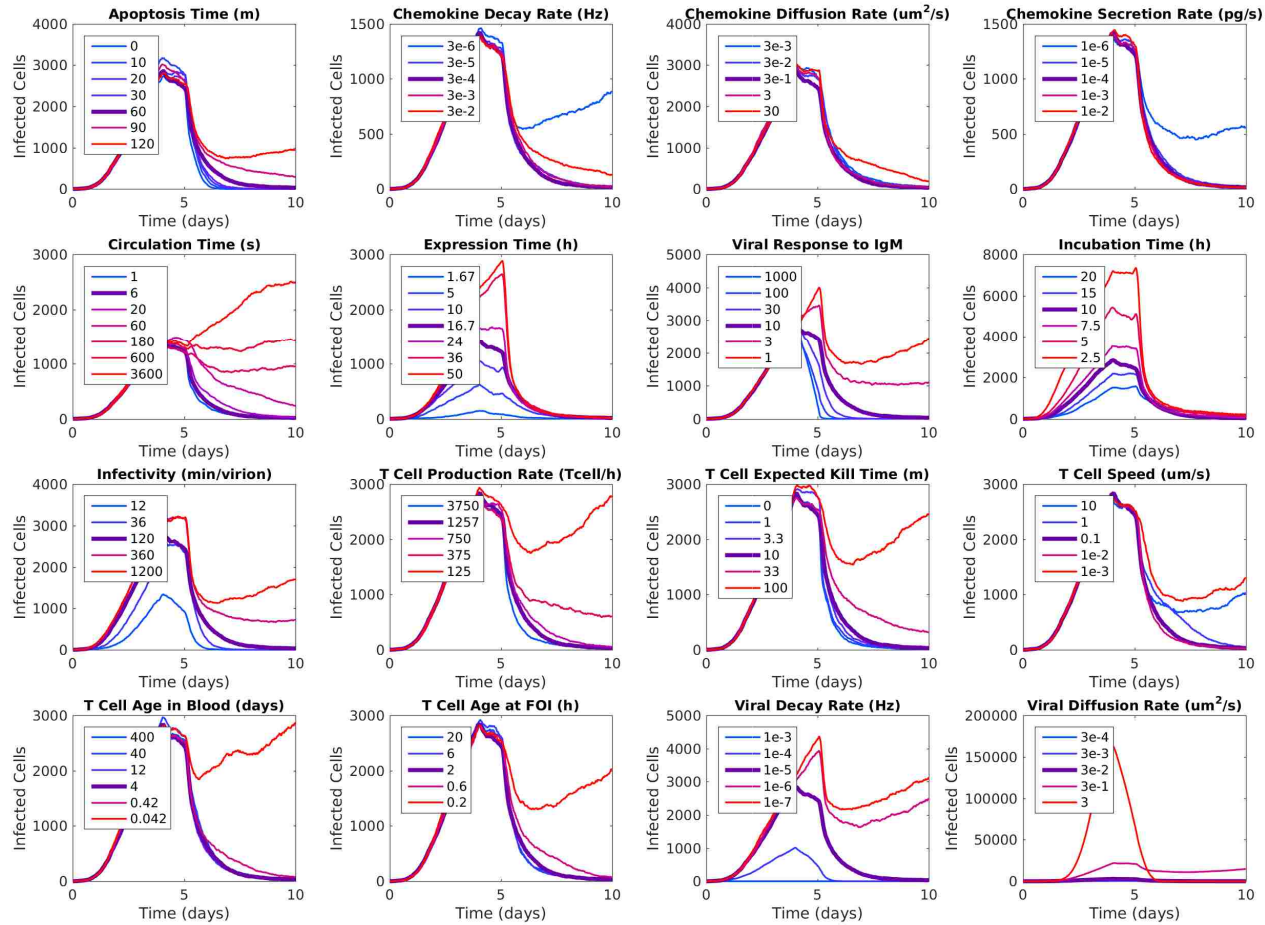


Figure A.2: **sH1N1 sensitivity analysis.** The sensitivity analysis results for 16 parameters applied to the sH1N1 model simulation. In each subgraph, the thicker purple line shows the results of the model using the default parameter value. The divergence between Apoptosis Time values of 90m and 120m (upper-left plot) highlights an important threshold of infection control.

## *Appendix A. CyCells Sensitivity Analysis*

its inclusion in this group is marginal and its effects are described in more detail in the results section of the main paper.

**Chemokine Decay Rate:** [  $3e-6$  Hz /  $3e-4$  Hz /  $3e-2$  Hz ] - The chemokine decay rate defines how quickly the chemokine is removed from the system. A higher rate of decay corresponds to a quicker rate of removal.

We tested values up to two orders of magnitude larger and smaller than the default parameter value even though these extremes are not biologically plausible.

Varying the chemokine decay rate results in stable model behavior for values one order of magnitude larger and smaller than the default. A value two orders of magnitude smaller results in divergent behavior for both the aH5N1 strain and the sH1N1 strain and a value two orders of magnitude larger results in divergent behavior for only the sH1N1 strain. This maximum decay rate used for all three strains corresponds to an implausible 18 second half-life. The minimal value corresponds to a similarly implausible 50 hour half-life. The fact that an extremely low decay rate can hinder clearance is interesting. In this case, lack of decay allows the chemokine to diffuse homogeneously across the entire infection, removing the concentration gradient required by T cells to find the active areas of the infection. This suggests that the quantity of chemokine is immaterial so long as there is enough for T cells to be able to detect it.

**Chemokine Diffusion Rate:** [  $3e-3$   $\mu\text{m}^2/\text{s}$  /  $0.3$   $\mu\text{m}^2/\text{s}$  /  $30$   $\mu\text{m}^2/\text{s}$  ] - The chemokine diffusion rate regulates how quickly the chemokine molecules spread out over the infected region. A larger diffusion coefficient corresponds to a more rapid rate of spread.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

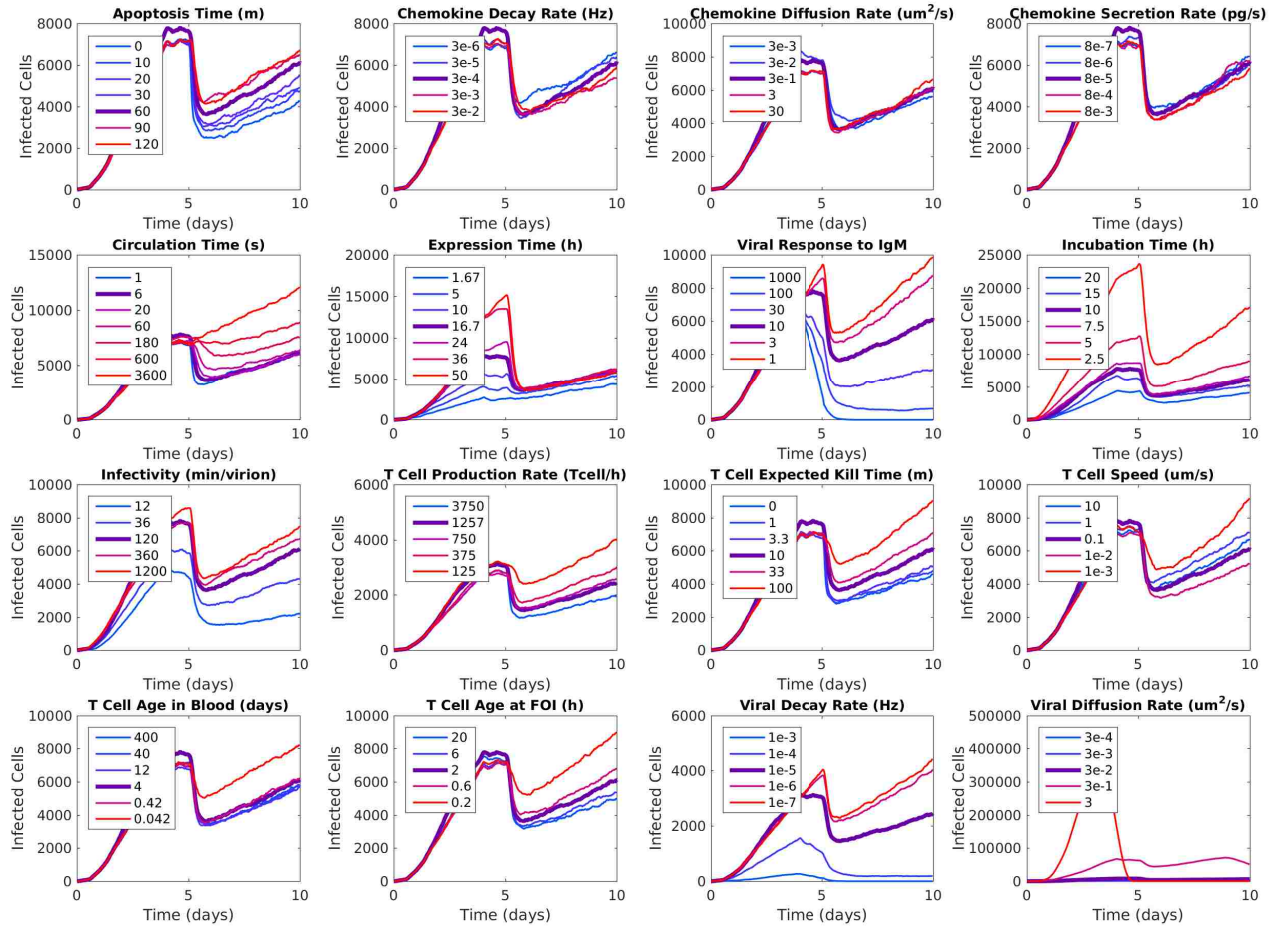


Figure A.3: **pH1N1 sensitivity analysis.** The sensitivity analysis results for 16 parameters applied to the pH1N1 model simulation. In each subgraph, the thicker purple line shows the results of the model using the default parameter value.

## Appendix A. CyCells Sensitivity Analysis

The chemokine diffusion rate only diverges in the aH5N1 strain for the highest value. Our estimate for the diffusion rate is already elevated as it is optimistically based off of the Stokes-Einstein equation using viscosity of water.

**Chemokine Secretion Rate:** [  $1e-7$  pg/s·cell /  $1e-5$  pg/s·cell /  $1e-3$  pg/s·cell ] -

The chemokine secretion rate defines how much chemokine each infected cell secretes per second.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

Chemokine secretion values differ between strains. It is important to note that aH5N1 and sH1N1 show a threshold at the same concentration: near  $1e-6$  pg/s·cell. This is an artifact of our artificially selected chemokine detection threshold, detailed in Section 4.2.1 and Figure 4.2. Because we initially picked a sensitivity threshold near the edge of the stable range of possible values, decreasing the total concentration of chemokine inadvertently crosses that arbitrary threshold and does not necessarily suggest an actual region of divergence.

**T Cell Circulation Time:** [  $1s$  /  $6s$  /  $1h$  ] - T cell circulation time defines how long a T cell takes to recirculate through the vascular system when it does not encounter the infection as it passes through the lung. This value should correspond to the vascular circulation time of a mouse as we used a mouse-sized model for computational reasons.

A value of 1 second was chosen as the minimum due to computational constraints, yet this value is likely already too small to be biologically plausible. We extended the range to up to 600 times the default value to examine what happens when the circulation time is so long that the T cells are effectively removed from the simulation

## Appendix A. CyCells Sensitivity Analysis

if they miss the focus of infection on their first pass through the lung.

T cell circulation times were tested over a very large range and does diverge for high values. Classifying this parameter as *bounded stable* may seem counterintuitive when examining the sensitivity figures. It is important to note that in this case, we tested extra values on the higher end of the baseline than other plots. While each strain does show divergence beyond a certain value, the intermediate values of 1 second and 20 seconds remain qualitatively similar to the main result, demonstrating a stable range surrounding the default value. Because vascular and lymph circulation times are unknown, parameter values of up to 1 minute and 3 minutes in mice may be plausible and warrant further study.

**T Cell Expected Kill Time:** [ 0m / 10m / 100m ] - The T cell expected kill time defines the rate at which T cells probabilistically induce apoptosis of virus-expressing epithelial cells in their immediate vicinity. A lower kill time corresponds to a higher kill rate.

We tested values only one order of magnitude higher than the default limiting the time to induce apoptosis of a virus-expressing cell to one hour. On the other side, we allowed T cells to induce apoptosis instantly to see if any delay in the induction of apoptosis was responsible for the inability for the immune response to clear the pandemic infection.

The parameter only diverges on the higher end in sH1N1. The extreme value corresponds to a T cell needing 100 minutes to induce the apoptosis of a single infected cell and is not biologically reasonable. The intermediate value corresponds to a time of 33.3 minutes and is also unlikely. Of interest, removing the parameter entirely by setting the kill time to instant was still not enough to allow the immune response to clear the pandemic infection.

## Appendix A. CyCells Sensitivity Analysis

**T Cell Speed:** [  $3.6 \mu\text{m}/\text{h}$  /  $6 \mu\text{m}/\text{m}$  /  $10 \mu\text{m}/\text{s}$  ] - T cell speed determines how fast T cells move over the epithelial monolayer.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

Divergence does occur in sH1N1 for the extreme values on both ends (but not for the intermediate ones). T cell movement in tissue has been observed [109]. Thus, we consider the extreme values biologically implausible.

**T Cell Age in Blood:** [  $1\text{h}$  /  $4\text{d}$  /  $400\text{d}$  ] - The expected age determines the decay rate of T cells in blood.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

T cell decay parameters allow the model to diverge only in the most extreme cases. Neither of these values are reasonable and the parameters show stable behavior otherwise. The default parameter appears to define a model as effective at clearing the infection as models with much larger values.

**T Cell Age at FOI:** [  $12\text{m}$  /  $2\text{h}$  /  $20\text{h}$  ] - The expected age determines the decay rate of T cells at the focus of infection (FOI).

We tested values up to one order of magnitude larger and smaller than the default parameter value.

T cell decay parameters allow the model to diverge only in the most extreme cases. Neither of these values are reasonable and the parameters show stable behav-



## Appendix A. CyCells Sensitivity Analysis

ior otherwise. While the default of 2 hours may seem short, it appears to generate similar behavior to models with larger values.

**T Cell Production Rate:** [ *125 cells/h* | *1257 cells/h* | *3750 cells/h* ] - The T cell production rate determines the rate at which T cells enter the blood stream starting at day 5 post-infection.

We tested values up to three times larger and ten times smaller than the default value. Values beyond a three-fold increase were not computationally tractable.

The T cell production rate does have a consistent response over its different values, but this effect is minimized above a certain rate. Thus, it is reasonable to assume that there is a threshold of T cell production beyond which the dynamics of the infection do not change. This is consistent with our observations of T cell clumping in areas of high chemokine concentration. Increasing the number of T cells in the system does not control the infection beyond a certain point because the T cells overlap in space and become redundant. The rapid decline of the infection size for the maximum value of  $3\mu m^2/s$  is simply an artifact of the infection exhausting the entire model space by day 3.

### A.1.2 Difference in Peak Only

Difference in peak only parameters affect the model's behavior up until the introduction of the T cell response, followed by a return to similar day 10 outcomes. Interestingly, these are the two parameters that define the transition delay between the different stages of epithelial cell infection. Because these terms refer to the timing of when infected cells release the virus particles, neither one directly affects the kinetics of the virus itself and thus the infection's behavior is still determined by the properties of the virus and the T cell response.

## Appendix A. CyCells Sensitivity Analysis

**Incubation Delay:** [ *5h* / *10h* / *20h* ] - Incubation delay describes how long an epithelial cell takes to transition to the virus-expressing state after it is initially infected. During this period, the cell, while infected, does not secrete virus.

We tested values up to a factor of two larger and smaller than the original parameter setting.

While a longer incubation time does not change the number of virions released by each infected cell, it does slow down the spread of the infection. Conversely, a shorter incubation time allows the infection to spread more quickly by allowing the infected cell to release the new virus in to the system earlier. Thus, changes in the incubation time do affect the infection profile before the introduction of T cells. The infection dynamics change dramatically upon the arrival of T cells. One of the main factors in determining whether or not the infection is cleared is the amount of new virus produced per infected cell. While this does depend somewhat on the overall size of the plaque, this value depends heavily on shutting down the expression of new virus, something T cells are unable to do during the cellular incubation phase as the virus-incubating cell has not expressed virus for T cells to detect. Thus, assuming the original plaque has not grown so large that it can no longer be effectively covered by the T cell response, the overall course of the infection will be unaffected by the length of the incubation period.

**Expression Delay:** [ *100m* / *1,000m* / *50h* ] - The expression delay is the amount of time a virus secreting cell will secrete virus in the absence of a T cell intervention. At the end of this time period, the expressing cell will transition to the dead state and remain inert for the remainder of the simulation.

The parameter was increased up by a factor of 3 and down by a factor of 10. In

## Appendix A. CyCells Sensitivity Analysis

each case, the change is extreme enough that the resulting values are not biologically likely.

Adjusting this parameter seems to affect the model behavior up until the arrival of the T cell response, after which the model converges back to the baseline behavior. This effect occurs because once T cells arrive, virus secreting cells no longer survive for their full lifespan. Rather, the delay becomes limited by the apoptosis time parameter.

### A.1.3 Sensitive Parameters

Sensitive parameters directly affect the result of the infection. There are five parameters classified as sensitive: viral response to IgM, infectivity, T cell production rate, viral decay, and viral diffusion. By comparing the three strains of influenza, we also know that the viral secretion rate affects the result. Of interest, a majority of these parameters are related to the behavior of the virus. The only sensitive parameter not related to the behavior or the virus is the T cell production rate.

**Viral Response to IgM:** [ 1 / 10 / 1,000 ] - Viral response to IgM simulates the presence of IgM at day 4 post-infection by increasing the decay rate of free virus particles by the given factor.

We examined a very large range of values, from the lowest possible value of 1 (effectively removing IgM), to an increased decay rate of three orders of magnitude.

While it is true that the extreme values have a large effect on the model behavior, it is more of interest that the intermediate values also seem to change the model behavior. This shows that there is no stable area around the default as is seen in the *bounded stable* parameters. The higher the strength of this parameter, the less virus there is in the system. Because the true effective strength of IgM is unknown in the

## Appendix A. CyCells Sensitivity Analysis

context of this model, it is not a target of our investigation.

**Viral Decay Rate:** [  $1e-7$  Hz |  $1e-5$  Hz |  $1e-3$  Hz ] - Viral decay determines how quickly free virus is removed from the system. A larger decay rate corresponds to faster removal.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

Viral decay has the exact same effect as the viral response to IgM parameter. Its value directly determines how much virus remains in the system over the course of the infection. There is a strong relationship between the decay rate and the resulting infection profile across all values tested.

**Infectivity:** [  $12$  m/virion |  $2$  h/virion |  $20$  h/virion ] - Infectivity describes the ability of the virus to infect healthy cells. A larger value corresponds to a virus particle needing less time to infect a nearby healthy epithelial cell.

We tested values up to one order of magnitude larger and smaller than the default parameter value.

The strength of the infectivity parameter is directly related to how much virus is present in the area of the healthy cells. Thus, viral density is linearly proportional to infectivity. Thus, it behaves similarly to the IgM and viral decay parameters as it directly determines the virulence of the virus.

**Viral Secretion Rate:** [  $5.4e-5$  PFU/s·cel |  $3.8e-4$  PFU/s·cel |  $5.1e-3$  PFU/s·cel ] - The viral secretion rate determines how quickly virus-secreting epithelial cells secrete new virus particles.

## Appendix A. CyCells Sensitivity Analysis

An explicit sensitivity analysis was not directly performed. The range of the parameters are taken from the respective secretion rates of the aH5N1, sH1N1, and pH1N1 influenzas (Table 3.8).

The viral secretion rate has a strong effect on the outcome of the infection. Similar to the previously discussed parameters, its value directly affects how much virus there is in the system. Thus, its effect is similar to the previous parameters.

**Viral Diffusion Rate:** [  $3e-4 \mu m^2/s$  |  $3e-2 \mu m^2/s$  |  $3 \mu m^2/s$  ] - The viral diffusion rate controls how quickly the virus spreads across the monolayer.

We tested values up to two orders of magnitude larger and smaller than the default parameter value.

The viral diffusion rate does not change how much virus there is in the system. Rather, it determines how fast the virus may spread across the alveoli. While its mechanism is different, its effect may be even stronger than the previous parameters. Increasing the viral diffusion rate so that the virus diffuses faster than the chemokine (an unlikely scenario) allows the virus to out-pace the body's ability to establish a chemical gradient around the infected cells. In this scenario T cells are constantly directed to locations behind the rapidly spreading viral cloud and would be unable to 'get ahead' of the infection.

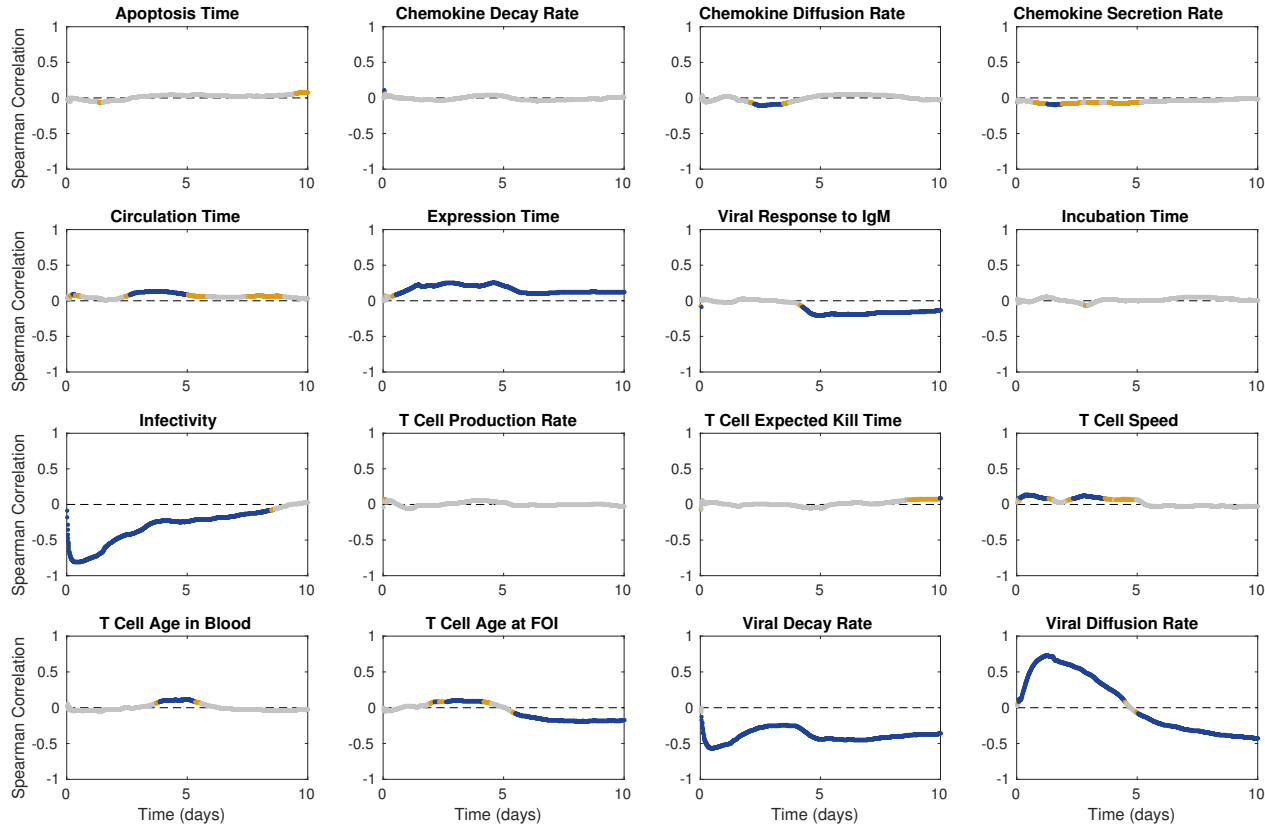


Figure A.4: **PRCC sensitivity analysis.** PRCC analysis generated Spearman correlation coefficients ( $\rho$ ) and significance values ( $p$ ) over 1441 time points. Plots show values of  $\rho$  vs. time for each parameter tested. Dark blue indicates significance levels at  $p < 0.01$ . Gold indicates significance of  $p < 0.05$ . Light gray indicates  $p \geq 0.05$ . Six parameters show significant effect over the time period where they were relevant to model dynamics: infected cell expression time, viral response to IgM, viral infectivity, T cell expected lifetime at the FOI, viral decay rate, and the viral diffusion rate.

## **A.2 Partial rank correlation coefficient sensitivity analysis**

We performed a partial rank correlation coefficient (PRCC) sensitivity analysis to complement our one factor at a time (OFAT) parameter analysis. Due to computational limitations, we computed PRCC values only for the sH1N1 strain, confirming that the results of the PRCC are consistent with the conclusions of the original sensitivity analysis and that viral parameters dominate the model output.

### **A.2.1 PRCC Design and Implementation**

PRCC analysis was performed as described in [38]. Samples were generated using Latin hypercube sampling (LHS) to generate 333 sample points for 17 parameters (those from Table 4.2, plus the sH1N1 viral secretion rate). Each parameter was assigned a confidence distribution (Table A.1) and sampled using its inverse cumulative distribution function to generate points with density proportional to the underlying probability distribution function. Each sample point was evaluated three times using unique random seeds for a total of 999 samples. One sample point resulted in undefined model behavior and was removed, leaving a total of 996 samples.

Output is defined as the number of infected cells remaining for each time point. Each model run generates 144,001 time points (10 day model with 6 second time steps). PRCC analysis was performed over every 100th time point, resulting in a time series of 1441 Spearman rank correlation coefficients ( $\rho$ ) and significance values ( $p$ ) for each parameter (Figs. A.4 - A.5).

## Appendix A. CyCells Sensitivity Analysis

Category	Parameter	Units	Distribution
Chemokine	Chemokine Decay Rate	$Hz$	LN(-3.4, 1.0)
	Chemokine Diffusion Rate	$\mu m^2/s$	LN(-0.5, 1.0)
	Chemokine Secretion Rate	$(pg/s \cdot cell)$	LN(-3.7, 0.1)
T Cell	Circulation Time	$seconds$	LN(-.75, 0.5)
	T Cell Kill Rate	$min$	LN(1.0, 0.5)
	T Cell Speed	$\mu m/min$	LN(-.75, 1),
	T Cell Age in Blood	$days$	LN(0.6, 1.0)
	T Cell Age at FOI	$min$	LN(2.1, 0.5)
	T Cell Production Rate	$cells/h$	LN(3.1, 0.15)
Delay	Apoptosis Time	$hours$	U(0.0, 2.0)
	Expression Time	$min$	LN(3.0, 0.15)
	Incubation Time	$hours$	LN(-1.0, 0.15)
Virus	Viral Response to IgM	—	LN(1.0, 0.5)
	Infectivity	$min/virion$	LN(2.1, 0.5)
	Viral Decay Rate	$day^{-1}$	LN(0, 1.0)
	Viral Diffusion Rate	$\mu m^2/s$	LN(-1.5, 1.0)
	Viral Secretion	$(PFU/s \cdot cell)$	LN(-3.4, 0.3)

Table A.1: **Latin hypercube sampling distributions.** LHS was used to generate sample points containing each of the 16 parameters from Table 4.2, plus the sH1N1 viral secretion rate. Parameters were sampled over biologically plausible ranges listed here for use in PRCC analysis. LN signifies a log-normal distribution. Values were sampled from a regular normal distribution with the listed parameters and then transformed to linear space from  $log_{10}$  space. U signifies a uniform distribution.

### A.2.2 PRCC Results

The analysis confirmed the conclusion of original sensitivity analysis that viral parameters dominate model output. Of the 17 parameters tested with PRCC analysis, 7 show significance of  $p < 0.01$  over the time period where the parameter was relevant to model dynamics (Figs. S6 - S7). All five of the viral parameters (viral response to IgM, viral infectivity, viral decay rate, viral diffusion rate, and viral secretion rate) are significant. Infected cell expression time and the T cell age at the focus of infection are also significant at  $p < 0.01$  (but have  $\rho$  values near 0). Viral infectivity, viral decay rate, viral diffusion rate, and viral secretion rate are the only parameters that



Appendix A. CyCells Sensitivity Analysis

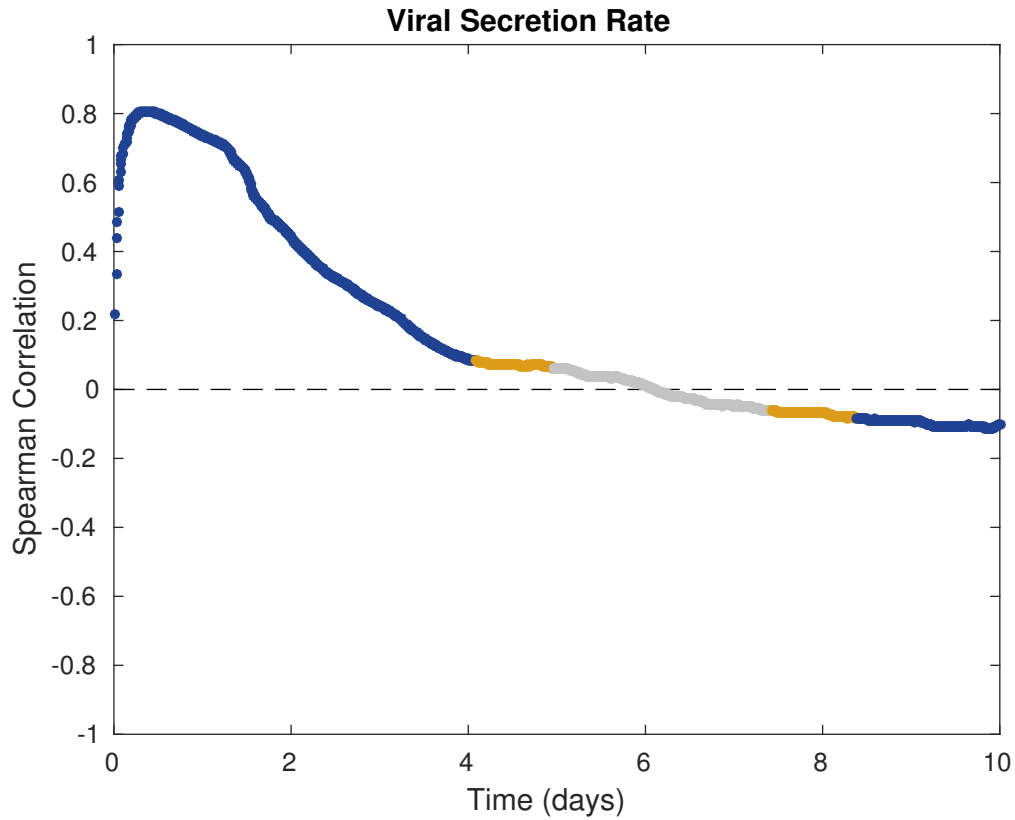


Figure A.5: **PRCC sensitivity analysis for the viral secretion rate.** Similar to Figure S6, PRCC analysis generated Spearman correlation coefficients ( $\rho$ ) and significance values ( $p$ ) over 1441 time points. Plot shows the values of  $\rho$  vs. time for the viral secretion rate. Dark blue indicates significance levels at  $p < 0.01$ . Gold indicates significance of  $p < 0.05$ . Light gray indicates  $p \geq 0.05$ . Viral secretion shows significant effect on model output. The decline of  $\rho$  over time reflects the early death of every modeled target cell (Figure A.6)).

have maximum absolute values of  $\rho$  above 0.5. Other parameters showed significant correlations (with low values of  $\rho$ ), but only over time periods where they were not active. Viral infectivity, viral diffusion rate, and viral secretion rate values of  $\rho$  cross or approach zero over time. This merely reflects that the modeled target cells have died (Fig. A.6) and so the decline in  $\rho$  is irrelevant.

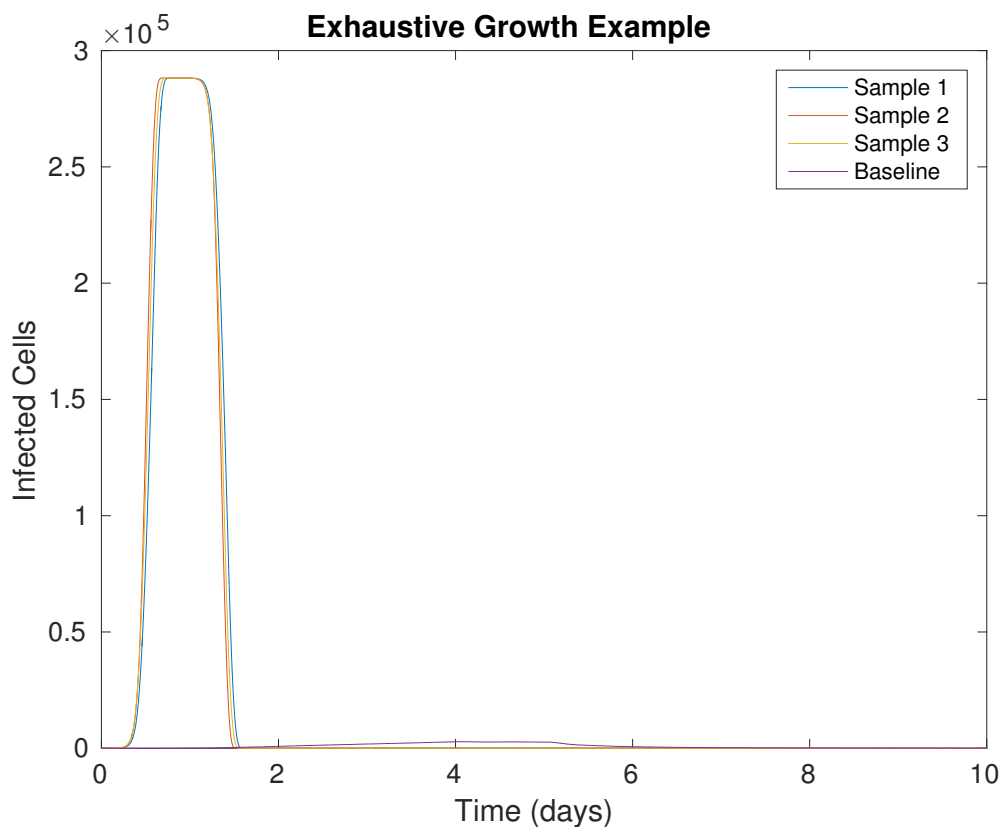


Figure A.6: **Exhaustive Growth:** Certain parameter sets generated by Latin hypercube sampling caused rapid uncontrolled infection growth (Samples 1-3). This resulted in the early death of every modeled target cell (saturation at day 1) and the inevitable decline of the infection size to zero. A larger, unbounded model environment would show unconstrained growth. This effect explains the decline of the  $\rho$  value of several viral parameters to or past zero over time in Figures A.4 and A.5.

# References

- [1] J. P. Hecker, M. E. Moses, Beyond pheromones: evolving error-tolerant, flexible, and scalable ant-inspired robot swarms, *Swarm Intelligence* 9 (1) (2015) 43–70.
- [2] C. Warrender, S. Forrest, F. Koster, Modeling intercellular interactions in early Mycobacterium infection., *Bulletin of Mathematical Biology* 68 (8) (2006) 2233–2261.
- [3] C. Beauchemin, Probing the effects of the well-mixed assumption on viral infection dynamics., *Journal of Theoretical Biology* 242 (2) (2006) 464–477.
- [4] G. V. Treyz, S. M. Treyz, *Automobile personal computer systems* (2004).
- [5] M. Weiser, *Hot topics-ubiquitous computing* (1993).
- [6] G. D. Abowd, E. D. Mynatt, Charting past, present, and future research in ubiquitous computing, *ACM Transactions on Computer-Human Interaction* 7 (1) (2000) 29–58.
- [7] G. Goggin, *Cell Phone Culture*, Routledge, 2006.
- [8] M.-A. Camarillo, Gonzalo and Garcia-Martin, *The 3G IP multimedia subsystem (IMS): merging the Internet and the cellular worlds*, Wiley, 2007.

## REFERENCES

- [9] Y. S. Y. Shin, K. C. K. Choi, T. Sakurai, Power optimization of real-time embedded systems on variable speed processors (2000).
- [10] W.-T. Shiue, C. Chakrabarti, Memory exploration for low power, embedded systems, in: Proceedings of the 36th annual ACM/IEEE Design Automation Conference, ACM, 1999, pp. 140–145.
- [11] J. L. J. Luo, N. K. Jha, Power-conscious joint scheduling of periodic task graphs and aperiodic tasks in distributed real-time embedded systems (2000).
- [12] B. M. Rogers, A. Krishna, G. B. Bell, K. Vu, X. Jiang, Y. Solihin, Scaling the bandwidth wall: challenges in and avenues for CMP scaling, *Computer Engineering* 37 (1) (2009) 371–382.
- [13] K. Argyraki, S. Ratnasamy, S. Baset, B.-G. Chun, K. Fall, G. Iannaccone, A. Knies, E. Kohler, M. Manesh, S. Nedeveschi, Can software routers scale?, Proceedings of the ACM workshop on Programmable routers for extensible services of tomorrow PRESTO 08 (2008) 21
- [14] T. Abdelzaher, J. Stankovic, Feedback control of data aggregation in sensor networks, in: 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601), Vol. 2, IEEE, 2004, pp. 1490–1495 Vol.2.
- [15] J.-Y. Chen, G. Pandurangan, D. Xu, Robust computation of aggregates in wireless sensor networks: distributed randomized algorithms and analysis, in: {IPSN} '05: Proceedings of the 4th international symposium on Information processing in sensor networks, 2005, p. 46.
- [16] A. Deshpande, S. Nath, P. B. Gibbons, S. Seshan, Cache-and-query for wide area sensor databases, Proceedings of the 2003 ACM SIGMOD international conference on on Management of data - SIGMOD '03 (2003) 503
- [17] S. Gill, Parallel programming, *The Computer Journal* 1 (1) (1958) 2–10.

## REFERENCES

- [18] D. Nassimi, S. Sahni, Bitonic sort on a mesh-connected parallel computer, *Computers, IEEE Transactions on* 100 (1) (1979) 2–7.
- [19] C. Cheng, R. Riley, S. P. R. Kumar, J. J. Garcia-Luna-Aceves, A loop-free extended Bellman-Ford routing protocol without bouncing effect, *ACM SIGCOMM Computer Communication Review* 19 (4) (1989) 224–236.
- [20] R. L. Cook, T. Porter, L. Carpenter, Distributed ray tracing, in: *ACM SIGGRAPH Computer Graphics*, Vol. 18, ACM, 1984, pp. 137–145.
- [21] G. M. Amdahl, Validity of the single processor approach to achieving large scale computing capabilities, in: *Proceedings of the April 18-20, 1967, spring joint computer conference*, ACM, 1967, pp. 483–485.
- [22] Y. Emek, J. Smula, R. Wattenhofer, Stone Age Distributed Computing, arXiv preprint arXiv:1202.1186.
- [23] M. Lanan, Spatiotemporal resource distribution and foraging strategies of ants (Hymenoptera: Formicidae), *Myrmecological News* 20 (2014) 53–70.
- [24] H. Mitchell, D. Levin, S. Forrest, C. a. a. Beauchemin, J. Tipper, J. Knight, N. Donart, R. C. Layton, J. Pyles, P. Gao, K. S. Harrod, A. S. Perelson, F. Koster, Higher level of replication efficiency of 2009 (H1N1) pandemic influenza virus than those of seasonal and avian strains: kinetics from epithelial cell culture and computational modeling., *Journal of virology* 85 (2) (2011) 1125–35.
- [25] T. P. Flanagan, K. Letendre, W. Burnside, G. M. Fricke, M. Moses, How ants turn information into food (2011).
- [26] M. A. Babyak, What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models, *Psychosomatic Medicine* 66 (3) (2004) 411–421.

## REFERENCES

- [27] W. G. Wilson, Resolving discrepancies between deterministic population models and individual-based simulations., *The American naturalist* 151 (2) (1998) 116–134.
- [28] A. J. Lotka, *Elements of Physical Biology*, Williams and Wilkins Company, 1925.
- [29] L. R. Brenan, Kathryn Eleda and Campbell, Stephen L and Petzold, Numerical solution of initial-value problems in differential-algebraic equations, 14th Edition, Society for Industrial and Applied Mathematics, 1987.
- [30] J. R. Dormand, P. J. Prince, A family of embedded Runge-Kutta formulae, *Journal of Computational and Applied Mathematics* 6 (1) (1980) 19–26.
- [31] H. J. Motulsky, L. A. Ransnas, Fitting curves to data using nonlinear regression: a practical and nonmathematical review., *The FASEB journal official publication of the Federation of American Societies for Experimental Biology* 1 (5) (1987) 365–374.
- [32] D. E. Goldberg, J. H. Holland, Genetic Algorithms and Machine Learning, *Machine Learning* 3 (2) (1988) 95–99.
- [33] S. Kirkpatrick, D. G. Jr., M. P. Vecchi, Optimization by simulated annealing, *Science* 220 (4598) (1983) 671–680.
- [34] J. J. Moré, The Levenberg-Marquardt algorithm: implementation and theory, in: *Numerical analysis*, Springer, 1978, pp. 105–116.
- [35] J. C. Helton, J. D. Johnson, C. J. Sallaberry, C. B. Storlie, Survey of sampling-based methods for uncertainty and sensitivity analysis, *Reliability Engineering and System Safety* 91 (10-11) (2006) 1175–1209.
- [36] C. F. J. Wu, Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis, *Annals of Statistics* 14 (4) (1986) 1261–1295.

## REFERENCES

- [37] C. Daniel, One-at-a-Time Plans, *Journal of American Statistical Association* 68 (342) (1973) 353–360.
- [38] S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, A methodology for performing global uncertainty and sensitivity analysis in systems biology (2008).
- [39] M. D. McKay, Latin hypercube sampling as a tool in uncertainty analysis of computer models, in: *Proceedings of the 24th conference on Winter simulation - WSC '92*, ACM Press, New York, New York, USA, 1992, pp. 557–564.
- [40] G. I. Bell, Mathematical model of clonal selection and antibody production, *Journal of Theoretical Biology* 29 (1970) 191–232.
- [41] G. A. Bocharov, A. A. Romanyukha, Mathematical model of antiviral immune response III. Influenza A virus infection (1994).
- [42] A. S. Perelson, Modelling viral and immune system dynamics, *Nature Reviews Immunology* 2 (2002) 28–36.
- [43] D. S. Callaway, A. S. Perelson, HIV-1 infection and low steady state viral loads., *Bulletin of mathematical biology* 64 (2002) 29–64.
- [44] P. Baccam, C. Beauchemin, C. a. Macken, F. G. Hayden, A. S. Perelson, Kinetics of influenza A virus infection in humans., *Journal of virology* 80 (2006) 7590–9.
- [45] C. A. A. Beauchemin, J. J. McSharry, G. L. Drusano, J. T. Nguyen, G. T. Went, R. M. Ribeiro, A. S. Perelson, Modeling amantadine treatment of influenza A virus in vitro., *Journal of Theoretical Biology* 254 (2) (2008) 439–451.
- [46] A. Handel, R. Antia, A Simple Mathematical Model Helps To Explain the Immunodominance of CD8 T Cells in Influenza A Virus Infections, *Journal of Virology* 82 (16) (2008) 7768–7772.

## REFERENCES

- [47] H. Miao, J. A. Hollenbaugh, M. S. Zand, J. Holden-Wiltse, T. R. Mosmann, A. S. Perelson, H. Wu, D. J. Topham, Quantifying the early immune response and adaptive immune response kinetics in mice infected with influenza A virus., *Journal of Virology* 84 (13) (2010) 6687–6698.
- [48] R. A. Saenz, M. Quinlivan, D. Elton, S. MacRae, A. S. Blunden, J. A. Mumford, J. M. Daly, P. Digard, A. Cullinane, B. T. Grenfell, J. W. McCauley, J. L. N. Wood, J. R. Gog, Dynamics of influenza virus infection and pathology., *Journal of Virology* 84 (8) (2010) 3974–3983.
- [49] I. Price, E. D. Mochan-Keef, D. Swigon, G. B. Ermentrout, S. Lukens, F. R. Toapanta, T. M. Ross, G. Clermont, The inflammatory response to influenza A virus (H1N1): An experimental and mathematical study., *Journal of theoretical biology* 374 (2015) 83–93.
- [50] H. Y. Lee, D. J. Topham, S. Y. Park, J. Hollenbaugh, J. Treanor, T. R. Mosmann, X. Jin, B. M. Ward, H. Miao, J. Holden-Wiltse, A. S. Perelson, M. Zand, H. Wu, Simulation and Prediction of the Adaptive Immune Response to Influenza A Virus Infection, *Journal of Virology* 83 (14) (2009) 7151–7165.
- [51] C. Beauchemin, J. Samuel, J. Tuszynski, A simple cellular automaton model for influenza A viral infections, *Journal of Theoretical Biology* 232 (2) (2005) 223–234. arXiv:0402012.
- [52] J. B. Beltman, A. F. M. Marée, R. J. De Boer, Spatial modelling of brief and long interactions between T cells and dendritic cells., *Immunology and Cell Biology* 85 (4) (2007) 306–314.
- [53] H. Zheng, B. Jin, S. E. Henrickson, A. S. Perelson, U. H. Von Andrian, A. K. Chakraborty, How antigen quantity and quality determine T-cell decisions in lymphoid tissue., *Molecular and Cellular Biology* 28 (12) (2008) 4040–4051.



## REFERENCES

- [54] R. M. A. Vroomans, A. F. M. Marée, R. J. De Boer, J. B. Beltman, Chemotactic migration of T cells towards dendritic cells promotes the detection of rare antigens, *PLoS Comput Biol* 8 (11) (2012) e1002763.
- [55] J. Textor, S. E. Henrickson, J. N. Mandl, U. H. von Andrian, J. Westermann, R. J. de Boer, J. B. Beltman, Random Migration and Signal Integration Promote Rapid and Robust T Cell Recruitment., *PLoS computational biology* 10 (8) (2014) e1003752.
- [56] B. Hölldobler, E. O. Wilson, *The Ants*, Vol. N1, Belknap Press, 1990.
- [57] K. Mody, K. E. Linsenmair, Finding its place in a competitive ant community: Leaf fidelity of *Camponotus sericeus*, *Insectes Sociaux* 50 (2003) 191–198.
- [58] B. Prabhakar, K. N. Dektar, D. M. Gordon, The Regulation of Ant Colony Foraging Activity without Spatial Information, *PLoS Computational Biology* 8.
- [59] G. Theraulaz, E. Bonabeau, A brief history of stigmergy., *Artificial Life* 5 (2) (1999) 97–116.
- [60] S. E. F. Evison, O. L. Petchey, A. P. Beckerman, F. L. W. Ratnieks, Combined use of pheromone trails and visual landmarks by the common garden ant *Lasius niger*, *Behavioral Ecology and Sociobiology* 63 (2008) 261–267.
- [61] R. Beckers, J. L. Deneubourg, S. Goss, J. M. Pasteels, Collective decision making through food recruitment, *Insectes Sociaux* 37 (1990) 258–267.
- [62] E. J. H. Robinson, F. L. W. Ratnieks, M. Holcombe, An agent-based model to investigate the roles of attractive and repellent pheromones in ant decision making during foraging, *Journal of Theoretical Biology* 255 (2008) 250–258.

## REFERENCES

- [63] T. J. Czaczkes, How to not get stuck-Negative feedback due to crowding maintains flexibility in ant foraging, *Journal of Theoretical Biology* 360 (2014) 172–180.
- [64] A. Dornhaus, Finding optimal collective strategies using individual-based simulations: colony organization in social insects, *Mathematical and Computer Modelling of Dynamical Systems* 18 (1) (2012) 25–37.
- [65] N. J. Sanders, D. M. Gordon, Resources and the flexible allocation of work in the desert ant, *Aphaenogaster cockerelli* (2002).
- [66] K. Letendre, M. E. Moses, Synergy in Ant Foraging Strategies: Memory and Communication Alone and in Combination, in: *Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation, GECCO '13*, ACM, New York, NY, USA, 2013, pp. 41–48.
- [67] N. R. Franks, T. Richardson, Teaching in tandem-running ants., *Nature* 439 (2006) 153.
- [68] B. Markó, W. Czechowski, Space use, foraging success and competitive relationships in *Formica cinerea* (Hymenoptera Formicidae) on sand dunes in southern Finland, *Ethology Ecology & Evolution* 24 (2012) 149–164.
- [69] M. W. Moffett, Foraging dynamics in the group-hunting myrmicine ant, *Pheidologeton diversus*, *Journal of Insect Behavior* 1 (3) (1988) 309–331.
- [70] S. C. Pratt, Efficiency and regulation of recruitment during colony emigration by the ant *Temnothorax curvispinosus*, *Behavioral Ecology and Sociobiology* 62 (2008) 1369–1376.
- [71] N. Pinter-Wollman, D. M. Gordon, S. Holmes, Nest site and weather affect the personality of harvester ant colonies, *Behavioral Ecology* 23 (5) (2012) 1022–1029.

## REFERENCES

- [72] A. Schmolke, Benefits of dispersed central-place foraging: an individual-based model of a polydomous ant colony., *The American naturalist* 173 (2009) 772–778.
- [73] A. Dornhaus, F. Klügl, C. Oechslein, F. Puppe, L. Chittka, Benefits of recruitment in honey bees: Effects of ecology and colony size in an individual-based model, *Behavioral Ecology* 17 (2006) 336–344.
- [74] Z. Cook, D. W. Franks, E. J. H. Robinson, Exploration versus exploitation in polydomous ant colonies, *Journal of Theoretical Biology* 323 (2013) 49–56.
- [75] F.-X. Dechaume-Moncharmont, A. Dornhaus, A. I. Houston, J. M. McNamara, E. J. Collins, N. R. Franks, The hidden cost of information in collective foraging., *Proceedings. Biological sciences / The Royal Society* 272 (2005) 1689–95.
- [76] M. Dorigo, M. Birattari, T. Stützle, Ant colony optimization, *IEEE Computational Intelligence Magazine* 1 (2006) 28–39.
- [77] G. Di Caro, M. Dorigo, AntNet: Distributed Stigmergetic Control for Communications Networks, *Journal of Artificial Intelligence Research* 9 (317–365) (1998) 317–365.
- [78] M. Dorigo, L. M. Gambardella, Ant colonies for the travelling salesman problem, *Biosystems* 43 (2) (1997) 73–81.
- [79] D. Merkle, M. Middendorf, H. Schmeck, Ant colony optimization for resource-constrained project scheduling (2002).
- [80] S. Fidanova, ACO algorithm for MKP using various heuristic information, in: *Numerical Methods and Applications*, Springer, 2003, pp. 438–444.
- [81] J. Svennebring, S. Koenig, Building terrain-covering ant robots: A feasibility study, *Autonomous Robots* 16 (2004) 313–332.

## REFERENCES

- [82] S. Garnier, F. T??che, M. Combe, A. Grimal, G. Theraulaz, Alice in pheromone land: An experimental setup for the study of ant-like robots, in: Proceedings of the 2007 IEEE Swarm Intelligence Symposium, SIS 2007, 2007, pp. 37–44.
- [83] R. Mayet, J. Roberz, T. Schmickl, K. Crailsheim, Antbots: A feasible visual emulation of pheromone trails for swarm robots, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 6234 LNCS, 2010, pp. 84–94.
- [84] D. Cassill, Rules of supply and demand regulate recruitment to food in an ant society, *Behavioral Ecology and Sociobiology* 54 (5) (2003) 441–450.
- [85] M. J. Krieger, J. B. Billeter, L. Keller, Ant-like task allocation and recruitment in cooperative robots., *Nature* 406 (6799) (2000) 992–5.
- [86] T. Schmickl, K. Crailsheim, Trophallaxis within a robotic swarm: Bio-inspired communication among robots in a swarm, *Autonomous Robots* 25 (1-2) (2008) 171–188.
- [87] S. Nouyan, R. Grob, M. Bonani, F. Mondada, M. Dorigo, Teamwork in self-organized robot colonies, *IEEE Transactions on Evolutionary Computation* 13 (4) (2009) 695–711.
- [88] D. Levin, S. Forrest, S. Banerjee, C. Clay, J. Cannon, M. E. Moses, F. T. Koster, A spatial model of the efficiency of T cell search in the influenza-infected lung, *Journal of Theoretical Biology* In Press.
- [89] A. M. Smith, F. R. Adler, A. S. Perelson, An accurate two-phase approximate solution to an acute viral infection model., *Journal of Mathematical Biology* 60 (5) (2010) 711–726.
- [90] C. Beauchemin, S. Forrest, F. T. Koster, Modeling Influenza Viral Dynamics in Tissue, *LNCS Artificial Immune Systems* 4163 (2006) 23–36.

## REFERENCES

- [91] R. H. Peters, *The ecological implications of body size*, Cambridge University Press, 1983.
- [92] B. J. Coburn, B. G. Wagner, S. Blower, Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1), *BMC Medicine* 7 (1) (2009) 30.
- [93] S. J. Patel, K. R. King, M. Casali, M. L. Yarmush, DNA-triggered innate immune responses are propagated by gap junction communication., *Proceedings of the National Academy of Sciences of the United States of America* 106 (31) (2009) 12867–72.
- [94] M. D. Gunn, K. Tangemann, C. Tam, J. G. Cyster, S. D. Rosen, L. T. Williams, A chemokine expressed in lymphoid high endothelial venules promotes the adhesion and chemotaxis of naive T lymphocytes., *Proceedings of the National Academy of Sciences of the United States of America* 95 (1) (1998) 258–263.
- [95] B. D. Medoff, S. Y. Thomas, A. Banerji, J. C. Wain, H. Zhang, C. M. Lilly, L. C. Ginns, A. D. Luster, Pathogenic T-cell recruitment into the airway in human disease., *Annals Of The New York Academy Of Sciences* 1062 (2005) 220–241.
- [96] T. Okada, M. J. Miller, I. Parker, M. F. Krummel, M. Neighbors, S. B. Hartley, A. O’Garra, M. D. Cahalan, J. G. Cyster, Antigen-Engaged B Cells Undergo Chemotaxis toward the T Zone and Form Motile Conjugates with Helper T Cells, *PLoS Biology* 3 (6) (2005) e150.
- [97] F. Castellino, A. Y. Huang, G. Altan-Bonnet, S. Stoll, C. Scheinecker, R. N. Germain, Chemokines enhance immunity by guiding naive CD8+ T cells to sites of CD4+ T cell-dendritic cell interaction., *Nature* 440 (7086) (2006) 890–5.

## REFERENCES

- [98] S. K. Bromley, T. R. Mempel, A. D. Luster, Orchestrating the orchestrators: chemokines in control of T cell traffic., *Nature immunology* 9 (9) (2008) 970–80.
- [99] M. Q. Zhao, M. H. Stoler, A. N. Liu, B. Wei, C. Soguro, Y. S. Hahn, R. I. Enelow, Alveolar epithelial cell chemokine expression triggered by antigen-specific cytolytic CD8(+) T cell recognition., *Journal of Clinical Investigation* 106 (6) (2000) R49–R58.
- [100] M. C. W. Chan, C. Y. Cheung, W. H. Chui, S. W. Tsao, J. M. Nicholls, Y. O. Chan, R. W. Y. Chan, H. T. Long, L. L. M. Poon, Y. Guan, J. S. M. Peiris, Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells, *Respiratory Research* 6 (1) (2005) 135.
- [101] M. R. Looney, J. Bhattacharya, Live imaging of the lung., *Annual review of physiology* 76 (2014) 431–45.
- [102] M. B. Headley, A. Bins, A. Nip, E. W. Roberts, M. R. Looney, A. Gerard, M. F. Krummel, Visualization of immediate immune responses to pioneer metastatic cells in the lung, *Nature* 531 (7595) (2016) 513–517.
- [103] M. J. Miller, S. H. Wei, M. D. Cahalan, I. Parker, Autonomous T cell trafficking examined in vivo with intravital two-photon microscopy, *Proceedings of the National Academy of Sciences of the United States of America* 100 (5) (2003) 2604–2609.
- [104] R. S. Allan, J. Waithman, S. Bedoui, C. M. Jones, J. A. Villadangos, Y. Zhan, A. M. Lew, K. Shortman, W. R. Heath, F. R. Carbone, Migratory dendritic cells transfer antigen to a lymph node-resident dendritic cell population for efficient CTL priming., *Immunity* 25 (1) (2006) 153–162.

## REFERENCES

- [105] E. Ingulli, C. Funatake, E. L. Jacovetty, M. Zanetti, Cutting edge: antigen presentation to CD8 T cells after influenza A virus infection., *The Journal of Immunology* 182 (1) (2009) 29–33.
- [106] A. Bachem, W. Hochstättler, M. Malich, The simulated trading heuristic for solving vehicle routing problems, *Discrete Applied Mathematics* 65 (1) (1996) 47–72.
- [107] K. J. Elbert, U. F. Schäfer, H. J. Schäfers, K. J. Kim, V. H. Lee, C. M. Lehr, Monolayers of human alveolar epithelial cells in primary culture for pulmonary absorption and transport studies., *Pharmaceutical Research* 16 (5) (1999) 601–608.
- [108] A. K. Abbas, A. H. H. Lichtman, S. Pillai, *Cellular and Molecular Immunology*, Saunders, 2011.
- [109] J. G. Egen, A. G. Rothfuchs, C. G. Feng, M. A. Horwitz, A. Sher, R. N. Germain, Intravital imaging reveals limited antigen presentation and T cell effector function in mycobacterial granulomas., *Immunity* 34 (5) (2011) 807–819.
- [110] S. Banerjee, M. Moses, Scale invariance of immune system response rates and times: perspectives on immune system architecture and implications for artificial immune systems, *Swarm Intelligence* 4 (4) (2010) 301–318.
- [111] S. Nandagopal, D. Wu, F. Lin, Combinatorial Guidance by CCR7 Ligands for T Lymphocytes Migration in Co-Existing Chemokine Fields, *PLoS ONE* 6 (3) (2011) 11.
- [112] S. Banerjee, D. Levin, M. Moses, F. Koster, S. Forrest, The Value of Inflammatory Signals in Adaptive Immune Responses, *ICARIS* 6825 (2011) 1–14.

## REFERENCES

- [113] M. S. Diamond, E. M. Sitati, L. D. Friend, S. Higgs, B. Shrestha, M. Engle, A critical role for induced IgM in the protection against West Nile virus infection., *The Journal of Experimental Medicine* 198 (12) (2003) 1853–1862.
- [114] V. V. Ganusov, R. J. De Boer, Estimating in vivo death rates of targets due to CD8 T-cell-mediated killing., *Journal of Virology* 82 (23) (2008) 11749–11757.
- [115] C. Desmetz, Y.-L. Lin, C. Mettling, P. Portalès, H. Rabesandratana, J. Clot, P. Corbeau, The strength of the chemotactic response to a CCR5 binding chemokine is determined by the level of cell surface CCR5 density., *Immunology* 119 (4) (2006) 551–61.
- [116] P. Gao, X.-y. Zhou, Y. Yashiro-ohtani, Y.-f. Yang, N. Sugimoto, S. Ono, T. Nagasawa, H. Fujiwara, T. Hamaoka, The unique target specificity of a nonpeptide chemokine receptor antagonist : selective blockade of two Th1 chemokine receptors CCR5 and CXCR3 cated in their migration to sites of inflammation ., *Journal of Leukocyte Biology* 73 (February) (2003) 273–280.
- [117] H. P. Mirsky, M. J. Miller, J. J. Linderman, D. E. Kirschner, Systems biology approaches for understanding cellular mechanisms of immunity in lymph nodes during infection (2011).
- [118] J. E. Christensen, A. Nansen, T. Moos, B. Lu, C. Gerard, J. P. Christensen, A. R. Thomsen, Efficient T-cell surveillance of the CNS requires expression of the CXC chemokine receptor 3., *Journal of Neuroscience* 24 (20) (2004) 4849–4858.
- [119] J. E. Christensen, C. De Lemos, T. Moos, J. P. Christensen, A. R. Thomsen, CXCL10 is the key ligand for CXCR3 on CD8+ effector T cells involved in immune surveillance of the lymphocytic choriomeningitis virus-infected central nervous system., *The Journal of Immunology* 176 (7) (2006) 4235–4243.



## REFERENCES

- [120] T. C. Dawson, M. a. Beck, W. a. Kuziel, F. Henderson, N. Maeda, Contrasting effects of CCR5 and CCR2 deficiency in the pulmonary inflammatory response to influenza A virus., *The American journal of pathology* 156 (6) (2000) 1951–9.
- [121] C. De Lemos, J. E. Christensen, A. Nansen, T. Moos, B. Lu, C. Gerard, J. P. Christensen, A. R. Thomsen, Opposing effects of CXCR3 and CCR5 deficiency on CD8+ T cell-mediated inflammation in the central nervous system of virus-infected mice., *The Journal of Immunology* 175 (3) (2005) 1767–1775.
- [122] S. Fadel, S. Bromley, B. Medoff, A. Luster, CXCR3-deficiency protects influenza-infected CCR5-deficient mice from mortality, *European journal of immunology* 38 (12) (2008) 3376–3387.
- [123] S. Gadhamsetty, A. F. M. Marée, J. B. Beltman, R. J. De Boer, A general functional response of cytotoxic T lymphocyte-mediated killing of target cells, *Biophysical Journal* 106 (8) (2014) 1780–1791.
- [124] J. R. Groom, A. D. Luster, CXCR3 ligands: redundant, collaborative and antagonistic functions., *Immunology and Cell Biology* 89 (2) (2011) 207–215.
- [125] J. R. Groom, A. D. Luster, CXCR3 in T cell function., *Experimental Cell Research* 317 (5) (2011) 620–631.
- [126] M.-F. Hsieh, S.-L. Lai, J.-P. Chen, J.-M. Sung, Y.-L. Lin, B. A. Wu-Hsieh, C. Gerard, A. Luster, F. Liao, Both CXCR3 and CXCL10/IFN-inducible protein 10 are required for resistance to primary infection by dengue virus., *The Journal of Immunology* 177 (3) (2006) 1855–1863.
- [127] R. S. Klein, E. Lin, B. Zhang, A. D. Luster, J. Tollett, M. A. Samuel, M. Engle, M. S. Diamond, Neuronal CXCL10 directs CD8+ T-cell recruitment and

## REFERENCES

- control of West Nile virus encephalitis., *Journal of Virology* 79 (17) (2005) 11457–11466.
- [128] J. E. Kohlmeier, T. Cookenham, S. C. Miller, A. D. Roberts, J. P. Christensen, A. R. Thomsen, D. L. Woodland, CXCR3 directs antigen-specific effector CD4<sup>+</sup> T cell migration to the lung during parainfluenza virus infection., *The Journal of Immunology* 183 (7) (2009) 4378–4384.
- [129] J. E. Kohlmeier, W. W. Reiley, G. Perona-Wright, M. L. Freeman, E. J. Yager, L. M. Connor, E. L. Brincks, T. Cookenham, A. D. Roberts, C. E. Burkum, S. Sell, G. M. Winslow, M. A. Blackman, M. Mohrs, D. L. Woodland, Inflammatory chemokine receptors regulate CD8(+) T cell contraction and memory generation following infection., *The Journal of Experimental Medicine* 208 (8) (2011) 1621–1634.
- [130] M. Thapa, R. S. Welner, R. Pelayo, D. J. J. Carr, CXCL9 and CXCL10 expression are critical for control of genital herpes simplex virus type 2 infection through mobilization of HSV-specific CTL and NK cells to the nervous system., *The Journal of Immunology* 180 (2) (2008) 1098–1106.
- [131] M. D. Wareing, A. B. Lyon, B. Lu, C. Gerard, S. R. Sarawar, Chemokine expression during the development and resolution of a pulmonary leukocyte response to influenza A virus infection in mice., *Journal of Leukocyte Biology* 76 (4) (2004) 886–895.
- [132] T. R. Wuest, D. J. J. Carr, Dysregulation of CXCR3 signaling due to CXCL10 deficiency impairs the antiviral response to herpes simplex virus 1 infection., *The Journal of Immunology* 181 (11) (2008) 7985–7993.
- [133] K. A. Pawelek, G. T. Huynh, M. Quinlivan, A. Cullinane, L. Rong, A. S. Perelson, Modeling within-host dynamics of influenza virus infection including immune responses, *PLoS Computational Biology* 8 (6).

## REFERENCES

- [134] R. W. Y. Chan, K. M. Yuen, W. C. L. Yu, C. C. C. Ho, J. M. Nicholls, J. S. M. Peiris, M. C. W. Chan, Influenza H5N1 and H1N1 Virus Replication and Innate Immune Responses in Bronchial Epithelial Cells Are Influenced by the State of Differentiation, *PLoS ONE* 5 (1) (2010) 9.
- [135] H. Zeng, C. Pappas, J. M. Katz, T. M. Tumpey, The 2009 pandemic H1N1 and triple-reassortant swine H1N1 influenza viruses replicate efficiently but elicit an attenuated inflammatory response in polarized human bronchial epithelial cells., *Journal of Virology* 85 (2) (2011) 686–696.
- [136] M. D. De Jong, C. P. Simmons, T. T. Thanh, V. M. Hien, G. J. D. Smith, T. N. B. Chau, D. M. Hoang, N. V. V. Chau, T. H. Khanh, V. C. Dong, P. T. Qui, B. V. Cam, D. Q. Ha, Y. Guan, J. S. M. Peiris, N. T. Chinh, T. T. Hien, J. Farrar, Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia., *Nature Medicine* 12 (10) (2006) 1203–1207.
- [137] S. Matsukura, F. Kokubu, H. Noda, H. Tokunaga, M. Adachi, Expression of IL-6, IL-8, and RANTES on human bronchial epithelial cells, NCI-H292, induced by influenza virus A., *The Journal of allergy and clinical immunology* 98 (6 Pt 1) (1996) 1080–1087.
- [138] U. Arndt, G. Wennemuth, P. Barth, M. Nain, Y. Al-Abed, A. Meinhardt, D. Gemsa, M. Bacher, Release of Macrophage Migration Inhibitory Factor and CXCL8/Interleukin-8 from Lung Epithelial Cells Rendered Necrotic by Influenza A Virus Infection, *Journal of Virology* 76 (18) (2002) 9298–9306.
- [139] I. Julkunen, K. Melén, M. Nyqvist, J. Pirhonen, T. Sareneva, S. Matikainen, Inflammatory responses in influenza A virus infection, *Vaccine* 19 Suppl 1 (Supplement 1) (2000) S32–7.
- [140] K. Lim, Y.-M. Hyun, K. Lambert-Emo, T. Capece, S. Bae, R. Miller, D. J.

## REFERENCES

- Topham, M. Kim, Neutrophil trails guide influenza-specific CD8+ T cells in the airways, *Science* 349 (6252) (2015) 4352.
- [141] N. Li Jeon, H. Baskaran, S. K. W. Dertinger, G. M. Whitesides, L. Van De Water, M. Toner, Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in a microfabricated device., *Nature Biotechnology* 20 (8) (2002) 826–830.
- [142] B. McDonald, K. Pittman, G. B. Menezes, S. A. Hirota, I. Slaba, C. C. M. Waterhouse, P. L. Beck, D. A. Muruve, P. Kubers, Intravascular danger signals guide neutrophils to sites of sterile inflammation., *Science* 330 (6002) (2010) 362–366.
- [143] A. Kacelnik, M. Bateson, Risky Theories – The Effects of Variance on Foraging Decisions, *Integrative and Comparative Biology* 36 (1996) 402–434.
- [144] D. Levin, J. P. Hecker, M. E. Moses, S. Forrest, Volatility and spatial distribution of resources determine ant foraging strategies, in: *European Conference on Artificial Life*, Vol. 13, The MIT Press, 2015, pp. 256–263.
- [145] T. P. Flanagan, K. Letendre, W. R. Burnside, G. M. Fricke, M. E. Moses, Quantifying the effect of colony size and food distribution on harvester ant foraging, *PLoS ONE* 7.
- [146] K. Jaffe, P. Howse, The mass recruitment system of the leaf cutting ant, *Atta cephalotes* (L.) (1979).
- [147] S. Goss, J. L. Deneubourg, The self-organising clock pattern of *Messor pergandei* (Formicidae, Myrmicinae), *Insectes Sociaux* 36 (1989) 339–347.
- [148] N. J. R. Plowes, K. Ramsch, M. Middendorf, B. Hölldobler, An empirically based simulation of group foraging in the harvesting ant, *Messor pergandei*, *Journal of Theoretical Biology* 340 (2014) 186–198.

## REFERENCES

- [149] R. A. Bernstein, Foraging strategies of ants in response to variable food density, *Ecology* 56 (1975) 213–219.
- [150] T. J. Czaczkes, F. L. W. Ratnieks, Pheromone trails in the Brazilian ant *Pheidole oxyops*: Extreme properties and dual recruitment action, *Behavioral Ecology and Sociobiology* 66 (2012) 1149–1156.
- [151] R. Droual, Anti-predator behaviour in the ant *Pheidole desertorum*: The importance of multiple nests (1984).
- [152] J. T. Mirenda, H. Topoff, Nomadic behavior of army ants in a desert-grassland habitat, *Behavioral Ecology and Sociobiology* 7 (1980) 129–135.
- [153] D. Holway, T. Case, Mechanisms of dispersed central-place foraging in polydomous colonies of the Argentine ant., *Animal behaviour* 59 (2000) 433–441.
- [154] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, H. E. Stanley, Lévy flight search patterns of wandering albatrosses, *Nature* 381 (1996) 413–415.
- [155] G. E. Uhlenbeck, L. S. Ornstein, On the theory of the Brownian motion, *Physical Review* 36 (1930) 823–841.
- [156] F. B. Knight, *Essentials of Brownian motion and diffusion*, American Mathematical Soc., 1981.
- [157] S. Redner, *A guide to first-passage processes*, Cambridge University Press, 2001.
- [158] M. R. Evans, S. N. Majumdar, Diffusion with stochastic resetting, *Physical Review Letters* 106 (2011) 1–4. arXiv:1102.2704,

## REFERENCES

- [159] M. R. Evans, S. N. Majumdar, Diffusion with resetting in arbitrary spatial dimension, *Journal of Physics A: Mathematical and Theoretical* 47 (28) (2014) 285001.
- [160] G. Ramos-Fernández, J. L. Mateos, O. Miramontes, G. Cocho, H. Larralde, B. Ayala-Orozco, Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*), *Behavioral Ecology and Sociobiology* 55 (2004) 223–230. arXiv:0301019,
- [161] J. M. Kleinberg, Navigation in a small world, *Nature* 406 (2000) 845.
- [162] G. M. Fricke, F. Asperti-Boursin, J. Hecker, J. Cannon, M. Moses, From microbiology to microcontrollers: Robot search patterns inspired by T cell movement, in: *Advances in Artificial Life, ECAL*, Vol. 12, 2013, pp. 1009–1016.
- [163] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, S. Havlin, M. G. E. Da Luz, E. P. Raposo, H. E. Stanley, Levy flights in random searches, *Physica A* 282 (2000) 1–12.
- [164] R. P. Brent, Algorithms for minimization without derivatives, Courier Corporation, 1973. arXiv:arxiv:1106.5394v1,
- [165] M. Dorigo, V. Maniezzo, A. Coloni, Ant system: Optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 26 (1996) 29–41.
- [166] G. Di Caro, F. Ducatelle, L. M. Gambardella, AntHocNet: an adaptive nature-inspired algorithm for routing in mobile ad hoc networks, *European Transactions on Telecommunications* 16 (5) (2005) 443–455.
- [167] V. Genovese, P. Dario, R. Magni, L. Odetti, Self Organizing Behavior And Swarm Intelligence In A Pack Of Mobile Miniature Robots In Search Of Pollu-

## REFERENCES

- tants, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Vol. 3, IEEE, 1992, pp. 1575–1582.
- [168] S. Nawaz, M. Hussain, S. Watson, N. Trigoni, P. N. Green, An underwater robotic network for monitoring nuclear waste storage pools, in: Sensor Systems and Software, Springer, 2009, pp. 236–255.
- [169] R. W. Deming, L. I. Perlovsky, Concurrent multi-target localization, data association, and navigation for a swarm of flying sensors, *Information Fusion* 8 (3) (2007) 316–330.
- [170] A. Dhariwal, G. Sukhatme, A. Requicha, Bacterium-inspired robots for environmental monitoring, in: IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004, Vol. 2, IEEE, 2004, pp. 1436–1443 Vol.2.
- [171] S. Berman, A. Halasz, M. Hsieh, V. Kumar, Optimized Stochastic Policies for Task Allocation in Swarms of Robots, *IEEE Transactions on Robotics* 25 (4) (2009) 927–937.