Mathematics of Planet Earth 4

Arianna Bianchi
Thomas Hillen
Mark A. Lewis
Yingfei Yi *Editors*

The Dynamics of Biological Systems

MPE

Springer

Mathematics of Planet Earth

Volume 4

Springer's Mathematics of Planet Earth series provides a variety of well-written books of a variety of levels and styles, highlighting the fundamental role played by mathematics in a huge range of planetary contexts on a global scale. Climate, ecology, sustainability, public health, diseases and epidemics, management of resources and risk analysis are important elements. The mathematical sciences play a key role in these and many other processes relevant to Planet Earth, both as a fundamental discipline and as a key component of cross-disciplinary research. This creates the need, both in education and research, for books that are introductory to and abreast of these developments.

More information about this series at http://www.springer.com/series/13771

Arianna Bianchi • Thomas Hillen
Mark A. Lewis • Yingfei Yi
Editors

The Dynamics of Biological Systems

MPE

Springer

Editors

Arianna Bianchi
Department of Mathematical
and Statistical Sciences
University of Alberta
Edmonton, AB, Canada

Thomas Hillen
Department of Mathematical
and Statistical Sciences
University of Alberta
Edmonton, AB, Canada

Mark A. Lewis
Department of Mathematical
and Statistical Sciences
University of Alberta
Edmonton, AB, Canada

Yingfei Yi
Department of Mathematical
and Statistical Sciences
University of Alberta
Edmonton, AB, Canada

ISSN 2524-4264 ISSN 2524-4272 (electronic)
Mathematics of Planet Earth
ISBN 978-3-030-22582-7 ISBN 978-3-030-22583-4 (eBook)
https://doi.org/10.1007/978-3-030-22583-4

Mathematics Subject Classification (2010): 35Q92, 37N25, 92B05, 92C15, 92C17, 92C37, 92C40, 92C42, 92D25, 92D30, 92E10

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Although the life sciences and mathematics have historically been separate, the applications of mathematics to solving scientific problems in the life sciences are now experiencing dramatic successes. Complex systems of biological processes on the microscale up to ecological scales can be understood using mathematical models. These range from predicting dynamics of cancers to revealing neuronal pathways used in brain function, to understanding infectious disease outbreaks, to predator-prey dynamics, species extinctions, and foreign species invasions.

A major mathematical tool in this field are *dynamical systems*. The interplay between dynamical systems and biology works in two directions. On the one hand, mathematics is used to reveal biological structures. Mathematical models for the biological processes are written as complex dynamical systems. The analysis of these dynamical systems then reveals emergent properties which then provide biological insight. On the other hand, the biological problems suggest new mathematical problems, which stimulate new areas for mathematical analysis, expanding the area of dynamical systems. Areas such as dynamic network theory, multiscale analysis, nonlinear invasions, pattern formation, and bifurcation theory have been enriched from exposure to biological problems. The design of appropriate mathematical models is quite an art, and it is one purpose of this textbook to introduce the reader to the mathematical modelling with dynamical systems.

The inspiration for this book came from a summer school, *Dynamics of Biological Systems*, part of the esteemed Séminaire de Mathématiques Supérieures (SMS) series of summer schools in mathematics. The goal of this summer school, held at the University of Alberta in 2016, was to investigate connections between dynamical and biological systems and to illuminate the rich interactions between science and mathematics that have been so successful to date. The focus was to understand the mathematical structures of dynamical systems that come from biological problems, then, relating the mathematical structures back to the biology to provide scientific insight. We were fortunate to have exceptional lecturers at the summer school, who have made significant contributions to ecological and biological modelling. Each was invited to write a chapter encapsulating the topic of his or her lectures. In this book, you will see the results of these invited contributions, plus an introductory chapter written by some of the book editors. The methods and topics of many of the chapters are directly relevant to global issues of ecosystems, such as extinctions, invasions, biodiversity, epidemics, and climate change, and we are grateful for the opportunity to contribute to the *Mathematics of Planet Earth.*

The purpose of the introductory chapter by T. Hillen and M.A. Lewis is to set the stage for the forthcoming topics. It reviews basic principles from the theory of dynamical systems such as stability, linearizations, and bifurcations. It also discusses the use of survival times and transition rates in modelling with ordinary and partial differential equations.

The area of systems biology aims to describe how the behavior of a biological system arises from the dynamics of its constituents. A large area within systems biology is the modelling of biochemical networks and food-web networks and their various methods of analysis. In Chap. 2, G. Yang and R. Albert introduce a topological description of networks, which allows the network dynamics to be described by a Boolean network.This methodology is scalable in the sense that large biochemical networks can be treated without much additional effort. The Boolean networks have been used successfully in many biological and medical contexts.

With the advent of new disease outbreaks, the accurate modelling of an epidemic is a modern-day challenge. Many models simplify disease dynamics so as to allow for straightforward mathematical analysis. However, in Chap. 3, M.Y. Li takes another approach. He considers epidemic models that can include heterogeneity of the host population, spatial distributions of the hosts, and detailed life cycles of the infectious agent. The resulting epidemic models are large scale, and new methods for analysis are needed. One such method, introduced here, is a graph-theoretical approach. This approach allows the construction of Lyapunov functions and a computation of the basic reproductive number R_0.

Another approach to consider epidemic models in heterogeneous landscapes is given by metapopulation models, as presented in Chap. 4 by Z. Feng and J.W. Glasser. Here, the host population is split into smaller groups (metapopulations) according to certain criteria. These metapopulations could reside in different locations (such as villages) or different social groups (such as parents, children, co-workers, etc.). Z. Feng and J. Glasser show how epidemic dynamics depend on disease transmission both within and between different groups. They analyze how this information can be used to device vaccination strategies.

In Chap. 5 by J. Wu, we encounter differential delay equations. J. Wu considers vector-borne infections, where the dynamics of the vector (e.g. mosquito or tick) must be included in the modelling. In many situations, this leads to differential equations with delay, and J. Wu explains how the resulting models can be analyzed and be applied to tick transmission of Lyme disease and bird transport of the avian flu virus.

In Chap. 6, authored by H. Qian, we enter the realm of stochastic population modelling. We again encounter the survival times first introduced in Chap. 1. Here, they are used to define discrete-time Markov processes for population dynamics and their corresponding Kolmogorov forward and backward equations. The analogy to classical thermodynamics is used to make a connection to stochastic biological modelling in the new framework of *mathematicothermodynamics*, which offers a new biological modelling paradigm.

The Turing model is one of the workhorses of biological pattern formation. Initially proposed by Alan Turing in 1957, the Turing model has now been used to describe species distributions, animal coat patterns, and vegetation patterns in landscapes. P.K. Maini and T.E. Wooley review this mechanism in Chap. 7, including beautiful examples of animal skin patterns, bone structures, and seashells. This is the first application of partial differential equations encountered in the book.

Chapter 8 takes the analysis of reaction-diffusion equations, which was started in the previous chapter, to a new level. K.Y. Lam and Y. Lou showcase the broad applicability of reaction-diffusion models for species competition, species persistence, heterogeneous landscapes, drift-diffusion problems, and evolution of dispersal. They carefully introduce the mathematical setting of eigenvalue problems, variational principles, super- and sub-solutions, and Lyapunov functions as tools for the analysis of reaction-diffusion models. Open problems for further research are suggested.

Chapter 9 by B. Perthame exposes the reader to the modelling of species movement by transport equations, also called kinetic equations. B. Perthame emphasises the roots of these models in physics and shows how physical scaling principles can be used in the biological context.The chapter includes the detailed functional analytic setting and a proof of weak convergence to the parabolic limit.

The chapters of this book cover a wide range of mathematical methods and biological applications. We are grateful to the high-quality contributions of leading experts in this area, which inspire readers to get involved in active research. We hope we were able to shine a bright light onto the beautiful tools that arise through the modelling with dynamical systems. We would be thrilled if this book sparks new ideas for the *mathematics of planet earth*.

Edmonton, AB, Canada/Siena, Italy — Arianna Bianchi
Edmonton, AB, Canada — Thomas Hillen
Edmonton, AB, Canada — Mark A. Lewis
Edmonton, AB, Canada — Yingfei Yi

Acknowledgements

We thank the Canadian Mathematical Society, Centre de Recherches Mathématiques, Fields Institute, Institut des Sciences Mathématiques, Mathematical Sciences Research Institute, Pacific Institute for the Mathematical Sciences, Springer, and the University of Alberta for their generous financial support. We also thank numerous student volunteers from the University of Alberta who helped with the programme. Finally, we give special thanks to Kim Budinski for her tireless help with organising the summer school programme from which the book arose.

Contents

Contributors

Editors

Arianna Bianchi Siena, Italy

Thomas Hillen University of Alberta, Edmonton, AB, Canada

Mark A. Lewis University of Alberta, Edmonton, AB, Canada

Yingfei Yi University of Alberta, Edmonton, AB, Canada

Authors

Réka Albert Pennsylvania State University, University Park, PA, USA

Zhilan Feng Purdue University, West Lafayette, IN, USA

John W. Glasser National Center for Immunization and Respiratory Diseases, CDC, Atlanta, GA, USA

Thomas Hillen University of Alberta, Edmonton, AB, Canada

King-Yeung Lam The Ohio State University, Columbus, OH, USA

Mark A. Lewis University of Alberta, Edmonton, AB, Canada

Michael Y. Li University of Alberta, Edmonton, AB, Canada

Yuan Lou Renmin University of China, Beijing, People's Republic of China
The Ohio State University, Columbus, OH, USA

Philip K. Maini University of Oxford, Oxford, UK

Benoît Perthame Sorbonne University, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions, Paris, France

Hong Qian University of Washington, Seattle, WA, USA

Thomas E. Woolley Cardiff University, Cardiff, UK

Jianhong Wu York University, Toronto, ON, Canada

Gang Yang Pennsylvania State University, University Park, PA, USA

Chapter 1
Dynamical Systems in Biology: A Short Introduction

Thomas Hillen and Mark A. Lewis

Abstract The contributions to this textbook are based on a summer school on Dynamics of Biological Systems as part of the series “Séminaire de Mathématiques Supérieures,” which was held at the University of Alberta in June 2016. The lectures cover a wide variety of topics and it would be presumptuous to assume that all readers are equally familiar with all the background material. Hence we use this introduction to lay down basic concepts on mathematical modelling, stability analysis, nondimensionalizations, partial and ordinary differential equations, basic population and epidemic models, random walk models, travelling wave solutions, and the critical domain size problem. Experienced researchers can easily skip this chapter.

1.1 Essentials of Survival-Time Analysis

Survival time is the foundation to many models for planet earth. Although classically formulated as time until death, survival time can be generalized to denote time until any event, whether earthquake, species extinction, passage of a protein through a membrane, or change of infection state. We start our analysis of survival time from the perspective of a dynamical system.

Mathematical models in the form of ordinary differential equations (ODEs) often use simple rate transition terms such as $\dot{x} = -\mu x$. In this context people often refer to a Poisson process or exponential distribution, and it is sometimes not clear what they mean. We like to use this section to understand where such terms in differential equations come from and what is really behind the assumption of a *Poisson process*. This will also clarify the relationship between rates and probabilities and highlight

T. Hillen (✉) · M. A. Lewis
University of Alberta, Edmonton, AB, Canada
e-mail: thillen@ualberta.ca; mark.lewis@ualberta.ca

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_1

a common misunderstanding of this relation. Furthermore, the general framework developed here allows us to make a connection to delay differential equations.[1]

1.1.1 Basic Notations

We are interested in individuals that can change their state. For example, susceptible individuals can get infected, prey individuals can be hunted, juvenile individuals can mature, mature individuals can reproduce, etc. We like to understand the expected time that an individual stays in a given state (i.e., time to get infected, time to get eaten, time to mature, etc.), and to use this survival time in our modelling. We introduce notation from [33]:

- Let a be the time that an individual spends in a given state. This time a has many different names, depending on the application. The most general formulation is the *sojourn time*, but other names are *waiting time*, *interevent time*, *survival time*, or *residency time* [1, 33, 35].
- Let $F(a)$ denote the probability that an individual has not left the state before or at time a. Often $F(a)$ is simply called the survival probability, where here survival must be understood as survival in a given state until the individual moves to the next state. We call F the *sojourn function* or *survival function* and we assume that $F(a)$ is non-increasing and $F(0) = 1$. If $\lim_{a\to\infty} F(a) = 0$, then each individual has to leave the state eventually. If $F(a) = 0$ for all $a > c$, then there is a maximum state duration time c and all individuals will have left before time c.
- If T denotes a random variable for the time to exit a given state, then

$$F(a) = P(T > a).$$

- The function $G(a) = 1 - F(a) = P(T \le a)$ denotes the probability to have left before time a.

1.1.2 Conditional Probabilities and Exit Rates

We are interested in the conditional probability to still remain in the state for h time units longer, given that the individual stayed already up to time a. This conditional probability is given by

[1] This section is based on the more detailed presentation in Thieme [33].

$$F(h|a) = \frac{F(a+h)}{F(a)}. \tag{1.1}$$

The conditional probability to exit exactly between time a and $a+h$, given that the individual was in the state at time a is then

$$\frac{F(a) - F(a+h)}{F(a)} = 1 - F(h|a).$$

If F is differentiable, then we define the *exit rate* as

$$\mu(a) = \lim_{h\to 0} \frac{F(a) - F(a+h)}{hF(a)} = -\frac{F'(a)}{F(a)}. \tag{1.2}$$

Note that since F is non-increasing the rate $\mu(a)$ is non-negative. If F is not differentiable, then we still use (1.2) with the distributional derivative of F.

Example 1.1 (Exponential Distribution) The first and most important example is the exponential distribution. That is, we assume that the exit time is exponentially distributed and the sojourn function is given by

$$F(a) = e^{-\gamma a}, \qquad \gamma > 0.$$

In this case we can easily compute the rate (1.2) as

$$\mu(a) = \gamma,$$

and the conditional probability (1.1)

$$F(h|a) = e^{-\gamma h} = F(h).$$

Hence the conditional probability of surviving h time units longer is independent of the time spent in the state. In fact, the exponential distribution is the only distribution with this property:

Theorem 1.1 (From Proposition 12.8 in Thieme [33]) *$F(h|a)$ is independent of a if and only if $F(a) = e^{-\gamma a}$, for some constant $\gamma \geq 0$.*

In case that the time increment $h = \Delta t$ is small, we can expand the exponential and find the probability of leaving the state in the interval $[t, t + \Delta t]$ as

$$G(\Delta t) = 1 - F(\Delta t|t) = 1 - e^{-\gamma\Delta t} \approx \gamma\Delta t + o(\Delta t).$$

In fact, the approximation

$$G(\Delta t) \approx \gamma\Delta t \tag{1.3}$$

is often used to explain the relationship of a rate to a probability. We see here that this relationship is an approximation of the exponential function for small time increments Δt.

We can consider a similar expansion for general (differentiable) survival probabilities $F(a)$ as

$$G(\Delta t) = 1 - F(\Delta t|t) \approx F'(0|t)\Delta t + F''(0|t)\frac{(\Delta t)^2}{2} + h.o.t.$$

Let us take a brief look at the use of the relation (1.3) as it is often found in the literature. Consider recovery from a disease and let us assume that 2 out of 20 individuals recover per day. Then the probability to recover in one day is $G(1) = 2/20 = 0.1$. The corresponding rate according to (1.3) is $\gamma = G(1)/1 = 0.1$. The rate here has units day^{-1}. The probability to recover in 1/2 a day equals $G(0.5) = 1/20$ and the corresponding rate is $\gamma = (1/20)/(1/2) = 0.1$. Similarly, the probability to recover in 2 days is $G(2) = 4/20$ and the rate is $\gamma = (4/20)/2 = 0.1$. We see that the rate remains constant, but the probability of change depends on the time interval chosen. It should be noted, though, that the rate has units day^{-1} and if these units are changed, to weeks^{-1}, for example, then the rate changes as well.

Let us consider a simple probabilistic model for the recovery process. If $I(t)$ denotes a random variable for the number of infected individuals at time t, then

$$I(t + \Delta t) = I(t) - G(\Delta t)I(t).$$

We subtract $I(t)$ and divide by Δt to obtain

$$\frac{I(t + \Delta t) - I(t)}{\Delta t} = -\frac{G(\Delta t)}{\Delta t}I(t).$$

Passing to the limit $\Delta t \to 0$ and using (1.3), we arrive at an ODE

$$\frac{d}{dt}I(t) = -\gamma I(t).$$

1.1.3 Age Structured Models

The survival time analyses from previous sections include age structure, as given by the dependence of the exit rate μ on age a in Eq. (1.2). We now focus on the issue of how to model age dependency in more detail via age structured models. These models can be applied to biological processes ranging from cells to populations of organisms.

In the general case we found the rate

$$\mu(a) = -\frac{F'(a)}{F(a)}.$$

In this case it can be shown that the population density satisfies an age structured model (see Thieme [33] for details). The *McKendrick* model describes the population density $u(t, a)$ of the number of individuals with state-age a at time t:

$$\begin{aligned} u_t + u_a &= -\mu(a)u \\ u(t, 0) &= B(t) \\ u(0, a) &= u_0(t), \end{aligned} \tag{1.4}$$

where the indices t and a describe partial derivatives. The function $B(t)$ describes the individuals that enter the state with state-age 0. In addition to (1.4) we also assume that no individual stays forever, i.e., $u(t, \infty) = 0$.

The total state contents are then

$$N(t) := \int_0^\infty u(t, a)da.$$

Example (Exponential Exit Times) In the case of exponential exit times $F(a) = e^{-\gamma a}$ we find $\mu(a) = \gamma$ and we can integrate (1.4) with respect to a:

$$\int_0^\infty u_t da + \int_0^\infty u_a da = -\mu \int_0^\infty u da$$

which gives a linear birth–death ODE for N:

$$\dot{N} = B(t) - \mu N(t). \tag{1.5}$$

For the general case of $F(a)$ it was shown in Thieme [33] that we can derive also an equation for $N(t)$.

Theorem 1.2 (Thieme [33]) *Assume $B(t)$ is continuous and $F(a)$ is continuously differentiable with $F'(a) \leq KF(a)$. Then*

$$\dot{N}(t) = B(t) - C(t) \tag{1.6}$$

with

$$C(t) = \int_0^t \mu(a)B(t-a)F(a)da + \int_t^\infty \mu(a)F(a)\frac{u_0(a-t)}{F(a-t)}da,$$

where $u_0(a)$ is the age distribution at time $t = 0$. If F is not differentiable, we can still write $C(t)$ using the integral over the measure $dF(a)$ as:

$$C(t) = -\int_0^t B(t-a)dF(a) - \int_t^\infty \frac{u_0(a-t)}{F(a-t)}dF(a). \tag{1.7}$$

Example 1.2 (Fixed Stage Duration) Another interesting example is the case where individuals stay in the state for exactly τ time units and then they leave immediately. In that case

$$F(a) = \begin{cases} 1 & a \le \tau \\ 0 & a > \tau \end{cases}. \tag{1.8}$$

The sojourn function F is not differentiable, but we can take the distributional derivative as

$$F'(a) = -\delta_\tau(a),$$

which means the measure in (1.7) is

$$dF(a) = -\delta_\tau(a)da.$$

Then (1.7) becomes

$$\begin{aligned} C(t) &= \int_0^t B(t-a)\delta_\tau(a)da + \int_t^\infty \frac{u_0(a-t)}{F(a-t)}\delta_\tau(a)da \\ &= \begin{cases} B(t-\tau) & t > \tau \\ u_0(\tau - t) & t < \tau \end{cases}. \end{aligned}$$

This leads for $t > \tau$ to a delay differential equation for N:

$$\dot{N}(t) = B(t) - B(t-\tau).$$

1.1.4 Summary of the Sojourn Time Analysis

- The time that individuals spend in a given state can have a general probability distribution $F(a)$.
- The most important case is the exponential distribution $F(a) = e^{-\gamma a}$. In this case the rate $\mu(a) = \gamma$ is constant and the conditional probabilities to live h time units longer do not depend on the actual survival time a. Typical transition terms in differential equation models are based on the exponential distribution. The

understanding that the transition probability in a small time interval $[t, t+\Delta t]$ is given by $P_{\Delta t} \approx \gamma \Delta t$ is in fact an approximation to the true value of

$$P_{\Delta t} = 1 - e^{-\gamma \Delta t}.$$

- A fixed stage duration (1.8) leads to delay differential equations.

1.2 Dynamical Systems and Linear Stability

Nonlinear dynamical systems can describe a broad array of processes in planet earth. Subcellular processes include enzyme kinetics, nerve impulses, and hormonal cycles. Physiological processes include the immune system, as well as organs such as the heart or kidney. At the level of populations, nonlinear dynamical systems can describe population trends, disease outbreaks, and species interactions. At even higher levels, planetary motion can be understood via nonlinear dynamical systems.

Linear stability analysis is the workhorse for the analysis of nonlinear dynamical systems, in particular in biological and medical applications [15, 29]. This method is explained in all standard textbooks of mathematical biology [4, 5, 26] and dynamical systems [15, 29]; hence here, we simply summarize the essential methodology and refer to the literature for details. We cannot resist, however, to include the two-dimensional trace–determinant stability criterion for discrete systems, since this is not included in standard texts (see [16]).

1.2.1 Linear Stability

For a differentiable function $f : \mathbf{R}^n \to \mathbf{R}^n$ consider continuous and discrete n-dimensional population models of the form

$$\dot{x} = f(x), \tag{1.9}$$

$$x_{k+1} = f(x_k), \tag{1.10}$$

respectively. A steady state (fixed point, equilibrium) of (1.9) satisfies $f(\hat{x}) = 0$, and for (1.10) $f(\hat{x}) = \hat{x}$. We denote the Jacobian matrix evaluated at an equilibrium point as $Df(\hat{x})$.

The linearization of the ODE model (1.9) at $\hat{x}$ is

$$\dot{x} = Df(\hat{x})x,$$

while the linearization of the discrete model (1.10) at a steady state is

$$x_{k+1} = Df(\hat{x})x_k.$$

The linearization tells us something about growth and decay of small perturbations around an equilibrium point. Hence here, $x(t)$ and x_k are no longer the solutions of the full nonlinear problems, but rather, they are small perturbations around the steady state $\hat{x}$. To be precise we should have used a different symbol, rather than x, but it is standard to keep the symbol x as in the original equation. If the perturbations converge to zero, then we call a steady state *locally asymptotically stable*, otherwise the equilibrium could be *stable* or *unstable*.

A well-known result on linear stability which is probably used in the majority of Math-Biology publications is

Theorem 1.3 *Consider a differentiable $f : \mathbf{R}^n \to \mathbf{R}^n$ and a dynamical system of the form (1.9) or (1.10). Let $\hat{x}$ denote an equilibrium point and $Df(\hat{x})$ be the Jacobian. Let λ_i, $i = 1, \ldots, n$ denote the eigenvalues of the Jacobian, counted with their multiplicity.*

1. *If $Re(\lambda_i) < 0$ for all $i = 1, \ldots, n$, then $\hat{x}$ is locally asymptotically stable for (1.9).*
2. *If there exists a λ_i such that $Re(\lambda_i) > 0$, then $\hat{x}$ is unstable for (1.9).*
3. *If $|\lambda_i| < 1$ for all $i = 1, \ldots, n$, then $\hat{x}$ is locally asymptotically stable for (1.10).*
4. *If there exists an eigenvalue λ_i with $|\lambda_i| > 1$, then $\hat{x}$ is unstable for (1.10).*

The reverse conclusions are in general not correct. Also note that eigenvalues with $\mathrm{Re}\lambda_i = 0$ for (1.9) and $|\lambda_i| = 1$ for (1.10) need special attention (see detailed bifurcation analysis in Perko [29], Hirsch and Smale [20], Guckenheimer and Holmes [15], Devaney [6]).

1.2.2 Linearization in Two Dimensions

It is worthwhile to study the 2D case in more detail. Assume $f : \mathbf{R}^2 \to \mathbf{R}^2$ be differentiable and let $A := Df(\hat{x})$ denote a linearization at $\hat{x}$. In two dimensions there is a relation between eigenvalues and trace and determinant of the form

$$\det A = \lambda_1 \lambda_2, \qquad \mathrm{tr} A = \lambda_1 + \lambda_2.$$

Written in terms of eigenvalues this gives the *quadratic formula for eigenvalues* or the *trace–determinant formula*:

$$\lambda_{1,2} = \frac{\mathrm{tr} A}{2} \pm \frac{1}{2}\sqrt{\mathrm{tr}^2 A - 4 \det A}. \tag{1.11}$$

In Theorem 1.3 we saw that the real part of this expression describes local stability. If $\mathrm{tr}^2 A - 4 \det A < 0$, then the real part is $\mathrm{tr} A/2$, and if $\mathrm{tr}^2 A - 4 \det A > 0$, then the real part is the whole expression in (1.11). A detailed analysis of the different cases leads to the well-known Zoo of qualitative behavior of steady states in 2D as illustrated in Fig. 1.1 (see [5, 9, 32]).

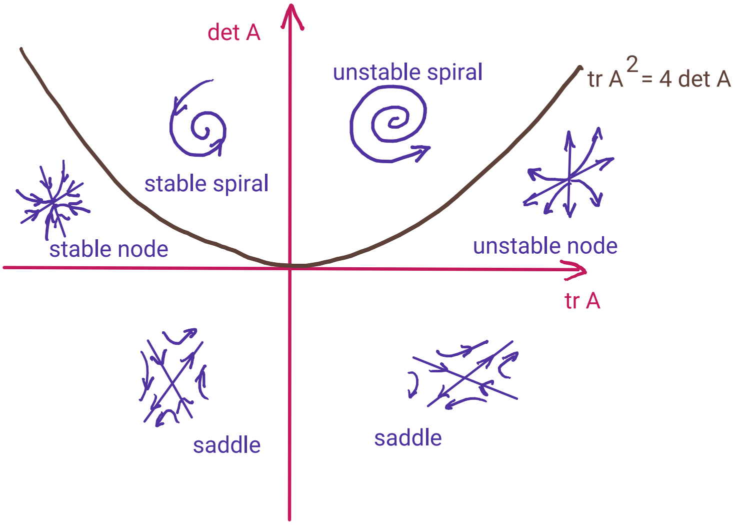


Fig. 1.1 The Zoo of qualitative behavior for two-dimensional ODEs

For the two-dimensional discrete system, stability is given by $|\lambda_{1,2}| < 1$.

- If $\lambda_1 = \bar{\lambda}_2$ are complex, then $\mathrm{tr}^2 A - 4 \det A < 0$, which implies $\det A > 0$. Moreover

$$|\lambda_i|^2 = \lambda_1 \lambda_2 = \det A.$$

Hence $|\lambda_i| < 1$ for complex eigenvalues is equivalent with

$$0 < \det A < 1. \tag{1.12}$$

- If λ_1 and λ_2 are real, then we study two cases. In case of $\mathrm{tr} A > 0$ we find $|\lambda_i| < 1$ for $i = 1, 2$ iff

$$\frac{\mathrm{tr} A}{2} + \frac{1}{2}\sqrt{\mathrm{tr}^2 A - 4 \det A} < 1$$

which can be transformed into the condition

$$\det A > \mathrm{tr} A - 1. \tag{1.13}$$

If $\mathrm{tr} A < 0$, then $|\lambda_i| < 1$ for $i = 1, 2$ is equivalent with

$$-\frac{\mathrm{tr} A}{2} - \frac{1}{2}\sqrt{\mathrm{tr}^2 A - 4 \det A} < 1$$

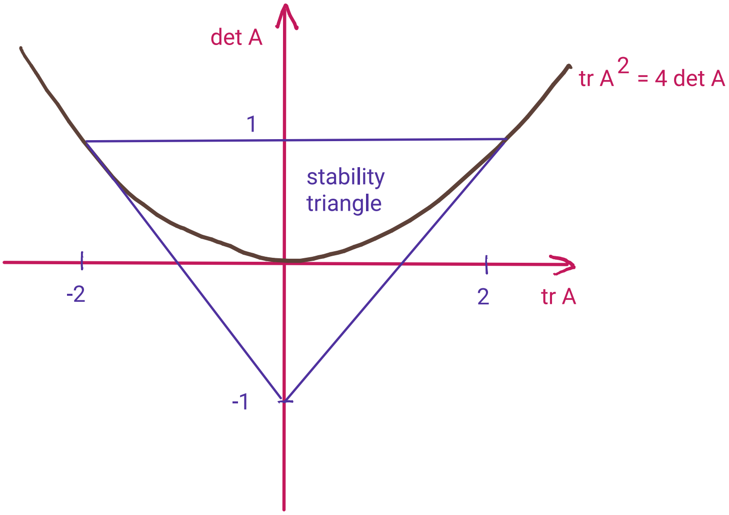


Fig. 1.2 Trace–determinant stability criterion for discrete maps in two dimensions

leading to the condition

$$\det A > -\mathrm{tr} A - 1. \tag{1.14}$$

- Conditions (1.12), (1.13), and (1.14) define a stability triangle as shown in Fig. 1.2.

1.3 Basic Epidemic Models

Mathematical theory has contributed greatly to our understanding of the dynamics of disease outbreaks and persistence. Whether focusing on human health threats, agricultural pests, or impacts of global change on disease states, mathematical epidemiology has provided a modelling structure and formalism to understand disease dynamics.

In epidemic modelling the population is often divided into susceptible $S(t)$, infected $I(t)$, and recovered $R(t)$ [2, 7, 16, 21]. Depending on the disease at hand, each of these classes can be further subdivided into classes of exposed, latent, vaccinated, quarantined, etc. Here we focus on the simple SIR model of susceptible, infected, and recovered and the simplified SI model of Kermack and McKendrick [16, 21]. The basic SIR model reads

$$\begin{aligned}\dot S &= -\beta S I + \gamma R\\ \dot I &= \beta S I - \alpha I\\ \dot R &= \alpha I - \gamma R,\end{aligned} \tag{1.15}$$

where $\beta > 0$ denotes the disease transmission coefficient, $\alpha > 0$ is the recovery rate, and $\gamma > 0$ denotes the rate of loss of immunity. Notice that this model (1.15) does not include birth or death terms and the total population $N = S+I+R$ remains constant during this epidemic.

If recovered individuals stay immune for all time, i.e., $\gamma = 0$, then we obtain the basic Kermack–McKendrick SI-model [21]

$$\begin{aligned}\dot{S} &= -\beta SI\\ \dot{I} &= \beta SI - \alpha I.\end{aligned}\tag{1.16}$$

A phase-plane analysis of this simple SI-model gives us insight into how an epidemic spreads and when it does spread. System (1.16) has a continuum of steady states of the form

$$\{(S^*, 0);\;\; S^* \geq 0\}.$$

The linearization of (1.16) at such a steady state is

$$\begin{pmatrix} 0 & -\beta S^* \\ 0 & \beta S^* - \alpha \end{pmatrix}$$

with eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \beta S^* - \alpha$. The eigenvalue λ_1 corresponds to the direction of the continuum of steady states (S-axis), while λ_2 describes the stability in the orthogonal direction (I-direction). If $S^* < \alpha/\beta$, then $\lambda_2 < 0$ and $(S^*, 0)$ is stable, whereas if $S^* > \alpha/\beta$, then $\lambda_2 > 0$ and $(S^*, 0)$ is unstable (see Fig. 1.3 on the left).

We can visualize the orbits in the phase plane by assuming that they lie on a graph $(S, I(S))$. Then

$$\frac{dI}{dt} = \frac{dI}{dS}\frac{DI}{dt},$$

which implies

$$\frac{dI}{dS} = \frac{\dot{I}}{\dot{S}} = \frac{\beta SI - \alpha I}{-\beta SI} = -1 + \frac{\alpha}{\beta S}.$$

Integrating this equation from S_0 to S gives

$$I(S) = I(S_0) - (S - S_0) + \frac{\alpha}{\beta}(\ln S - \ln S_0).\tag{1.17}$$

Orbits of (1.16) lie on these curves (1.17) and are visualized in Fig. 1.3 on the right. If $S_0 > \alpha/\beta$, then an epidemic outbreak takes place. The remaining susceptible

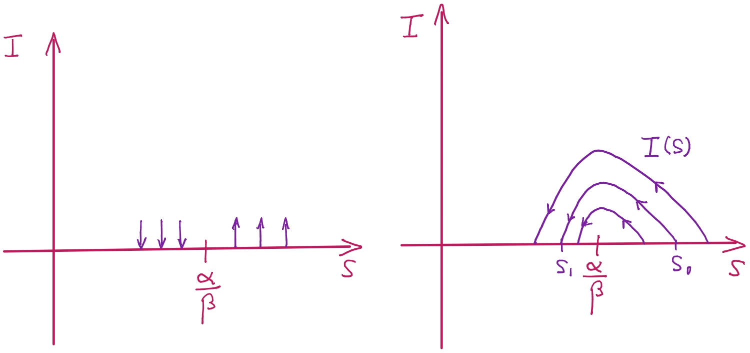


Fig. 1.3 Left: Stability of the equilibria along the S-axis. Right: Sketch of the orbits of the SI epidemic model (1.16)

population, S_1 after the epidemic has passed, can be computed by finding the second solution of $I(S) = 0$. If $S_0 < \alpha/\beta$, then there is no outbreak.

1.3.1 Basic Reproduction Number

In the previous example it turned out that the ratio $\frac{\alpha}{\beta}$ or its inverse $R_0 = \frac{\beta}{\alpha}$ is of particular importance. If we consider a population that is normalized to $S + I \leq 1$ and we consider a fully susceptible population $S^* = 1$, then the disease-free equilibrium $(S^*, 0)$ of (1.16) is unstable for $R_0 = \frac{\beta}{\alpha} > 1$ and stable for $R_0 = \frac{\beta}{\alpha} < 1$. Hence R_0 acts as a bifurcation parameter that distinguishes between outbreak (for $R_0 > 1$) and no outbreak (for $R_0 < 1$). R_0 is called the *basic reproduction number* and it can be computed for many more general epidemic models [2, 7, 16]. A general method to compute R_0 was introduced by Diekmann [8] and van den Driessche and Watmough [34]. In the general framework an epidemic model is split into two processes: new infections, expressed through an operator F, and transition between infected compartments, expressed by an operator V:

$$\dot{x} = F(x) - V(x).$$

If DF and DV denote the linearizations of these operators in the disease-free equilibrium, then van den Driessche and Watmough [34] could show that they have the general form

$$DF = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \qquad DV = \begin{pmatrix} B & 0 \\ \star & \star \end{pmatrix},$$

where A and B are matrices of dimension $m \times m$, where m is the number of infected compartments. We use $\star$ to denote entries that are not important for our argument. The *next generation matrix* is then defined as $A\ B^{-1}$. It roughly measures the new infections that arise, while infected individuals have not yet recovered. Van den Driessche and Watmough [34] could show that the basic reproduction number R_0 arises as the spectral radius of the next generation matrix:

$$R_0 = \rho(A\ B^{-1}).$$

If we use this abstract framework in the context of the simple SI-model (1.16), we notice that we have only one infected compartment I, hence the above matrices A and B have dimension 1×1, i.e., they are scalar. To compute them we simply study the infected equation and linearize with respect to I:

$$\dot{I} = \beta SI - \alpha I = FI - VI$$

with operators $FI = \beta SI$ and $VI = \alpha I$. Linearizing these operators at $(S^*, 0) = (1, 0)$ gives

$$A = \beta \qquad B = \alpha,$$

and the next generation matrix becomes

$$AB^{-1} = \frac{\beta}{\alpha}.$$

Then

$$R_0 = \rho(DFDV^{-1}) = \frac{\beta}{\alpha},$$

as we found before.

1.4 Bifurcations

Bifurcations arise in dynamical systems when a small variation of a model parameter leads to qualitatively different behavior of the corresponding dynamical system. Bifurcations can be local or global. In a local bifurcation the stability of an equilibrium might change, or the number of equilibria might change as a parameter changes [15, 29]. We will discuss the elementary local bifurcations in this section. Global bifurcations arise as the global phase portrait changes as a parameter varies. Global bifurcations arise near heteroclinic and homoclinic orbits and they can lead to new periodic orbits, invariant tori, or chaotic behavior. We will not study global bifurcations here and we refer the interested reader to the literature [13–15, 28].

We consider a dynamical system in $\mathbf{R}^n$ of the form

$$\dot{x} = f(x, \mu),$$

where μ is a parameter and $f(., \mu) : \mathbf{R}^n \to \mathbf{R}^n$ is continuously differentiable. Of course, f can depend on more than one parameter, but here, to explain the basic bifurcation principles, we focus on one scalar parameter μ. We call the class of continuously differentiable vector fields on $\mathbf{R}^n$ as $C^1(\mathbf{R}^n)$, equipped with the usual C^1-norm.

To describe a bifurcation mathematically we need to formalize the understanding that vector fields are "close" and also the notion of "qualitatively the same" (see [13–15, 28]).

Definition 1.1

1. Two vector fields $f, g \in C^1(\mathbf{R}^n)$ are *topological equivalent* if there exists a homeomorphism $H : \mathbf{R}^n \to \mathbf{R}^n$ that maps orbits of $\dot{x} = f(x)$ onto orbits of $\dot{y} = g(y)$ and keeps the orientation of these orbits.
2. A vector field $f \in C^1(\mathbf{R}^n)$ is *structurally stable* if there exists a neighborhood N_f of f in $C^1(\mathbf{R}^n)$ such that each $g \in N_f$ is topologically equivalent to f.
3. Let $f(x, \mu)$ depend continuously on the parameter μ and $f(., \mu) \in C^1(\mathbf{R}^n)$ for each μ. The parameter value μ_0 is a *bifurcation value* if for each $\varepsilon > 0$ there are values $\mu_1 < \mu_0, \mu_2 > \mu_0$ with $|\mu_i - \mu_0| < \varepsilon$ for $i = 1, 2$, such that $f(x, \mu_1)$ and $f(x, \mu_2)$ are not topologically equivalent.

If $f(x, \mu)$ depends continuously on μ, then the set $\{f(., \mu); |\mu - \mu_0|\} < \varepsilon\}$ for ε small enough forms a subset of the neighborhood N_f from item 2.

A trivial example of a bifurcation arises in the model of linear growth and decay

$$\dot{x} = \mu x.$$

For $\mu > 0$ we obtain exponential growth, for $\mu = 0$ we have constant solutions, and for $\mu < 0$ solutions decay exponentially. The bifurcation value is $\mu_0 = 0$.

1.4.1 Elementary Local Bifurcations

Here we consider one-dimensional vector fields $f(., \mu) : \mathbf{R} \to \mathbf{R}$ and

$$\dot{x} = f(x, \mu). \tag{1.18}$$

Local bifurcations in one dimension arise at non-hyperbolic fixed points. Remember, a fixed point is called *hyperbolic*, if the linearization has no eigenvalue with real part equal to zero.

Definition 1.2 $\bar{x}$ is a bifurcation point and $\bar{\mu}$ a bifurcation value of (1.18) if

$$f(\bar{x}, \bar{\mu}) = 0, \qquad \frac{\partial}{\partial x} f(\bar{x}, \bar{\mu}) = 0.$$

Notice that $f(\bar{x}, \bar{\mu}) = 0$ means that $\bar{x}$ is an equilibrium and $f'(\bar{x}, \bar{\mu})$ means that this equilibrium is not hyperbolic.

To consider the three elementary bifurcations in one dimension, we simply study the corresponding *normal forms*. Normal forms are essentially the simplest explicit form that has the desired behavior. There is a comprehensive normal form theory, which we will not discuss here (see [15]).

1. **Saddle-node bifurcation:** The normal form of a saddle-node bifurcation is

$$\dot{x} = \mu - x^2. \tag{1.19}$$

Here the vector field is $f(x, \mu) = \mu - x^2$, the equilibria are $\bar{x} = \pm\sqrt{\mu}$, and they exist only for $\mu \geq 0$. The fixed points have the linearization

$$\frac{\partial}{\partial x} f(\bar{x}, \bar{\mu}) = -2\bar{x} = \mp\sqrt{\mu}.$$

Hence the non-hyperbolic fixed point is $\bar{x} = 0$ for $\mu = 0$. In Fig. 1.4 we show the phase line plots for three values of $\mu < 0, \mu = 0, \mu > 0$. We see that for $\mu < 0$ we have no equilibrium, while at $\mu = 0$ we have exactly one equilibrium, which is non-hyperbolic and for $\mu > 0$ we have two equilibria, where $-\sqrt{\mu}$ is unstable and $\sqrt{\mu}$ is asymptotically stable. The lower figure in Fig. 1.4 shows the corresponding bifurcation diagram. We plot the equilibrium points as function of the bifurcation parameter μ where we indicate stable fixed points by a solid line and unstable fixed points by a dashed line.

2. **Transcritical bifurcation:** The normal form of a transcritical bifurcation is

$$\dot{x} = \mu x - x^2. \tag{1.20}$$

Here the vector field is $f(x, \mu) = \mu x - x^2$, the equilibria are $\bar{x} = 0$ and $\bar{x} = \mu$. The fixed points have the linearization

$$\frac{\partial}{\partial x} f(\bar{x}, \bar{\mu}) = \begin{cases} \mu & \text{for } \bar{x} = 0 \\ -\mu & \text{for } \bar{x} = \mu. \end{cases}$$

Hence the non-hyperbolic fixed point is $\bar{x} = 0$ for $\mu = 0$. In Fig. 1.5 we show the phase line plots for three values of $\mu < 0, \mu = 0, \mu > 0$. We see that for $\mu < 0$ the fixed point $\bar{x} = \mu$ is unstable, while 0 is stable. At $\mu = 0$ the fixed point 0 is non-hyperbolic and for $\mu > 0$ the point 0 is unstable and μ is asymptotically stable. The lower figure in Fig. 1.4 shows the corresponding bifurcation diagram. The origin switches stability as the other fixed point transitions through 0.

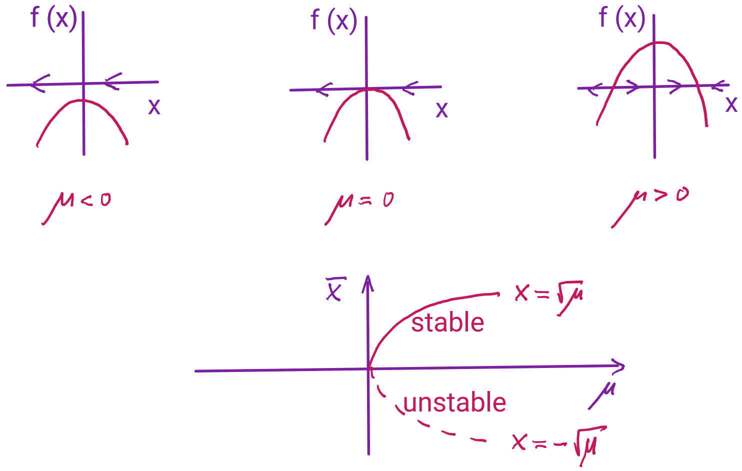


Fig. 1.4 Top row: qualitative different phase line plots. Bottom row: bifurcation diagram of a saddle-node bifurcation

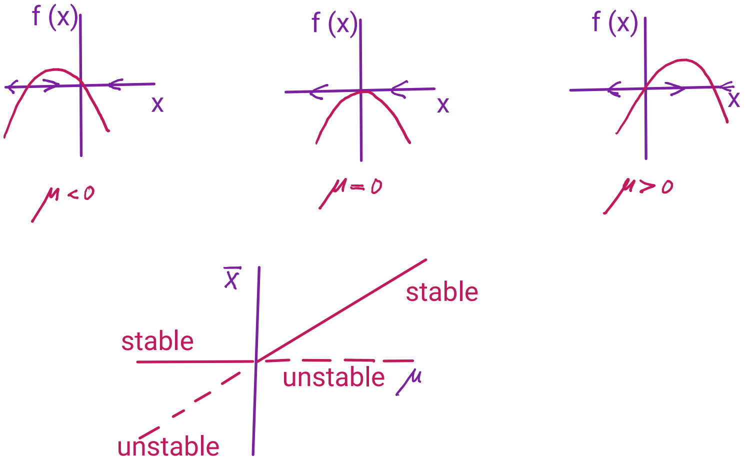


Fig. 1.5 Top row: qualitative different phase line plots. Bottom row: bifurcation diagram of a transcritical bifurcation

3. **Pitchfork bifurcation:** The normal form of a pitchfork bifurcation is

$$\dot{x} = \mu x - x^3. \tag{1.21}$$

Here the vector field is $f(x, \mu) = \mu x - x^3$, the equilibria are $\bar{x} = 0$ and $\bar{x} = \pm\sqrt{\mu}$. The latter only exist for $\mu \geq 0$. The fixed points have the linearization

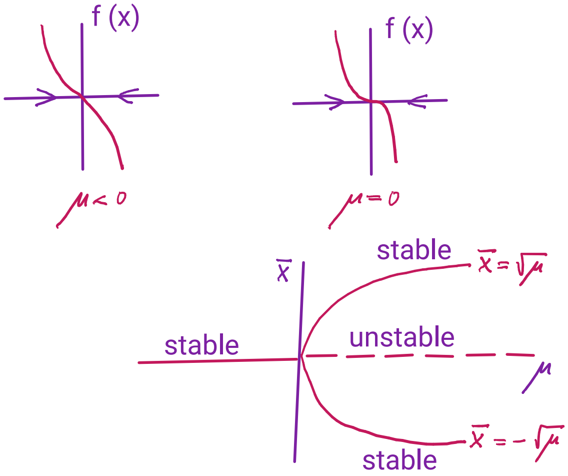


Fig. 1.6 Top row: qualitative different phase line plots. Bottom row: bifurcation diagram of a pitchfork bifurcation

$$\frac{\partial}{\partial x} f(\bar{x}, \bar{\mu}) = \begin{cases} -2\mu & \text{for } \bar{x} = -\sqrt{\mu} \\ \mu & \text{for } \bar{x} = 0 \\ -2\mu & \text{for } \bar{x} = \sqrt{\mu}. \end{cases}$$

Hence the non-hyperbolic fixed point is $\bar{x} = 0$ for $\mu = 0$. In Fig. 1.6 we show the phase line plots for three values of $\mu < 0$, $\mu = 0$, $\mu > 0$. We see that for $\mu < 0$ we have one stable equilibrium at $\bar{x} = 0$. At $\mu = 0$ we still have a stable equilibrium at $\bar{x} = 0$, but it is non-hyperbolic. For $\mu > 0$ we have three equilibria, where $\bar{x} = 0$ is unstable and $\bar{x} = \pm\sqrt{\mu}$ are both asymptotically stable. The lower figure in Fig. 1.6 shows the corresponding bifurcation diagram.

Bifurcations in one-dimensional vector fields cannot lead to periodic solutions. However, this is possible in two dimensions and higher and the prototype of a bifurcation that creates periodic orbits is the *Hopf bifurcation*. Again we look at a normal form

$$\dot{x} = -y + x(\mu - x^2 - y^2)$$
$$\dot{y} = x + y(\mu - x^2 - y^2).$$

Using planar polar coordinates (r, θ) the above system can be written as

$$\dot{r} = \mu r - r^3 \tag{1.22}$$
$$\dot{\theta} = 1. \tag{1.23}$$

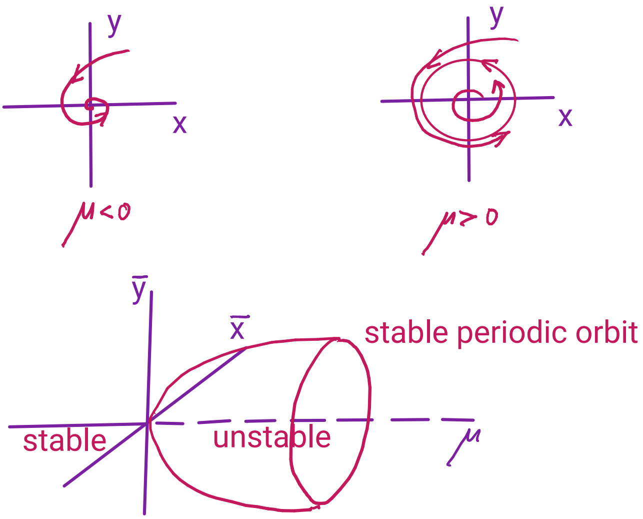


Fig. 1.7 Top row: qualitative different phase plots. Bottom row: bifurcation diagram of a Hopf bifurcation

The radial equation (1.22) has the normal form of a pitchfork bifurcation (1.21) and the bifurcation point $\bar{r} = 0$ corresponds to the origin $(\bar{x}, \bar{y}) = (0, 0)$. The angular equation (1.23) describes rotation around the origin with constant speed 1. Hence the equilibrium point $\bar{r} = \sqrt{\mu}$ becomes a periodic orbit for the system (1.22) and (1.23). The point $\bar{r} = -\sqrt{\mu}$ does not exist, since $r \geq 0$. The creation of a periodic orbit from a pitchfork bifurcation is shown in Fig. 1.7.

1.5 Diffusion as a Random Walk

The idea of diffusion originates from the movement of pollen particles via Brownian motion. However, in practice, diffusion can be used to model movement patterns where there is a random component [3, 27]. Whether animals, cells, or even stock prices, the underlying mathematical model is similar.

1.5.1 Diffusion Process

We define $p(x, t)$ to be the probability density function for a particle moving randomly on a lattice with lattice spacing λ and time step τ. At each time step the

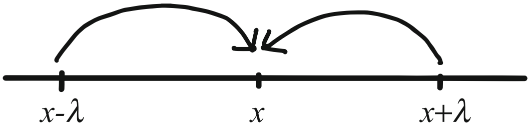


Fig. 1.8 A single step of the random walk process

particle jumps to the left with probability $1/2$ and jumps to the right with probability $1/2$. The *master equation* describes how $p(x, t)$ changes during one step of the random walk [1]. *Master equation*:

$$p(x, t + \tau) = \frac{1}{2}\, p(x - \lambda, t) + \frac{1}{2}\, p(x + \lambda, t). \tag{1.24}$$

Here an individual at x at time $t + \tau$ can have arrived from either the left or from the right (Fig. 1.8).

We consider the case where the space and time steps are small and approximate $p(x, t)$ in Eq. (1.24) using Taylor series expansions in x and t

$$\begin{aligned} p(x, t) + \tau \frac{\partial p}{\partial t}(x, t) + \frac{(\tau)^2}{2} \frac{\partial^2 p}{\partial t^2}(x, t) + \text{h.o.t.} \\ = \tfrac{1}{2} \left\{ p(x, t) - \lambda \frac{\partial p}{\partial x}(x, t) + \frac{(\lambda)^2}{2} \frac{\partial^2 p}{\partial x^2}(x, t) + \text{h.o.t.} \right. \\ \left. + \; p(x, t) + \lambda \frac{\partial p}{\partial x}(x, t) + \frac{(\lambda)^2}{2} \frac{\partial^2 p}{\partial x^2}(x, t) + \text{h.o.t.} \right\}. \end{aligned} \tag{1.25}$$

This simplifies to

$$\frac{\partial p}{\partial t} + \frac{\tau}{2} \frac{\partial^2 p}{\partial t^2} = \frac{(\lambda)^2}{2\tau} \frac{\partial^2 p}{\partial x^2} + \text{h.o.t.} \tag{1.26}$$

We consider the limit where the space and time steps approach zero. If we choose the limit carefully so that $\lambda, \tau \to 0$ so that $\frac{(\lambda)^2}{2\tau} \to D$, then Eq. (1.26) yields the *diffusion equation*

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}. \tag{1.27}$$

The probability density function for the initial location of the particle gives the initial condition $p(x, 0) = p_0(x)$ for Eq. (1.27).

The limit $\lambda, \tau \to 0$ so that $\frac{(\lambda)^2}{2\tau} \to D$ is referred to as the *diffusion limit* and leads to the diffusion equation (1.27). See [27] for a discussion of alternative limits and resulting models.

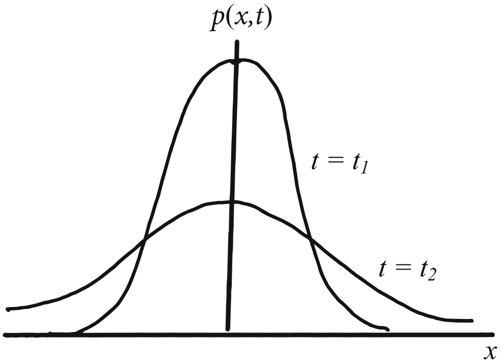


Fig. 1.9 Fundamental solution to the diffusion equation (1.27)

1.5.2 Fundamental Solution to the Heat Equation

If we consider an individual released at $x = 0$ at time $t = 0$, then $p_0(x) = \delta(x)$. The solution to (1.27) for this initial condition is called the *fundamental solution to the heat equation* and is given by

$$p(x,t) = \frac{1}{2\sqrt{\pi Dt}} e^{-\frac{x^2}{4Dt}}. \tag{1.28}$$

Figure 1.9 shows the solution (1.28) for different values of $t > 0$, and can be interpreted as a Gaussian distribution with zero mean and variance $2Dt$. The growth in the variance over time represents increasing uncertainty regarding the location of the particle as time progresses. This solution (1.28) can be found through Fourier transform methods [19]. However, it is straightforward to verify that (1.28) satisfies the diffusion equation (1.27) plus the point source initial condition.

1.5.3 Biased Random Walk

It may be that the probability of jumping to the left or right is not exactly 0.5, as assumed above. For example, there could be drift in a given direction due to underlying wind or water flow [25]. Alternatively, an individual may prefer one direction over another because the environment is more favorable, or because there is a stronger chemical cue in that direction. Regardless of the reason, a slight bias in movement ($R = 0.5 + \gamma\lambda$, $L = 0.5 - \gamma\lambda$) yields an advection–diffusion equation

$$\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x} = D\frac{\partial^2 p}{\partial x^2}, \tag{1.29}$$

where the limiting process is as described above, but with $\frac{\gamma(\lambda)^2}{2\tau} \to \gamma D = v$. A point source initial condition $p_0(x) = \delta(x)$ leads to the fundamental solution

$$p(x,t) = \frac{1}{2\sqrt{\pi D t}} e^{-\frac{(x-vt)^2}{4Dt}}, \tag{1.30}$$

which is a Gaussian with variance $2Dt$, shifting to the right with velocity v.

1.5.4 Correlated Random Walk in One Dimension

We have seen already that there is a close connection between reaction–advection–diffusion equations and random walks. Here we like to present the approach to derive the RD equations from a correlated random walk. Since the scaling of the parameters in this case is different than before, we encounter an intermediate partial differential equation which is called the one-dimensional correlated random walk (CRW). While these CRW equations can be scaled to become the standard diffusion model, it is also useful to consider the CRW model in itself. In fact, we will show here that the CRW system is a transport equation and whence prepare the chapter of B. Perthame on transport equations in biology (see also [16, 18, 30]).

In the framework of a correlated random walk we want to keep track of the correlations in movement directions. We introduce $u^{\pm}(x,t)$ for densities of individuals who arrived at x at time t by moving right (left). The master equation with step size τ and space step δ is

$$u^+(x,t+\tau) = pu^+(x-\delta,t) + (1-p)u^-(x-\delta,t), \tag{1.31}$$

$$u^-(x,t+\tau) = pu^-(x+\delta,t) + (1-p)u^+(x+\delta,t), \tag{1.32}$$

where $p = 1 - \lambda\tau$ is the probability of persisting in the direction of movement and λ is the turning rate.

In the limit

$$\lim_{\tau,\delta\to 0} \frac{\delta}{\tau} = \gamma, \tag{1.33}$$

one obtains the equations for a *one-dimensional correlated random walk (CRW)*

$$\begin{aligned} u_t^+ + \gamma u_x^+ &= \lambda(u^- - u^+), \\ u_t^- - \gamma u_x^- &= \lambda(u^+ - u^-). \end{aligned} \tag{1.34}$$

We can find an equivalent system by using the total population density $u = u^+ + u^-$ and the population flux $v = \gamma(u^+ - u^-)$ to get to the system

$$u_t + v_x = 0, \qquad v_t + \gamma^2 u_x = -2\lambda v, \tag{1.35}$$

which is also known as *Cattaneo system* [17]. Then, dividing the equation for v by 2λ and letting $\lambda, \gamma \to \infty$ with the parabolic limit

$$\lim_{\lambda,\gamma\to\infty} \frac{\gamma^2}{2\lambda} = D < \infty, \tag{1.36}$$

we obtain, again, the diffusion equation

$$u_t = Du_{xx}. \tag{1.37}$$

We like to show now that the one-dimensional model for correlated random walk (1.34) is also a transport equation as introduced in a later chapter by B. Perthame. He considers the space and time evolution of a particle density $f(x, \xi, t)$, where $\xi \in V \subset \mathbf{R}^n$ denotes the actual velocity of this particle. The particle can change direction and, as assumed by B. Perthame, it often chooses a new direction that corresponds to a given equilibrium density, called the *Maxwellian* $M(\xi)$. The equations are

$$\frac{\partial}{\partial t} f(x,\xi,t) + \xi \cdot \nabla f(x,\xi,t) = k\Big(n(x,t)M(\xi) - f(x,\xi,t)\Big), \tag{1.38}$$

where $k > 0$ is a constant and

$$n(x,t) = \int_V f(x,\xi,t)d\xi$$

is the total population density. The analysis and modelling with (1.38) will be presented later. Here we only like to show that the one-dimensional correlated random walk (1.34) has exactly the form (1.38). Let us go back to (1.34) and write it in an equivalent form as

$$\begin{aligned} u_t^+ + \gamma u_x^+ &= 2\lambda\Big(\tfrac{1}{2}(u^+ + u^-) - u^+\Big), \\ u_t^- - \gamma u_x^- &= 2\lambda\Big(\tfrac{1}{2}(u^+ + u^-) - u^-\Big). \end{aligned} \tag{1.39}$$

We have two speeds $V = \{-\gamma, +\gamma\}$ and we can write

$$u^-(x,t) = f(x,-\gamma,t), \qquad u^+(x,t) = f(x,+\gamma,t).$$

The equilibrium distribution $M(\xi)$ satisfies $\lambda(u^- - u^+) = \lambda(M(-\gamma) - M(+\gamma)) = 0$, hence $M(-\gamma) = M(+\gamma)$. Furthermore, M is normalized, i.e.,

$$1 = \int_V M(\xi)d\xi = M(-\gamma) + M(+\gamma)$$

which implies

$$M(\xi) = \frac{1}{2}, \qquad \xi \in \{-\gamma, +\gamma\}.$$

The total population density is

$$n(x,t) = \int_V f(x,\xi,t)d\xi = u^+(x,t) + u^-(x,t).$$

Finally we set $k = 2\lambda$. In this case system (1.38) consists of exactly two equations and these are given in (1.39).

1.6 Reaction–Diffusion Models

An alternative method for deriving advection–diffusion equations comes from continuum modelling. This approach has the advantage that nonlinear reaction terms, such as birth and death, are easy to include. The underlying variable is the population density, $u(x,t)$. Unlike $p(x,t)$ above, this quantity is not considered to be a probabilistic measure of the spatial location of an individual, but rather is deterministic measure and is therefore applicable to the case where there are many similar individuals in a population, and we are interested in the change in population density over time. The population density could describe reacting chemicals, interacting species, infective individuals, and range of other possibilities.

1.6.1 Balance Laws

A *balance law* describes the rate of change in number of individuals in a given arbitrary region Ω with a smooth boundary. The *flux* in population density, $J(x,t)$, is a vector describing the flow of individuals with units of density times velocity. The rate of change of individuals in Ω is given as a function of the flux across the boundary of Ω, Γ, and the population growth rate, $f(u(x,t))$ (units density per unit time) within Ω (Fig. 1.10). Mathematically, the balance law is written as

$$\frac{\partial}{\partial t}\int_\Omega u(x,t)dV = -\int_\Gamma J(x,t)\cdot n\,dS + \int_\Omega f(u(x,t))dV. \tag{1.40}$$

The *divergence theorem*, which states that

$$\int_\Gamma J(x,t)\cdot n\,dS = \int_\Omega \mathrm{div} J(x,t)dV, \tag{1.41}$$

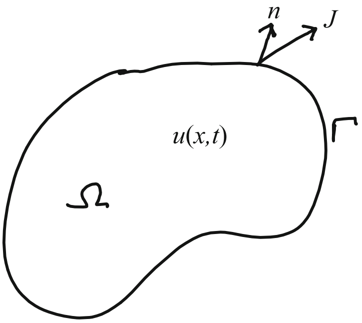


Fig. 1.10 The arbitrary region Ω used in the formulation of a conservation law

can be used to rewrite Eq. (1.40) as

$$\int_{\Omega}\left(\frac{\partial u}{\partial t}+\operatorname{div}J-f(u)\right)dV=0. \tag{1.42}$$

Equation (1.42) must hold for every arbitrary test region Ω. The only way that this can occur is when the integrand is identically zero (almost everywhere), so

$$\frac{\partial u}{\partial t}+\operatorname{div}J=f(u). \tag{1.43}$$

This is referred to as a balance law equation, it balances the change of the particle density via flux J and growth or decay f. The final step needed to formulate a reaction–advection–diffusion equation requires an explicit connection between the flux vector J and the population density and/or its gradient. This connection is developed in the next section.

1.6.2 Modelling the Flux

Typically there are two components to the flux: diffusive and advective. Diffusive flux describes population density moving from high to low, with a magnitude proportional to the gradient of the density so that $J = -D\nabla u$. Advective flux describes population density moving via bulk transport with velocity $\mathbf{v}$ so that $J = \mathbf{v}u$. Both of these processes can act at the same time, so combining the two terms gives

$$J=-D\nabla u+\mathbf{v}u, \tag{1.44}$$

advection–diffusion flux. Substituting the advection–diffusion flux into the balance law equation (1.43) yields the advection–diffusion–reaction equation

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v}u) = \nabla \cdot (D\nabla u) + f(u), \tag{1.45}$$

or in one dimension

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(vu) = \frac{\partial}{\partial x}\left(D\frac{\partial u}{\partial x}\right) + f(u). \tag{1.46}$$

Note that the case where v and D are constant and the reaction term f is zero gives an equation with the same terms as the probabilistic model of a random walk with bias (1.29).

1.6.3 Modelling the Growth

The simplest spatio-temporal models in mathematical biology couple simple diffusive flux with nonlinear growth in one spatial dimension. These models take the form of 1D reaction–diffusion equations [3, 26, 27]

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + f(u). \tag{1.47}$$

The best known examples choose the nonlinear growth function $f(u)$ to be a polynomial. For example, quadratic f,

$$f(u) = ru\left(1 - \frac{u}{K}\right) \tag{1.48}$$

gives rise to logistic growth, with r the intrinsic growth rate, and K the carrying capacity. Here population is assumed to be positive. The growth rate $f(u)$ is zero at the extinction ($u = 0$) and carrying capacity ($u = K$) steady states, is positive below the carrying capacity ($0 < u < K$), and is negative above the carrying capacity ($u > K$). When this quadratic growth function is coupled to diffusion via Eq. (1.51) the resulting model is referred to as Fisher's equation or Fisher-KPP equation [11, 22].

A variant on the logistic growth function is given by the cubic function:

$$f(u) = ru\left(1 - \frac{u}{K}\right)\left(\frac{u - C}{K}\right). \tag{1.49}$$

As before, steady states are given by the extinction ($u = 0$) and carrying capacity ($u = K$). However, an additional threshold steady state ($u = C$) is also included. The growth rate is negative below this threshold steady state ($0 < u < C$), is positive at intermediate densities ($C < u < K$), and is negative at high population densities ($u > K$). This bistable equation describes the so-called *Allee dynamics*,

dynamics which reflect the inability of populations to exhibit net growth until they surpass the critical threshold for density, C [3, 27].

1.6.4 Systems of Reaction–Diffusion Equations

Often spatially distributed populations do not act in isolation, but interact with other distinct populations. Examples are numerous and include populations of interacting chemicals, competing species, and predator and prey populations [3, 24, 27, 31]. The simplest situation is that of two interacting species, each of which diffuses at a different rate in one spatial dimension. In this case, the model for the two population densities, $u(x,t)$ and $v(x,t)$, becomes

$$\frac{\partial u}{\partial t} = D_1 \frac{\partial^2 u}{\partial x^2} + f(u,v), \tag{1.50}$$

$$\frac{\partial v}{\partial t} = D_2 \frac{\partial^2 v}{\partial x^2} + g(u,v). \tag{1.51}$$

1.7 Travelling Waves

One of the key behaviors of scalar reaction–diffusion equations is that of spatial spread. A population that grows locally may spread spatially into new environments as it diffuses from one location to the next and then continues to grow in the new location. Examples range from biological invaders to waves of chemicals across the surface of cells.

In this case, a relevant mathematical structure to consider is that of a *travelling wave* connecting the zero equilibrium ($u = 0$) to the carrying capacity ($u = K$). A travelling wave moves across the spatial domain with a given speed, c, while retaining a fixed shape (Fig. 1.11). In this case, the solution can be written in the form $u(x,t) = U(x - ct)$. The travelling wave speed can be interpreted biologically as the rate at which the population invades a new environment.

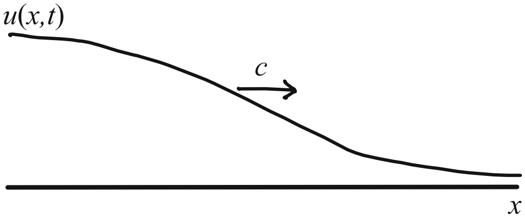


Fig. 1.11 A travelling wave moves across the domain with speed c while retaining its shape

The travelling wave ansatz, $u(x,t) = U(x - ct)$, can be used to translate the terms in Eq. (1.51) into ordinary derivates

$$\frac{\partial}{\partial t}u(x,t) = -cU', \quad \frac{\partial^2}{\partial x^2}u(x,t) = U'',$$

where $'$ indicates differentiation with respect to the travelling wave variable $z = x - ct$. With these substitutions Eq. (1.51) becomes

$$cU' + DU'' + f(U) = 0. \tag{1.52}$$

Our requirement that the travelling wave connects the zero equilibrium to the carrying capacity yields boundary conditions as $U(-\infty) = K, \; U(+\infty) = 0$. The travelling wave problem can be stated in terms of a question: For what value(s) of c does a non-negative travelling wave exist, satisfying

$$cU' + DU'' + f(U) = 0, \quad U(-\infty) = 1, \; U(+\infty) = 0? \tag{1.53}$$

We investigate the travelling wave problem (1.53) by means of introducing the dummy variable $V = U'$ and then analyzing the resulting system of ordinary differential equations in the phase plane [10]. Equation (1.52) becomes

$$U' = V, \tag{1.54}$$

$$V' = -\frac{1}{D}\left(cV + f(U)\right). \tag{1.55}$$

The boundary conditions in (1.53) define a heteroclinic orbit in the (U, V) phase plane going from $(1, 0)$ to $(0, 0)$. Thus the travelling wave problem can be rephrased as asking whether there are values of c such that a heteroclinic orbit from $(1, 0)$ to $(0, 0)$ exists for (1.54)–(1.55) where $U \geq 0$ along the orbit.

For Fisher's equation (1.48), linearization about the leading edge of the wave ($U = 0$, $V = 0$) shows that $(0, 0)$ is a stable spiral for $0 < c < c^* = 2\sqrt{rD}$ (Fig. 1.12), and is a stable node for $c \geq c^*$ (Fig. 1.13). Linearization about the trailing edge of the wave ($U = 1$, $V = 0$) shows that $(1, 0)$ is always a saddle (Figs. 1.12 and 1.13).

The fact that $(0, 0)$ is a stable spiral for $0 < c < c^*$ means that there exists a neighborhood of $(0, 0)$ where the associated travelling wave solution $U(z)$ must go negative (Fig. 1.12). The fact that $(0, 0)$ is a stable node for $c \geq c^*$ allows for the possibility of a heteroclinic orbit with an associated travelling wave solution that is non-negative (Fig. 1.13). The actual proof that $c \geq c^*$ leads to a non-negative travelling wave requires a little more work in the phase plane. Details of the proof are given in [10]. The speed

$$c^* = 2\sqrt{rD} \tag{1.56}$$

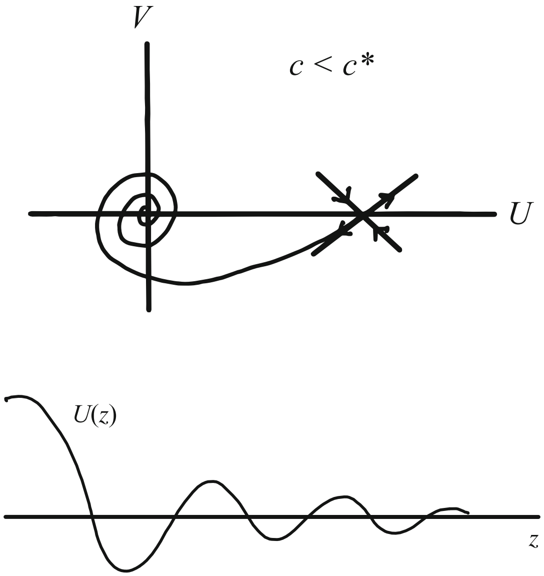


Fig. 1.12 When $c < c^* = 2\sqrt{rD}$ the origin is a stable spiral. This means that $U(z)$ goes negative and there is no non-negative travelling wave solution $U(z)$

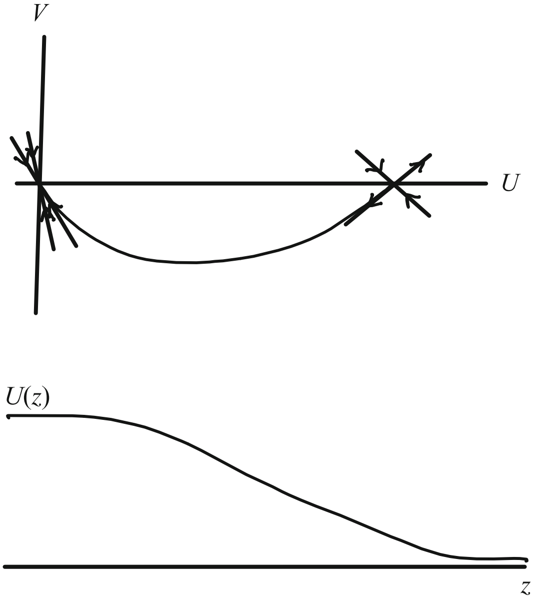


Fig. 1.13 When $c \geq c^* = 2\sqrt{rD}$ the origin is a stable node and there exists a non-negative travelling wave solution $U(z)$

is referred to as the *minimum wave speed* for Fisher's equation. Any travelling wave solution to Fisher's equation must move at a speed which is at least c^*. It is possible to use a dimensional analysis to show that the formula for c^* makes sense from the perspective of units. The quantity r has units of time^{-1}, whereas D has units of length$^{2}\times$ time^{-1}. Thus $2\sqrt{rD}$ has the appropriate units for speed: length $\times$ time^{-1}.

A similar type of analysis for the case with Allee dynamics (1.49) shows that there is a saddle–saddle connection from $(1, 0)$ to $(0, 0)$, and in this case there is a unique wave speed c which gives this heteroclinic orbit [10].

There is a large number of methods to find and analyze travelling waves and invasion speeds and a specialized literature is available. For more details, see [3, 12, 26].

1.8 Critical Domain Size Problem

Consider a population that lives on a patch of habitat with hostile boundaries. We may be interested in how large the patch needs to be in order to support the population. This question arises in the analysis of terrestrial reserves and marine protected areas, which are designed to help ensure the long-term persistence of species at risk.

The mathematical formulation of the equation is given by

$$\begin{aligned}\frac{\partial u}{\partial t} &= D\frac{\partial^2 u}{\partial x^2} + f(u) \qquad (1.57)\\ u(0,t) &= 0,\ \ u(l,t) = 0\\ u(x,0) &= u_0(x),\end{aligned}$$

where $u_0(x)$ is non-negative and not identically zero. We again choose a simple quadratic growth function (1.48), which yields Fisher's equation in (1.57).

For the critical domain problem with Fisher's equation, the following questions are equivalent [3, 5, 23]:

1. How large must a patch be to support a population?
2. What is the critical domain size l_c such that a nontrivial stationary solution (steady state) exists for $l > l_c$?
3. What is the critical domain size l_c such that $u \equiv 0$ is stable for $l < l_c$ and unstable for $l > l_c$?

The formulation of each question takes a different perspective on the same problem. Question 1 takes the biological perspective. Question 2 takes the perspective of existence of a nontrivial steady-state solution $U(x)$ (Fig. 1.14). This can be understood through analysis of

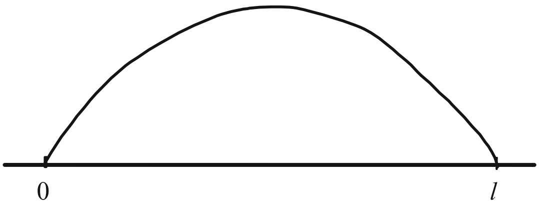


Fig. 1.14 A nontrivial stationary solution to Eq. (1.57)

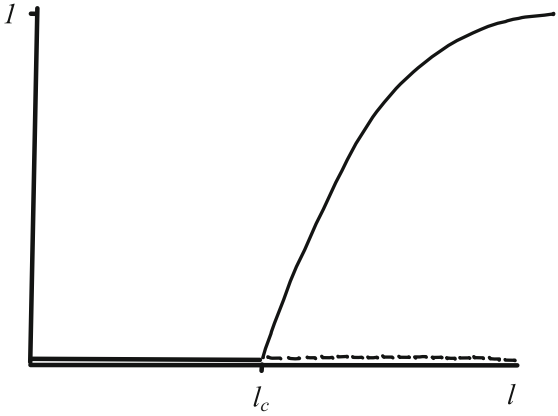


Fig. 1.15 Bifurcation diagram shows the maximum height of the equilibrium solution versus the domain size *l*

$$\begin{aligned} &DU'' + rU\left(1 - \frac{U}{K}\right) \\ &U(0) = 0,\ U(l) = 0, \end{aligned} \tag{1.58}$$

which gives rise to a bifurcation diagram of the form given in Fig. 1.15. Details of the analysis can be found in [23].

Question 3 takes the perspective of the stability of $u \equiv 0$. This can be understood through analysis of the linearized model

$$\begin{aligned} \frac{\partial u}{\partial t} &= D\frac{\partial^2 u}{\partial x^2} + ru \\ u(0,t) &= 0,\ u(l,t) = 0 \\ u(x,0) &= u_0(x). \end{aligned} \tag{1.59}$$

Because (1.59) is a linear equation it can be solved by the method of separation of variables [19]. This yields

$$u(x,t) = \sum_{k=1}^{\infty} B_k e^{\left(r - D\left(\frac{k\pi}{l}\right)^2\right)t} \sin\left(\frac{k\pi}{l}x\right), \tag{1.60}$$

where the constants B_k are determined by initial condition $u(x, 0)$. The population will grow if $r - D(k\pi/l)^2 > 0$ for some wave number k and will not grow if $r - D(k\pi/l)^2 < 0$ for all wave numbers k. The fastest growing mode is associated with wave number $k = 1$, so the population grows if and only if l exceeds

$$l_c = \pi\sqrt{\frac{D}{r}}. \tag{1.61}$$

This leads to an unstable trivial equilibrium $u \equiv 0$. Hence l_c is referred to as the *critical domain size*. The population will not grow if $l < l_c$ and will grow if $l > l_c$. As with the travelling wave problem, it is possible to check to make sure that the units make sense. As before, the quantity r has units of time^{-1} whereas D has units of length^2×time^{-1}. Thus $\pi\sqrt{D/r}$ has the appropriate units for domain size: length.

1.9 Nondimensionalization

Nondimensionalization is one of the longest words used in biological dynamics theory. This is why we have left it to the end of the chapter. Besides of being an extraordinary word, it gives an important method for the analysis of biological models. It makes the analysis independent of physical dimensions such as time, space, size, etc. It reduces the number of independent parameters and it can often be used to identify large and small quantities, indicating fast and slow time scales.

As an example for nondimensionalization we consider the standard Fisher's equation for a spatially distributed population $u(x, t)$, given by Eqs. (1.51) and (1.48), and reproduced here for the reader's convenience:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + ru\left(1 - \frac{u}{K}\right). \tag{1.62}$$

The parameter $r > 0$ denotes the population growth rate, $K > 0$ denotes the carrying capacity, and $D > 0$ is the diffusion coefficient. Hence this is a three parameter model. The spatial domain is either an interval in $\mathbf{R}$ with appropriate boundary conditions or the whole line $\mathbf{R}$. We assume that the parameters have the following units:

$$[D] = \frac{m^2}{s}, \qquad [r] = \frac{1}{s}, \qquad [K] = \#\,\text{cells}, \qquad [u] = \#\,\text{cells}.$$

We use the tilde˜to indicate dimensionless quantities and we start by defining

$$\tilde{u} = \frac{u}{K} \quad \text{with} \quad [\tilde{u}] = \frac{\#\,\text{cells}}{\#\,\text{cells}} = 1.$$

Then

$$\tilde{u}_t = D\frac{u_{xx}}{K} + r\frac{u}{K}\left(1 - \frac{u}{K}\right) = D\tilde{u}_{xx} + r\tilde{u}(1 - \tilde{u}). \tag{1.63}$$

Next we scale time with the growth rate r:

$$\tilde{t} = rt, \qquad [\tilde{t}] = \frac{s}{s} = 1.$$

Then $d\tilde{t} = rdt$ such that $\frac{\partial}{\partial t} = r\frac{\partial}{\partial \tilde{t}}$ and (1.63) becomes

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} = \frac{D}{r}\tilde{u}_{xx} + \tilde{u}(1 - \tilde{u}). \tag{1.64}$$

Finally, we rescale space x as

$$\tilde{x} = \sqrt{\frac{r}{D}}x, \qquad [\tilde{x}] = \sqrt{\frac{1/s}{m^2/s}}m = 1,$$

such that $\frac{\partial}{\partial \tilde{x}} = \sqrt{\frac{D}{r}}\frac{\partial}{\partial x}$. Then (1.64) becomes

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} = \tilde{u}_{\tilde{x}\tilde{x}} + \tilde{u}(1 - \tilde{u}).$$

Finally, we remove the tilde $\tilde{}$, as is standard in most papers, and we get the dimensionless Fisher's equation

$$u_t = u_{xx} + u(1 - u). \tag{1.65}$$

Notice that this model has no free parameter. Now we can analyze the nondimensional model (1.65) and draw conclusions for the original model (1.62). For example, it is known that (1.65) on **R** admits travelling wave solutions of the form $u(x,t) = \phi(x - ct)$ for all speeds that are larger or equal to the minimal speed $c^* = 2$ [3, 26], i.e., formula (1.56) for $r = 1, D = 1$. Detailed travelling wave analysis will be presented in one of the subsequent chapters. If we are interested to compare this minimum speed to a given experiment or observation, we need to know what the minimal speed is in dimensional parameters. Hence we use the tilde $\tilde{}$ again and start with

$$\tilde{c}^* = 2.$$

In original coordinates this becomes

$$2 = \tilde{c}^* = \frac{d\tilde{x}}{d\tilde{t}} = \frac{\sqrt{r/D}dx}{rdt} = \frac{1}{\sqrt{Dr}}c^*.$$

Hence

$$c^* = 2\sqrt{Dr},$$

the Fisher-speed (1.56) for general r, $D > 0$.

1.10 The Dynamical Systems Toolkit

In summary, the tools of biological dynamics are based heavily on the theory of dynamical systems. Thus dynamical systems provide foundational models for mathematical biology as well as for many other areas in the mathematics of planet earth. While they can describe nonlinear interactions between individuals in populations, dynamical systems can also arise from probabilistic concepts, such as failure times and random walks. Models take the form of ordinary and partial differential equations and discrete-time maps. Much of the qualitative theory for these systems relates to equilibria and their stability. The stability of equilibria can be investigated via linearization and bifurcation theory. This becomes particularly simple for phase-plane analysis of two-dimensional systems of ordinary differential equations. Alternatively, special solutions such as travelling waves can sometimes give useful insight regarding long-term behaviors of biological populations, such as population spread. Finally, nondimensionalization is a useful tool for reducing the number of parameters in a model, thereby facilitating easier analysis.

References

1. L.J.S. Allen, *An Introduction to Stochastic Processes with Applications to Biology* (Prentice Hall, Upper Saddle River, 2003)
2. F. Brauer, C. Castillo-Chavez, *Mathematical Models for Communicable Diseases* (SIAM, Hoboken, 2013)
3. N.F. Britton, *Reaction–Diffusion Equations and Their Applications to Biology* (Academic Press, London, 1986)
4. N.F. Britton, *Essential Mathematical Biology* (Springer, Heidelberg, 2003)
5. G. de Vries, T. Hillen, M. Lewis, J. Müller, B. Schönfisch, *A Course in Mathematical Biology* (SIAM, Philadelphia, 2006)
6. R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd edn. (Addison-Wesley, Reading, 1989)
7. O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley, Chichester, 2000)
8. O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**, 365–382 (1990)
9. L. Edelstein-Keshet, J. Watmough, Grunbaum. D, Do travelling band solutions describe cohesive swarms? An investigation for migratory locust. J. Math. Biol. **36**, 515–549 (1998)
10. P. Fife, *Mathematical Aspects of Reacting and Diffusing Systems* (Springer, New York, 1979)

11. R.A. Fisher, The advance of advantageous genes. Ann. Eugenics **7**, 355–369 (1937)
12. B.H. Gilding, R. Kersner, *Travelling Waves in Nonlinear Diffusion-Convection Reaction* (Birkhauser, Basel, 2004)
13. M. Golubitsky, D.G. Schaeffer, *Singularities and Groups in Bifurcation Theory: Vol. I.* Applied Mathematical Sciences vol. 51 (Springer, New York, 1985)
14. M. Golubitsky, I.N. Stewart, D.G. Schaeffer, *Singularities and Groups in Bifurcation Theory: Vol. II.* Applied Mathematical Sciences vol. 69 (Springer, New York, 1988)
15. J. Guckenheimer, P.J. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, Heidelberg, 1983)
16. K.P. Hadeler, *Topics in Mathematical Biology* (Springer, Heidelberg, 2018)
17. T. Hillen, On the L^2-moment closure of transport equations: the Cattaneo approximation. Discr. Cont. Dyn. Syst. B **4**(4), 961–982 (2004)
18. T. Hillen, Existence theory for correlated random walks on bounded domains. Can. Appl. Math. Q. **18**(1), 1–40 (2010)
19. T. Hillen, E. Leonard, H. van Roessel, *Partial Differential Equations; Theory and Completely Solved Problems* (Wiley, Hoboken, 2012)
20. M.W. Hirsch, S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra* (Academic Press, New York, 1974)
21. W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics. Proc. R. Soc. Ser. A **115**, 700–721 (1927). [Reprinted in: G. Oliveira–Pinto, B.W. Conolly, *Applicable Mathematics of Nonphysical Phenomena* (Ellis Horwood, Chichester, 1982), pp. 222–247]
22. A.N. Kolmogorov, I.G. Petrovskii, Piskunov N.S, A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. Bjol. Moskovskovo Gos. Univ. **17**, 1–72 (1937)
23. M. Kot, *Elements of Mathematical Ecology* (Cambridge University Press, Cambridge, 2001)
24. M.A. Lewis, B. Li, H.F Weinberger, Spreading speed and the linear determinacy for two-species competition models. J. Math. Biol. **45**(3), 219–233 (2002)
25. F. Lutscher, E. Pachepsky, M.A. Lewis, The effect of dispersal patterns on stream populations. SIAM Rev. **478**, 749–7725 (2005)
26. J.D. Murray, *Mathematical Biology* (Springer, Berlin, 1989)
27. A. Okubo, S.A. Levin, *Diffusion and Ecological Problems: Modern Perspectives* (Springer, Berlin, 2002)
28. J. Palis, W. de Melo, *Geometric Theory of Dynamical Systems* (Springer, New York, 1982)
29. L. Perko, *Differential Equations and Dynamical Systems*. Texts in Applied Mathematics (Springer, Berlin, 2001)
30. B. Perthame, *Transport Equations in Biology* (Birkhäuser, Basel, 2007)
31. J. Smoller, *Shock Waves and Reaction-Diffusion Equations* (Springer, Berlin, 1982)
32. S. Strogatz, *Nonlinear Dynamics and Chaos* (Westview Press, Boulder, 2000)
33. H.R. Thieme, *Mathematics in Population Biology* (Princeton University Press, Princeton, 2003)
34. P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
35. N.G. van. Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007)

Chapter 2
Modeling of Molecular Networks

Gang Yang and Réka Albert

Abstract In biomolecular systems, various non-identical molecules interact in diverse ways. The field of systems biology aims to understand how the components and interactions of biological systems give rise to the system's behavior and phenotypes. Researchers have used molecular networks and dynamic models to represent and understand biological systems. In this chapter, we introduce the network representation and the graph measures that quantify its topological properties. We describe how to build a discrete dynamic (Boolean) model of a biological system from experimental data, and how to use the model to provide insights into emergent phenomena and make useful predictions. We also introduce methods to bridge the network's topological and dynamical properties. We use real biological system involved in complex disease to demonstrate the theoretical framework. Discrete dynamical models, especially Boolean networks, benefit from the current high-throughput technologies and large amounts of qualitative data and provide insight to large-scale systems, where continuous modeling is not possible yet.

2.1 Introduction

Decades of research in molecular biology established a large amount of information about the structure and function of individual molecules in cells. It is now known that various non-identical (macro)molecules such as DNA, RNA, proteins, small molecules interact in diverse ways [9, 50]. The totality of interactions among various molecular components gives rise to cellular functions such as movement or proliferation. Thus, cells are an example of complex interacting systems, as are organs, individuals, or populations. In order to understand such systems, researchers are increasingly using networks to represent the components of the system and their interactions [4, 10, 48, 50, 67].

G. Yang · R. Albert (✉)
Department of Physics, Pennsylvania State University, University Park, PA, USA
e-mail: gzy105@psu.edu; rza1@psu.edu

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics
of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_2

A network (or graph) is a mathematical abstraction, consisting of nodes, which represent different elements, and edges, which specify the pairwise relationships between the elements [10, 48]. In molecular networks, nodes are genes, RNA, proteins, and small molecules; edges indicate interactions and regulatory relationships [4, 67]. Edges can be symmetrical (representing a mutual relationship) or directed (representing mass or information flow from a source to a target). The latter type of edges can also have a sign, representing positive (activating) or negative (inhibitory) influences. The network representation allows the use of graph measures to characterize the organization of the molecular interaction networks. In Sect. 2.2 of this chapter, we will introduce different types of molecular networks and present informative graph measures.

Complex systems exhibit several emergent dynamical properties, such as homeostasis, multi-stability, or synchronization [4, 50, 67]. To understand and explain these emergent behaviors of the system, the network needs to be complemented by a dynamical model. In Sect. 2.3, we describe how to build a discrete dynamical model of a molecular network and how to use the model to make predictions. In Sect. 2.4, we explore methods to connect the topological properties of the interaction network with the emergent dynamics of the complex system.

2.2 The Structure of Biomolecular Networks

2.2.1 Introduction to Biomolecular Networks: Classifications and Examples

Let us review the kinds of interactions possible inside a cell. Genes are transcribed into mRNAs, which are translated into proteins. Proteins called *transcription factors* can activate or inhibit the transcription (also called *expression*) of genes. Proteins interact with each other and may form protein complexes. Proteins called *enzymes* catalyze chemical reactions of the metabolism. Molecules from the environment are metabolized or are sensed by receptor proteins [4, 50, 67]. Biologists usually try to group these interactions and separately define four types of networks, namely *gene regulation*, *protein–protein interaction*, *signal transduction*, and *metabolic networks*, but they are in fact interconnected [4, 50, 67]. In the following we exemplify three types of intracellular networks, in the order of increasing diversity.

Protein–protein interaction networks are formed by biochemical events and/or electrostatic interactions between proteins. Several methods now exist to detect such interactions on a large scale, such as two-hybrid screening [41], biomolecular fluorescence complementation (BiFC) [34], and co-immunoprecipitation (Co-IP) [44]. Such networks have been built for several organisms including *S. cerevisiae*, *Drosophila*, *C. elegans*, and humans [23, 50, 51, 64]. For example, 1870 proteins and 2240 identified direct physical interactions between them are mapped in the *S. cerevisiae* protein–protein interaction network. The network is built through studying combined, non-overlapping data, obtained by systematic two-hybrid analyses [30, 64].

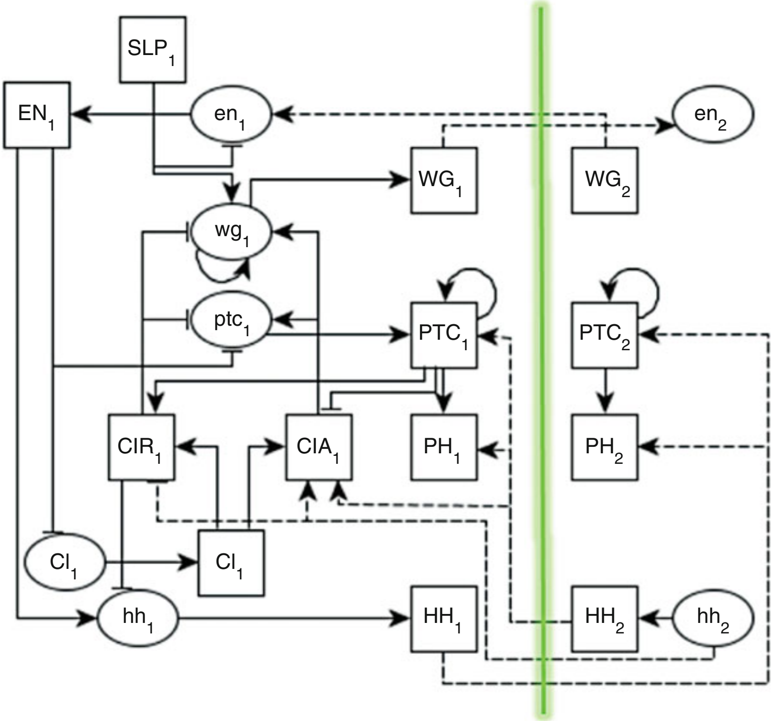


Fig. 2.1 Drosophila segment polarity gene network model. Four cells with periodic boundary conditions are considered and mainly the first cell is shown. The green line indicates a cell boundary. Ellipses represent mRNAs and squares represent proteins. Positive edges terminate in arrow-heads and negative edges terminate in blunt segments. Solid lines indicate intracellular regulation and dashed lines indicate intercellular regulation. Figure is adapted from [3]

A *gene regulatory network* is a set of genes and gene products (mRNA and proteins) that interact with each other and with other molecules in the cell to regulate gene expression levels. For example, genes and their interactions involved in embryonic pattern formation in the fruit fly *Drosophila melanogaster* are mapped into the Drosophila segment polarity network, as shown in Fig. 2.1 [3]. Various dynamical models have been built to understand the embryonic development process [3, 65].

Signal transduction is the process through which living cells receive and respond to various external stimuli. A diverse set of interacting (macro)molecules participate in this process, such as enzymes, other types of proteins, and small molecules. Signal transduction is crucial in the maintenance of cellular homeostasis, in a cell's communications with its surroundings, and in cell behavior such as growth, survival, apoptosis, and movement [22]. Many complex diseases, such as developmental disorders, diabetes, and cancer, arise from mutations or alterations in the expression of signal transduction pathway components [28, 58]. Figure 2.2 depicts an example

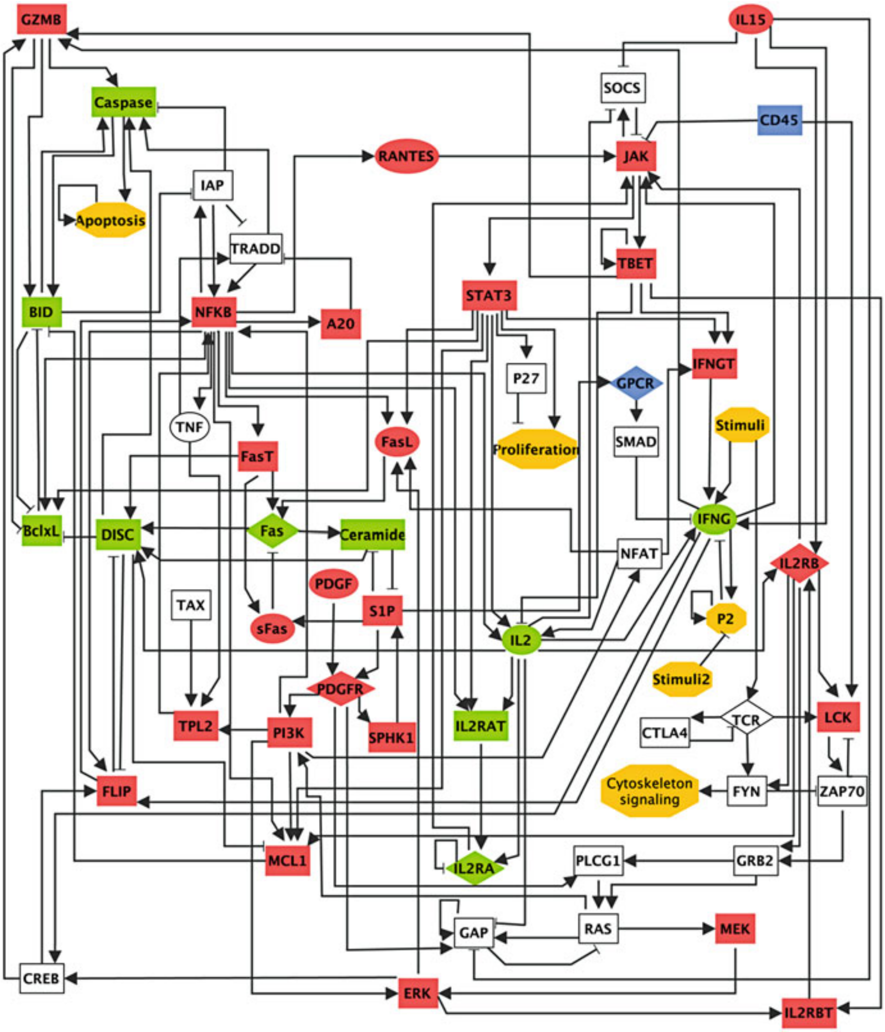


Fig. 2.2 A signal transduction network involved in activation-induced cell death of white blood cells called cytotoxic T cells. The key signals are Stimuli (representing the presence of pathogens) together with the external molecules interleukin 15 (IL15) and platelet derived growth factor (PDGF). These signals correspond to source nodes, which only have outgoing edges. The key output node of the network is Apoptosis, expressing programmed cell death. Nodes that, like Apoptosis, have no outgoing edges are called sink nodes. The shape of the nodes indicates their cellular location: rectangles indicate intracellular components, ellipses indicate extracellular components, and diamonds indicate receptors. Conceptual nodes are represented by yellow hexagons. The color of the nodes indicates the known status of these nodes in abnormally surviving T-LGL cells as compared to normal T cells: red indicates abnormally high expression or activity, green means abnormally low expression or activity, and blue indicates inconclusive or contradictory evidence. An arrow-head or a short perpendicular bar at the end of an edge indicates activation or inhibition, respectively. Details about the name of the nodes can be found in [53, 72]. Figure is reproduced from [53]

of a real signal transduction network, describing the activation-induced cell death of white blood cells called cytotoxic T cells [53, 72]. This network was used to study the disruption of activation-induced cell death in the disease T-LGL leukemia, causing the survival of a fraction of activated T cells, which later start attacking healthy cells. The network has 60 nodes and 142 edges. In Fig. 2.2, cellular location is indicated by the shape of the node: rectangles indicate intracellular components, ellipses indicate extracellular components, and diamonds indicate receptors. In addition, hexagonal nodes are conceptual nodes used to summarize connections with other signal transduction mechanisms or cell behaviors [53, 72].

2.2.2 Network Topological Properties

The totality of the nodes and edges of a network is referred to as the *network structure* or *network topology*. The structural (topological) analysis enables us to trace the propagation of information in the network and determine the key mediators. This initial analysis invokes graph theoretical measures, such as centrality measures, shortest paths, and network motifs, to describe the organization of the network [2, 10, 48].

Centrality measures were introduced to describe the importance of individual nodes in the network. The simplest centrality measure is the node *degree*, which is the number of edges connected to the node. For directed networks, the *in-* and *out-degree* of a node is defined as the number of edges coming into or going out of the node, respectively [2, 10, 48]. For example, in the T-LGL leukemia network shown in Fig. 2.2, node CREB (bottom left corner) has in-degree 2 and out-degree 2. In some molecular networks, especially signal transduction networks, it is possible that nodes have an auto-regulatory *loop*, an edge that both starts and ends at the same node. This loop usually represents a stabilizing, or on the contrary, destabilizing, self-influence. For example, the conceptual node Apoptosis has a self-loop, indicating that after commitment to apoptosis (programmed cell death) the process is self-sustaining.

In directed networks, nodes with in- or out-degree of zero are given special names. The nodes with only outgoing edges (with the potential exception of loops) are called *sources*, and nodes with only incoming edges (again, with the potential exception of loops) are *sinks* of the network. In signal transduction networks, source nodes generally correspond to external signals, while sink nodes denote responses or outcomes of the process [4]. For example, in Fig. 2.2, the nodes Stimuli, IL15, and PDGF are source nodes and have no incoming edges, and indeed they represent external signals acting on T cells. Proliferation, Cytoskeleton signaling, and Apoptosis are sink nodes and have no outgoing edges except the loop of Apoptosis, and indeed they represent outcomes of the signal transduction process: the increase in the number of cells due to cell growth and division, the reorganization of the cytoskeleton necessary for movement, and the genetically determined process of cell destruction [53, 72].

Statistical quantities, such as the degree distribution, can be formed to summarize the information of all nodes in the network [2, 10, 48]. The node degree distribution $P(k)$ is a function that, for each degree k, gives the fraction of nodes that have k edges. Similarly, we can define an in-degree and out-degree distribution for directed networks. The degree distribution reveals a lot of information about the structure of the network. For example, in a random network, where the probability of having an edge between each pair of nodes is the same, the node degree distribution will be close to a binomial distribution [2, 10, 48]. However, a variety of molecular networks have a degree distribution that follows a power law, for example, the metabolites in the *E. coli* metabolic network have an in-degree distribution $P(k) \sim k^{\gamma_{in}}$, where $\gamma_{in} = 2.2$ [29]. The heterogeneity encompassed in this so-called scale-free degree distribution has a significant impact on the network's dynamical properties, such as its controllability and stability with respect to perturbation [7, 42, 71].

The nodes whose degree is in the top 1–5% of the nodes are termed *hubs* [10, 48]. These hub nodes often play an important role in the network. For example, the node representing the NFκB protein has an out-degree of 11 and an in-degree of 4, and is a hub of the T-LGL network in Fig. 2.2. This is expected since NfκB is a transcription factor that is known to be important in cellular responses to various stimuli and in cell survival [21].

A *path* exists between two nodes if there is a sequence of adjacent edges connecting them. In directed networks, the adjacency needs to be directional as well [10, 48]. Thus in a directed network the existence of a path from A to B does not imply that a path from B to A exists. For example, as shown in Fig. 2.2, there is a path from Caspase to the conceptual node Apoptosis; however, there is no path from Apoptosis to Caspase.

In networks that can have both positive and negative edges, the *sign of a path* is positive if there is no or an even number of negative edges in the path and is negative if there is an odd number of negative edges [4, 67]. For example, as shown in Fig. 2.2, the path from Stimuli2 to P2 is negative and the path from Stimuli2 to IFNG is positive since the path consists of two negative edges.

A path containing two or more edges that begins and ends at the same node is called a *circuit* or *cycle* (if it does not repeat nodes or edges). The *length* of a path or a cycle is defined to be the number of its edges (loops can be considered as cycles of length one). A directed cycle is also called a feedback loop. The sign of a cycle is defined the same way as the sign of a path. For example, as shown in Fig. 2.2, the cycle between S1P, PDGFR, and SPHK1 is a positive feedback loop, while the cycle between TCR and CTLA4 is a negative feedback loop.

An undirected network is connected if there is a path between any two nodes. A disconnected network is made up by two or more connected components (subgraphs). A directed network is *strongly connected* if for any two nodes u and v in the network, there is a directed path both from u to v and from v to u. If a network is not strongly connected, it is informative to identify *strongly connected components* of the network. Having no strongly connected components (SCCs) indicates that the network has an acyclic structure (i.e., it does not contain feedback loops), while having a large SCC implies that the network has a central core. The core can be obtained by iteratively removing source and sink nodes until no nodes can be removed from the network. A directed network is weakly connected if it is connected when we disregard the edge directions. Signaling networks tend to have a strongly connected core of considerable size[43]. For example, the network in Fig. 2.2 has a strongly connected component of 44 nodes, which represents 75% of all nodes.

We can define the *in-component* of an SCC as the nodes that can reach the SCC, and the *out-component* of an SCC as the nodes that can be reached from the SCC. In biological networks, nodes in each of these subsets tend to have a common task. In signaling networks, the nodes of the in-component represent signals or their receptors and the nodes of the out-component are usually responsible for the transcription of target genes or for phenotypic changes [43]. For example, the in-component of the T-LGL network on Fig. 2.2 includes six source nodes, while its out-component consists of three sink nodes and P27.

Another useful centrality measure is betweenness centrality. The *betweenness centrality* of node k is given by

$$g_k = \sum_{i \neq j \neq k} \frac{C_k(i, j)}{C(i, j)}, \tag{2.1}$$

where $C(i, j)$ is the number of shortest paths between node i and j and $C_k(i, j)$ is how many of these pass through node k [20]. For example, one step of the calculation of the betweenness centrality of CIA would have i as CI and j as wg, thus $C_{CIA}(CI, wg) = 1$ and $C(CI, wg) = 2$, the ratio of the two quantities is 1/2. The betweenness centrality is the sum of such ratios among all possible pairs. Betweenness centrality tends to be a better importance measure than node degree.

A network module has many inside edges but few edges going outside the module. There are several more specific definitions of modules, and many methods to identify network modules [10, 48]. One method of module detection is based on adjacent k-cliques, where a k-clique is a complete undirected network of k nodes [49]. Two k-cliques are adjacent if they share $k - 1$ nodes. The k-clique module is the union of all k-cliques that can be reached from each other through a series of adjacent k-cliques. Palla et al. applied this method to detect modules in the protein–protein interactions network of *S. cerevisiae*, and demonstrated that the proteins in the detected modules have a shared functional classification [49].

Network motifs are recurring patterns of interconnection with well-defined topologies [9]. Among these motifs are *feed-forward loops* (in which a pair of nodes is connected by both an edge or short path and a longer path) and *feedback loops* (directed cycles). For example, in the T-LGL leukemia network shown in Fig. 2.2, nodes STAT3, P27, and Proliferation form an incoherent feed-forward loop, since the two paths from STAT3 to Proliferation have different signs. Feed-forward loops are more abundant in transcriptional regulatory and signaling networks of different organisms compared to randomized networks that keep each node's degree. They were found to support several functions such as filtering of noisy input signals, pulse generation, and response acceleration [9]. Positive feedback loops were found to support multi-stability, while negative feedback loops can cause pulse generation or oscillations [62].

Software packages for network visualization and analysis include yEd Graph Editor, Cytoscape [57], NetworkX [24], and Pajek [11].

2.3 Logical Modeling of the Dynamics of Biomolecular Networks

2.3.1 Introduction

Network representation and analysis provide insight into the connectivity between inputs and outputs and the importance of mediator nodes in the molecular system. However, as each node represents a specific molecular species in the biomolecular network, it also has an abundance associated with it and this abundance can change in time. Thus we need a second, dynamic layer in addition to the static network representation to model cell behavior. We assign each node a variable x_i to represent its state or abundance. The value of this state variable (or, simply said, the state of the node) will depend on the state of the node's regulators (which are specified by the network). Then the states of the nodes (or of a subset) can be used to represent a certain cell function or behavior [4, 67]. For example, in the T-LGL network a high value for the state variable of Apoptosis indicates that the cell committed to the cell death process, and a zero or low value of Apoptosis, coupled with abnormal values of other nodes (shown as node colors in Fig. 2.2), indicates the abnormal survival state of leukemic cells.

Dynamical models can be classified into continuous or discrete depending on whether the state variables are continuous or discrete. In continuous dynamical models, the rate of change (time derivative) of each node state x_i is expressed as a function of other variables in the molecular network. Thus the regulatory relationships are described by a system of ordinary differential equations [32, 35]. Continuous models are optimal for well characterized systems, where the mechanistic details for each interaction, the regulatory functions' form, and their parameters' values are well known through collecting a sufficient amount of quantitative information (usually through decades of experimental work). However, this is usually not the case in biomolecular systems involving large numbers of heterogeneous molecules; in most cases, not all interactions have been established, the underlying mechanisms are not sufficiently known and the kinetic parameter values are difficult to measure or estimate. Thus continuous modeling is not a good fit for these types of systems.

Discrete dynamical models use discrete variables to represent logic categories of node abundance and describe the future state of each node as a function of the states of its regulators in the biomolecular network. The discrete models only require qualitative or relative measurements, use no or very few kinetic parameters, and yet can provide a qualitative dynamic description of the system [67]. Also, there is increasing evidence that the responses to signals in molecular networks (the so-called *dose–response curves*) show sigmoidal functional forms, which provides a rationale to describe the responses with discrete variables. For example, the MAP kinase cascade has sigmoidal regulatory functions at each level, and overall leads to a step-like input–output relationship [63]. Certain network motifs show parameter-independent input–output characteristics or outcomes that are robust to changes in parameter values [63, 65]. Taken together, this evidence makes it possible for us to use discrete models to capture the characteristics of real molecular systems. These discrete dynamical models including Boolean network models [33], multi-valued logical models [8], and Petri nets [15] have been employed to study various systems in unicellular organisms, plants, animals, and humans [3, 14, 39, 53, 55, 56, 58, 61, 72].

Choosing the right dynamical model involves finding a balance between modeling detail and scalability. The hypothesis behind discrete dynamical models is that for certain classes of systems, the kinetic details of individual interactions are less important than the organization of the regulatory network [3, 33, 63]. Boolean networks are the simplest discrete dynamic models. In the following, we introduce the definitions of a Boolean network, sketch the steps in constructing a Boolean model of a molecular network, and discuss several obstacles and possible solutions.

In a Boolean network, each node state x_i is a *binary variable*, either 0 or 1. The value $x_i = 1$ (ON) represents that the node (i.e., gene, protein, or molecule) is active or expressed or is above a certain concentration threshold, while the value $x_i = 0$ represents that the node is inactive, not expressed, or is below a certain concentration threshold. The threshold may not need to be specified as long as it is clear that such threshold exists, above which the component will effectively regulate the downstream nodes. The state of the entire system will be represented as a vector $(x_1, \ldots, x_N)$. The regulation relationships are described by the governing equations $x_i^* = f_i$, which means that the future node state x_i^* is determined by the Boolean regulatory function f_i (also called *Boolean rule*) of its regulators. There are two ways to specify the Boolean function. The first intuitive way is to write it in terms of the logic operators AND, OR, and NOT. For example, $x_4^* = f_4 = (x_1 \text{ OR } x_2) \text{ AND } (\text{NOT } x_3)$ means that x_4 will be ON when x_3 is OFF and simultaneously at least one of x_1 or x_2 is ON. The implicit order of precedence of logical operators may be used: NOT has higher precedence than AND, and AND has higher precedence than OR. Thus the above Boolean rule can also be written as $f_4 = (x_1 \text{ OR } x_2) \text{ AND NOT } x_3$, but it is different from $f_4 = x_1 \text{ OR } x_2 \text{ AND NOT } x_3$. The second way to express a Boolean function is through a truth table, where we specify the output value for each possible input configuration. If node x_i has k regulators, we will have 2^k input configurations since each regulator has two possible states. For example, the three basic logic operators, NOT (third column), OR (fourth column), and AND (last column), can be written as shown in Table 2.1.

Table 2.1 Truth tables illustrating the NOT, OR, and AND operators

x_A	x_B	f_C = NOT x_A	f_D = x_A OR x_B	f_E = x_A AND x_B
0	0	1	0	0
0	1	1	1	0
1	0	0	1	0
1	1	0	1	1

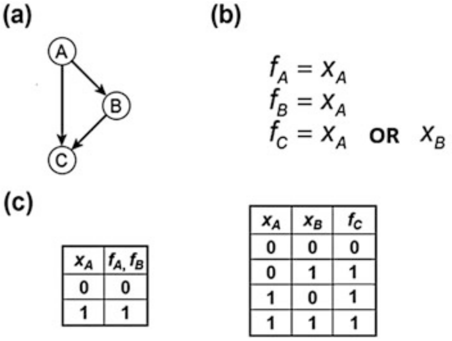
The first two columns indicate all the possible configurations for the two input nodes A and B. The third to fifth columns give the output value of the corresponding input configuration in the same row for the three functions NOT x_A, x_A OR x_B, x_A AND x_B, respectively

2.3.2 *Procedures to Construct Boolean Dynamic Models*

We first outline the whole procedure to develop a Boolean model of a biomolecular network, then give the details in the following paragraphs. One starts to build the Boolean model by establishing the list of nodes and of the known interactions and regulatory relationships among these nodes. One then needs to determine the Boolean regulatory function of each node. One also needs to determine the relevant initial conditions and choose an updating scheme to model the passing of time. Model construction is followed by model analysis, including determining the long-term behavior of the model. The model's results need to be compared with established experimental results. If there are discrepancies, one needs to iteratively revise the Boolean model, including the network topology or the Boolean regulatory functions until the model is consistent with known behavior. Then one can use the Boolean model to make novel predictions awaiting experimental confirmation.

The first step in constructing the Boolean network is to collect information about the network nodes and interactions. The modeler needs to integrate and assemble information from several experiments, for example, high-throughput gene expression, proteomics and metabolomics data, or detailed studies of individual interactions [52, 67]. High-throughput phosphoproteomics, protein–DNA interaction, and genetic interaction studies can be used for two purposes: to determine the meaning of the binary states of components in known conditions (in a comparative manner or by using a threshold) or to infer casual relationships between components. These casual relationships can be represented by a directed edge from one node to another in the network. Often the sign of the edge, positive (activating) or negative (inhibitory), can also be inferred. We can construct the molecular network if the totality of relevant information is sufficient [52, 67]. The readers interested in how to deal with incomplete information can refer to [5, 31, 40].

The next step is to determine the Boolean regulatory function for each node. When there are multiple regulators for a node, we select the function that best represents the existing knowledge about their action. The OR function is used if the node can be activated by any of its regulators. The AND function is used if the node needs all of its regulators to be activated. If the Boolean regulatory function involving several regulators cannot be fully determined, one needs to take a trial and



x_A	f_A, f_B
0	0
1	1

x_A	x_B	f_C
0	0	0
0	1	1
1	0	1
1	1	1

Fig. 2.3 A Boolean model of a simple signal transduction network. (**a**) The graphical representation of the network. The edges with arrow-heads represent positive regulation. Note that the Boolean regulatory function for node C is not uniquely determined by the network representation. (**b**) The Boolean regulatory functions for each node in the model. (**c**) The truth tables of the Boolean regulatory functions given in (**b**)

error approach to select the function that can successfully reproduce the existing experimental results (both at the node and at the whole network level).

For example, let us determine a compatible Boolean regulatory function for the three-node feed-forward motif shown in Fig. 2.3. A natural choice for source node A is $x_A^* = f_A = x_A$ as it represents that the signal of the system maintains a certain state for a certain period of time. As A positively regulates B, $x_B^* = f_B = x_A$. Node A and B positively regulate C. Then there are two compatible choices for the Boolean regulatory function of node C: $f_C = x_A$ OR x_B, and $f_C = x_A$ AND x_B. The results of knockout experiments (wherein one node is set into the OFF state) can help us determine which one is more appropriate. Let's assume that providing A and simultaneously knocking out B resulted in the activation of C. This means that A alone can activate C, and thus $f_C = x_A$ OR x_B.

The next step is to determine the relevant initial condition for the system, e.g., the system's natural resting state. When the relevant initial condition is not known, one can sample from different initial conditions in the state space. We note that the biologically relevant initial conditions may occupy a small region in the state space.

One also needs to choose a time implementation or *updating regime* for the system to evolve. Time is often implemented as a discrete variable, that is, the node states are updated at fixed time steps and their values are kept the same between time steps [52, 67]. The timescale of the processes represented as edges can vary from fractions of a second to several hours depending on the biological process [9]. Mathematically, we use the vector $(x_1(t), \ldots, x_n(t))$ to represent the state of the system at time t. Then we determine the value of each node state $x_i(t + \tau_i)$ in the next time step based on the Boolean regulatory function, that is, $x_i(t + \tau_i) = f_i(x_{k_1}(t), \ldots, x_{k_i}(t))$, where the τ_i is the time step for node i and $k_1, \ldots, k_i$ are the regulators of node i.

As we need to update all the nodes to obtain the system's trajectory, we also need to specify the order of updating each node. The simplest updating regime is synchronous updating, wherein all the nodes are updated simultaneously. This is equivalent to setting $\tau_1 = \cdots = \tau_n$ as a time step [52, 67]. Thus the synchronous updating regime implicitly assumes that the timescales of all biological events are approximately the same so that the state change of each node is synchronized. In biological systems that include biological events of different timescales (e.g., include both transcriptional and post-translational regulation) it is not appropriate to use synchronous updating.

In order to take into account variations in timescale, different asynchronous updating regimes were developed. In deterministic asynchronous updating, a fixed timescale or time delay is used for each node. In the stochastic asynchronous regime, the system is updated in a random way. To be specific, in random order asynchronous update, a random permutation of a sequence of all the nodes is generated for each round and the nodes are updated in the order of the simulated permutation; this process is repeated until convergence [52, 67]. Thus, in this regime, every node will be updated once during each round.

Another popular stochastic asynchronous update method is general asynchronous update, where a randomly selected node is updated in each time step [52, 67]. Thus, in contrast with random order asynchronous regime, it is possible that one node is updated several times before another node gets updated next. However, since the node is randomly selected, the expected number of updates is the same for all nodes. If we know that nodes should be updated with different frequencies, we can use an update probability distribution.

Let us continue with the three-node motif in Fig. 2.3 to illustrate two deterministic and two stochastic updating regimes. In synchronous updating, the state transitions will be $x_A^* = x_A(t+1) = f_A(t), x_B^* = x_B(t+1) = f_B(t), x_C^* = x_C(t+1) = f_C(t)$, where the nodes' future states (at time $t+1$) are determined simultaneously, using their current node state at time t. In a deterministic asynchronous updating regime, say $\tau_A = 1$, $\tau_B = 2$, $\tau_C = 3$, the system will be updated in a pattern with period of 6: A alone, A and B together, A and C together, A and B together, A alone, A, B, and C together. For random order asynchronous updating, there are 3!=6 ways of ordering these three nodes, at each time one ordering will be randomly selected. The order of update in our example can be A, C, B; A, B, C; B, C, A; A, C, B; $\ldots$, where semicolons indicate the end of a time step. In general asynchronous update, a possible update order for the three nodes system could be $A, B, C, B, B, C, B, A \ldots$. Notice that node B has been updated four times until A was updated again in this particular realization.

After the model is completely specified, we need to determine its long-term behavior. Since the Boolean network is a finite system, the state of the system will evolve into a single state (steady state) or a set of recurring states (a complex attractor). These steady states or recurring states are collectively called *attractors* [52, 67]. Attractors of molecular networks have corresponding biological meanings. A steady state or a group of steady states with similar function can be associated

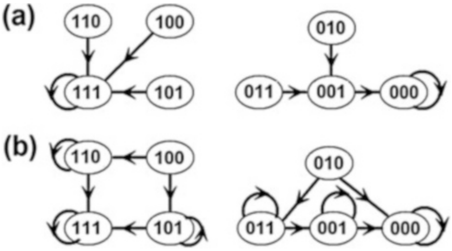


Fig. 2.4 State transition graphs of the Boolean model presented in Fig. 2.3. A node represents a state of system, written in the order A, B, C; thus, 111 represents $x_A = 1, x_B = 1, x_C = 1$. A directed edge between two states indicates a possible transition from the first state to the second in one update specified in the updating scheme. A loop (an edge that starts and ends at the same state) indicates that the state does not change during update. (**a**) The state transition graph under synchronous update. The two states that have loops are the fixed points of the system. (**b**) The state transition graph under general asynchronous update (update one random node at a time). Though several states have loops, only the two states that have no outgoing edges are fixed points of the system

with a cell state or phenotype. Complex attractors can be interpreted as cyclic or oscillatory behavior such as the cell cycle, circadian rhythms, or Ca^{2+} oscillations [1, 39, 40].

A compact visualization of all possible trajectories is given by the *state transition graph* (STG), wherein each node is a possible state of the system, and each directed edge represents a possible transition from one state to another state in one update [52, 67]. The STG will contain 2^N nodes for a Boolean network with N nodes as it contains all the possible states in the state space. For example, the state transition graph of the three node Boolean model in synchronous updating regime is shown in Fig. 2.4.

In the state transition graph, a steady state will be a node with a loop and no other outgoing edge; a complex attractor will be a strongly connected component without an outgoing edge. For example, in Fig. 2.4, the state 111 and state 000 only have a loop and no other outgoing edges, indeed they are the steady state of the three node system in Fig. 2.4. Notice that this criterion can be used to identify steady states, however, it won't be an efficient way as it requires to map the entire state transition graph first.

For each attractor, all the states that can reach the attractor in the state transition graph are called its basin of attraction. For example, in Fig. 2.4, the basin of the steady state of 111 includes state 100, 110, 101, and 111, while the basin of the steady state of 000 includes state 000, 010, 001, and 011.

It is an interesting question whether a chosen updating regime will affect the properties of attractors and of the state transition graph. Let us start with the steady state (fixed point) type of attractor. In a steady state the future state, i.e., the outcome of the Boolean regulatory function, equals the current state for each node. This requirement is time independent; therefore, steady states are independent of updating regimes. Indeed, in Fig. 2.4 the steady states of the Boolean model under the two updating regimes are the same. However, based on the example above, one can readily see that the state transition graph is different for the two different updating regimes. In synchronous updating, since all nodes are updated simultaneously, each state can only have one outgoing edge. Due to this, the complex attractor in synchronous updating regime is also called a limit cycle as the set of recurring states repeats in a fixed order [67]. Also, the basin of attraction for each attractor will be separated as one state can only follow a unique path in the STG (see Fig 2.4a). While in asynchronous updating, each node can have multiple outgoing edges due to the different updating order, as illustrated in Fig 2.4b. Thus states in the complex attractor can appear in an aperiodic manner.

Some limit cycles can only be observed under synchronous updating and any perturbation of the updating timescales will eliminate the attractor [37]. One can readily see this in the example shown in Fig. 2.5: the limit cycle between the states 001 and 010 is not observed under general asynchronous update, where two successive states can only differ in one node's state. The figure also exemplifies that the basin of attraction of the attractors may overlap due to the randomness in the updating regime. The state transition graph can be seen as a graphical representation of a corresponding Markov Chain model, where each node is a state in the Markov chain and each edge corresponds to a transition with non-zero probability between states. If complete randomness is guaranteed, the system is taking a random walk on the state transition graph, which specifies a unique Markov Chain model [67].

At the end, one needs to compare the model's results with established experimental results. If there are discrepancies, one needs to revise the Boolean network or the Boolean regulatory function [12, 18, 67]. Boolean network should qualitatively

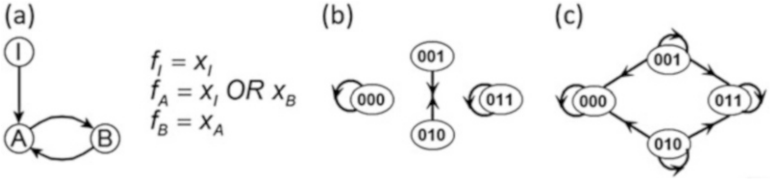


Fig. 2.5 A simple three node network. (**a**) The network representation and corresponding Boolean rules. Node A and B form a positive feedback loop. Node B and source node I can independently activate node A. (**b**) The network's (partial) state transition graph under synchronous update when the signal is set as OFF ($x_I = 0$). The states are specified in the node order I, A, B. (**c**) The state transition graph under general asynchronous update when the signal is set as OFF ($x_I = 0$). The figure is adapted from [4]

reproduce properties demonstrated in biological systems including homeostasis or multi-stability [9, 33]. In the next subsection, we use two biological examples to illustrate this point.

Dynamical models can also be used to make novel predictions, such as predictions about the effect of perturbations and about network control strategies [13, 67, 69, 70]. These predictions can provide insight about the biological system and guide future experiments. In perturbation analysis, we determine the change in the attractors induced by external or internal perturbations, including knockout or constitutive expression/activity of a node. Node knockout can be modeled as fixing the corresponding node in the OFF state, while constitutive expression/activity can be modeled as fixing the node in the ON state. Transient perturbations can be modeled as temporary changes in the node's state and letting the system evolve naturally [13]. The perturbation analysis can predict changes in the attractors and their basin induced by each possible perturbation. Thus those perturbations that lead to a dramatic cascading effect will be identified, which helps us identify components key to maintaining a phenotype in a biological system. In a signal transduction network involved in a disease, the identified key components could be targets of therapeutic interventions [40, 53, 58].

Several software tools are available for Boolean dynamic modeling of biological systems. The CoLoMoTo (Consortium for Logical Models and Tools) platform provides resources for logical modeling, including software tools and biological models [17]. Among them, SBML qual is an open-source model library, promoting a standard format to analyze and exchange qualitative models [17]. GINsim is a free Java software application for logical modeling of regulatory and signaling networks [16, 47]. It allows users to define a model or import models in various formats. It also supports simulations of logical models and generates state transition graphs under various updating regimes. The R package BoolNet provides attractor search and robustness analysis methods for synchronous, asynchronous, and probabilistic Boolean models [46]. In addition, BooleanNet is a python package that can be used to simulate synchronous and random order asynchronous models and to determine their state transition graph [6]. There are other existing simulation and analysis software tools for logical models, including ADAM [27], the Cell Collective [26], CellNetAnalyzer [36], CellNOpt [60], ChemChains [25], Odefy [38], SimBoolNet [73], and SQUAD [19].

2.3.3 Two Biological Network Examples

It has been shown that Boolean models can capture characteristic dynamic behavior, such as excitation–adaptation behavior and multi-stability, as continuous models do [63]. For example, positive feedback loops support multi-stability, coherent feed-forward loops support the filtering of noisy input signals, and incoherent feed-forward loops support excitation–adaptation behaviors [9, 62, 63]. The reader interested in the details of these examples can refer to [63].

Here we illustrate the capacity of Boolean network models to capture dynamic behavior using two biological network examples. The first one is the **T cell large granular lymphocyte leukemia (T-LGL) network**, which is mentioned in Sect. 2.2.1 and shown in Fig. 2.2. T-LGL leukemia is a rare blood cancer. While normal T cells undergo activation-induced cell death (apoptosis) after successfully fighting a virus, leukemic T-LGL cells survive. Through an extensive literature search, Zhang et al. constructed a Boolean network model of T-LGL leukemia, which can reproduce the abnormal survival of T-LGL cells and other known experimental results of the system [72]. The details of the T-LGL model, including the Boolean regulatory functions, can be found in [72]. Zhang et al. chose a stochastic asynchronous updating regime. The model has two steady states under the relevant source node initial condition (Stimuli, IL15, and PDGF are ON and Stimuli2, CD45, and TAX are OFF) [72]. The two steady states, respectively, correspond to the apoptosis of T cells and survival of the abnormal T cells as seen in T-LGL leukemia. This exemplifies that the Boolean model successfully reproduces the qualitative experimental results and captures the multi-stability of a real system. Full analysis of the state space was not possible due to the large size of the network; thus, follow-up work employed network simplifications to reduce the network size and the state space. Two kinds of network reductions were applied, both of which have been shown to preserve the attractor repertoire of the system [54]. First, one can determine and eliminate the nodes whose state stabilizes due to their regulation by sustained signals. Second, one can iteratively collapse nodes with one incoming and one outgoing edge, for example, node MCL1 could be removed in Fig. 2.6a. One can obtain a reduced network with 18 nodes after applying the first kind of reduction and a reduced network with six nodes after both reductions. Now it is much easier to visualize the state space of the T-LGL leukemia network, which is shown in Fig. 2.7. Perturbation analysis of the Boolean model in Fig. 2.6b reveals that permanently reversing the node state of S1P, Ceramide, or DISC in the T-LGL leukemia steady state can eliminate the T-LGL steady state and lead to apoptosis [53]. Nodes such as S1P, Ceramide, or DISC can be called key mediators of the T-LGL state. These key mediators are candidate therapeutic targets, which is supported by experiments (one of which was performed to test this prediction). Similar analysis of the original 60-node Boolean model in Fig. 2.2 identifies 15 key mediators in the original network, which are also candidate therapeutic targets [53].

The second example is the **epithelial-to-mesenchymal transition (EMT) network**. EMT is a cell fate change, during which epithelial cells lose their original adhesive property, leave their primary site, invade neighboring tissue, and migrate to distant sites as mesenchymal cells [58]. EMT plays an important role in pathological

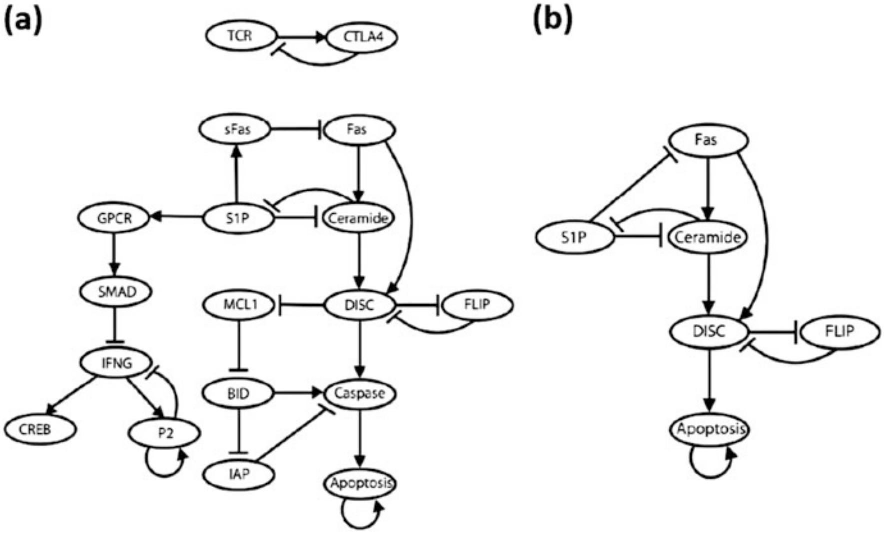


Fig. 2.6 Reduced T-LGL leukemia signaling network. An arrow-head indicates a positive edge, and a blunt segment indicates a negative edge. (**a**) The 18-node network obtained by removing stabilized nodes due to the sustained state of source nodes. (**b**) The 6-node subnetwork obtained by merging mediator nodes from the bottom subgraph in part A. This figure is reproduced from [53]

processes, including the invasion process in hepatocellular carcinoma (HCC); thus, it is important to understand this signaling process and design strategies to suppress it. A hallmark of EMT is the loss of E-cadherin, a cell adhesion protein. EMT can be induced by transforming growth factor-β (TGFβ), growth factors, and other external signals [58]. Through extensive literature search, Steinway et al. built a Boolean network model of EMT in the context of HCC invasion [58]. This network contains 69 nodes and 134 edges. In this network, E-cadherin is the sole negative regulator of the sink node, which is a conceptual node to represent the occurrence of EMT. The model is updated in a ranked asynchronous updating regime to account for the fact that the relevant signal transduction events occur substantially faster than the involved transcriptional events [58]. Simulations of the Boolean model can reproduce the EMT driven by TGFβ: starting from an epithelial state (which is an attractor of the signal-free system), and activating the TGFβ signal, the system will evolve and finally stabilize into a mesenchymal state [58]. With TGFβ fixed to be ON, the model can be reduced to a network with 19 nodes and 70 edges after applying similar network reduction techniques as in the T-LGL leukemia network. The mesenchymal state is the only steady state of the reduced network, which is confirmed by exploration of the state space of the reduced network. This

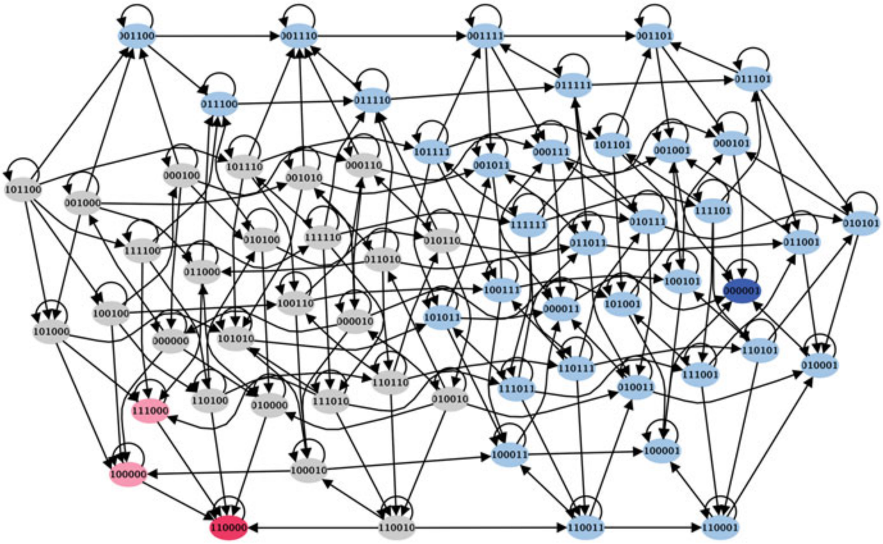


Fig. 2.7 The state transition graph of the reduced 6-node subnetwork of T-LGL leukemia network shown in Fig. 2.6b. There are 64 possible states in the state space. The dark blue node represents the normal steady state (apoptosis of T cells) and the red node represents the T-LGL leukemia steady state. The light blue states are transient states that will evolve into the normal steady state (dark blue) and the pink states are transient states that will evolve into the leukemia steady state (red). Gray states are transient states that can evolve into either steady state. This figure is reproduced from [53]

suggests that the system will ultimately adopt a mesenchymal state in response to TGFβ. A systematic search revealed that there are seven nodes, whose individual knockout can prevent TGFβ driven EMT (i.e., inhibit the transition from the epithelial state into the mesenchymal state in response to TGFβ). These seven nodes are all transcription factors that directly regulate E-cadherin. The effectiveness of the knockout of these transcription factors was already established experimentally, but unfortunately currently it is not possible to target these transcription factors by drugs. There are also six combinations of two node knockouts (not involving any of the previous seven nodes) that can suppress TGFβ driven EMT. All these six knockout pairs require the inhibition of the SMAD complex. If constitutive activation is also considered, one new single-node intervention target, miR200, and one new two-node combination are identified [58] (Fig. 2.8).

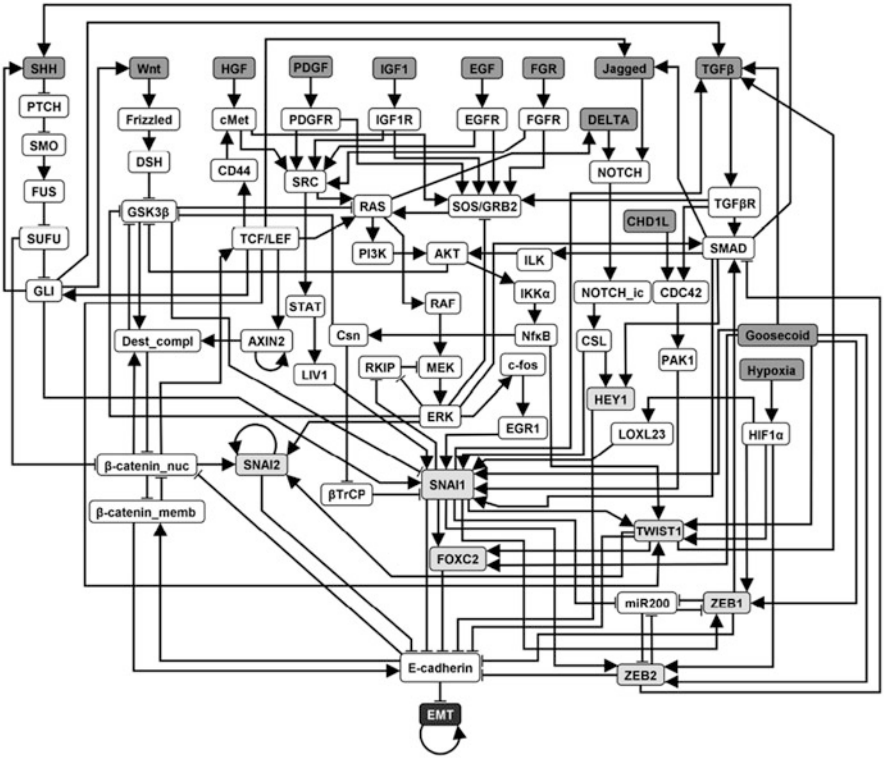


Fig. 2.8 The EMT signaling network in HCC, which consists of 69 nodes and 134 edges. Signals are in dark gray fill, transcriptional regulators of E-cadherin are in light gray fill. The output node EMT is marked with black background. Positive edges are drawn with arrow-heads and negative edges terminate in blunt segments. The figure is reproduced from [58]

2.4 Connecting the Structure and Dynamics of Molecular Networks

There is increasing evidence that the dynamics of certain systems is not sensitive to the details of the interactions and to the kinetic parameter values, which inspired researchers to explore the effect of the underlying network topology on the network dynamics [9, 33, 63]. Multiple lines of research have been devoted to shed light upon this subject; we will introduce several tools developed to analyze this relationship in this section [59, 66, 69, 70].

We first discuss network structural features that influence the attractor repertoire of Boolean models. As hypothesized and later verified by researchers, feedback loops play an important role in determining the network attractors [62]. René Thomas conjectured that a necessary condition for a system to have multi-stability is the existence of a positive feedback loop and a necessary condition for a system

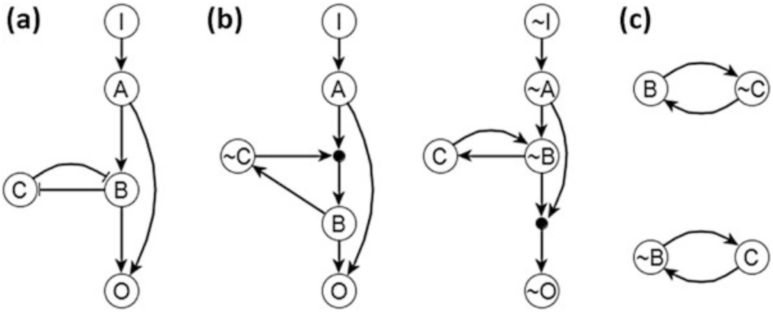


Fig. 2.9 Illustration of the expanded network of a simple network. (**a**) A hypothetical signal transduction network similar to the reduced 6-node T-LGL leukemia network in Fig. 6b. (**b**) The expanded network of the given network in (**a**). The composite node is denoted by a solid circle. (**c**) The stable motif of the given network under a sustained signal input $x_I = 1$. The figure is adapted from [4]

to have sustained oscillations is the existence of a negative feedback loop [62]. This indicates that the sign of the cycles in the network determines the dynamical behavior of a system. An additional important feature, which is not explicitly represented by the interaction network, is the possible dependence or combinatorial effect of multiple incoming edges to a node. This motivates us to integrate a network representation with the Boolean regulatory functions of each node into a so-called expanded network [66].

We introduce the concept of expanded network with the example in Fig. 2.9, which consists of five nodes, the input I, and the regulated nodes O, A, B, and C with the regulatory functions $f_A = x_I$, $f_B = x_A$ AND (NOT x_C), $f_C =$ NOT x_B, $f_O = x_A$ OR x_B. First, we introduce a complementary node for each original node in the system to represent the negation (deactivation) of the original node, denoted by the real node's name preceded with $\sim$. In all the Boolean regulatory functions, all the NOT functions are replaced by the negated state of the respective node (i.e., its complementary node) since the NOT function is a unary operator. The edges in the expanded network are redistributed according to the updated rules. For example, $f_C =$ NOT $x_B = x_{\sim B}$ and thus a corresponding edge is drawn from $\sim B$ to C in the expanded network. The Boolean regulatory function of a complementary (negated) node is the logical negation of the regulatory function of the original node. For example, $f_{\sim C} =$ NOT (NOT $x_B) = x_B$ and thus a corresponding edge is drawn from C to $\sim B$ in the expanded network. Thus the Boolean rules for all the complementary nodes in Fig. 2.9 are

$$f_{\sim A} = \text{ NOT } x_I = x_{\sim I},$$

$$f_{\sim B} = (\text{NOT } x_A) \text{ OR } x_C = x_{\sim A} OR\ x_C,$$

$$f_{\sim C} = x_B,$$

$$f_{\sim O} = (\text{NOT } x_A) \text{ AND } (\text{NOT } x_B) = x_{\sim A} \text{ AND } x_{\sim B}.$$

Second, to differentiate AND rules from OR rules when considering the relationship of edges pointing toward the same target node, we introduce a composite node for each set of edges that are linked by an AND function. In order to uniquely determine the edges in the expanded network, the regulatory functions need to be specified in disjunctive normal format, that is, a disjunction of conjunctive clauses, in other words, grouped AND clauses separated by OR clauses. For example, (A AND B) OR (A AND C) is in a disjunctive normal form, while A AND (B OR C) is not. Algorithmically, the desired disjunctive normal form can be formed by a disjunction of all conditions that give output one in the Boolean table and then simplified to the disjunction of prime implicants (Blake canonical form) by the Quine–McCluskey algorithm [45]. Now we add a composite node for each AND clause in the Boolean regulatory function, denoted by a solid black node in Fig. 2.9. For example, the composite node in the left part of Fig. 2.9b represents the expression x_A AND (NOT x_C), which activates node B; the composite node in the right part of Fig. 2.9b represents the expression (NOT x_A) AND (NOT x_B), which induces the complementary node $\sim O$. Notice that one can read all the regulatory functions from the topology of the expanded network. The AND rule is indicated by a composite node with multiple regulators, while all the other edges represent independent activation (parts of an OR function).

As the expanded network contains the essential information that determines the network dynamics, the expanded network serves as a basis for network reduction and attractor analysis, i.e., the dynamical information. One approach is through analyzing the stable motifs of the expanded network [69]. A stable motif is defined as the smallest strongly connected component (SCC) satisfying the following two properties: (1) The SCC cannot contain both a node and its complementary node and (2) If the SCC contains a composite node, it must also contain all of its input nodes [69]. The first requirement guarantees that the SCC does not contain any conflict in node states and the second requirement guarantees that all the conditional dependence is satisfied and the SCC is self-sufficient in activating each node state inside the stable motif. Thus the stable motif represents a group of nodes that can sustain their states irrespective of other outside nodes' states. The corresponding node states implied by the stable motif can be directly read out: the original node represents the ON (1) state and the complementary node represents the OFF (0) state [69]. For example, in the top stable motif of Fig. 2.9c, the stable motif represents that B is ON and C is OFF.

Once we find a stable motif, we can plug in these node states into the Boolean regulatory functions and obtain a simplified network corresponding to this stable motif. We can identify all the stable motifs of this simplified network and repeat the process. The results of this iterative process can be represented as a stable motif succession diagram [70]. For example, the stable motif succession diagram of the T-LGL network is shown in Fig. 2.10 [70]. After iterative identification of stable motifs and network reduction, we will obtain one of the two final outcomes: all the nodes will be in a fixed state (either in a stable motif or fixed during network reduction) or some nodes are not in a fixed state and will be expected to have oscillatory behavior. In the first scenario, we obtained a fixed point (steady state). In the second scenario,

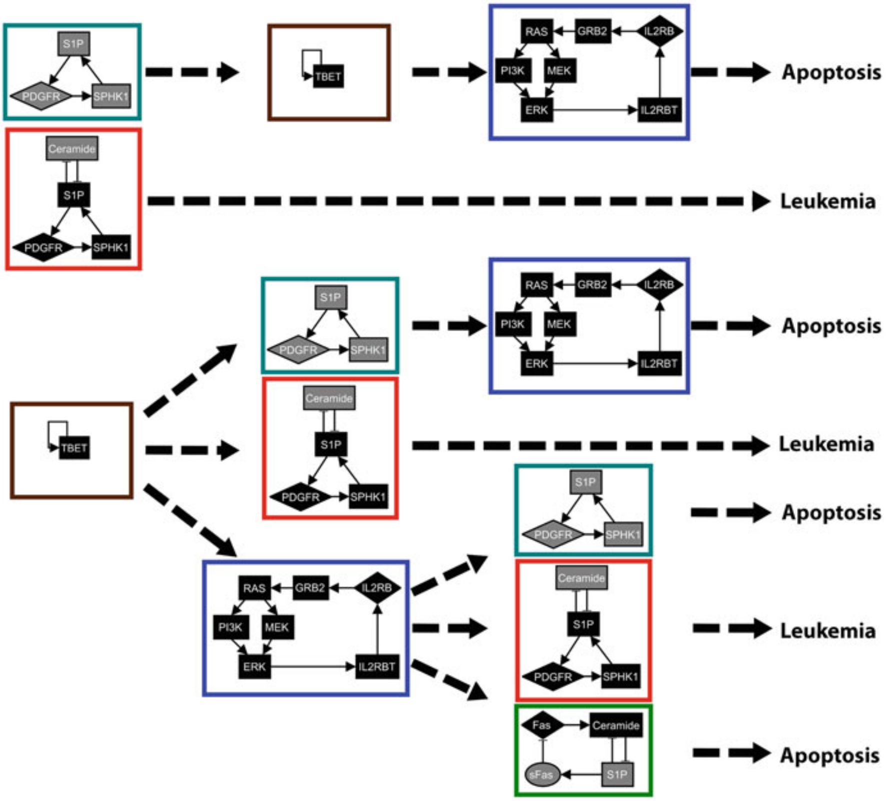


Fig. 2.10 Stable motif succession diagram for the T-LGL leukemia network. Each colored rectangle represents a different stable motif. Inside the box, gray shaded nodes indicate nodes in the ON state and black shaded nodes indicate nodes in the OFF state. There are two possible steady-state attractors: the normal state of cell death (apoptosis) and the diseased state (T-LGL leukemia). The attractor to which the sequence of stable motifs leads is marked at the rightmost. A dashed line pointing from a stable motif to a second stable motif means that the second stable motif can be found in the reduced network due to stabilization of the first stable motif. A dashed line pointing from a stable motif to an attractor means that applying network reduction with the fixed stable motif will lead to the attractor. The figure is reproduced from [70]

we obtained a quasi-attractor, which tells us the fixed node states and potential oscillatory nodes (which visit both of their states as part of a complex attractor)[70]. Thus stable motif analysis can be used as a preliminary analysis or substitute for attractor analysis depending on the level of detail we care about. For example, in the T-LGL network, successive stabilization of stable motif shown in Fig. 2.10 will ultimately drive the system to one of the two steady states: the apoptosis steady state or the T-LGL leukemia steady state [70].

We are not only interested in building the molecular network to understand the underlying biological process, but also in designing interventions or therapeutic strategies to drive the system from an initial state to a desired state or attractor.

The stable motif succession diagram readily implies a control strategy called stable motif control as the sequential stabilization of each stable motif in the stable motif succession diagram guarantees that the system will reach the desired attractor. We can control the system by controlling the corresponding stable motifs [70]. For example, sequential stabilization of the three motifs in the first line in Fig. 2.10 will drive the system to the normal steady state (apoptosis). However, the control strategy does not need control of all the nodes involved as two types of reductions can be done [70]. First, not all stable motifs need to be controlled. If there is a branch-free line of stable motifs after a particular stable motif, or if all the branches lead to the same steady state in the succession diagram, then the stable motifs after this particular stable motif do not need to be controlled. For example, in the first sequence of stable motif in Fig. 2.10, one only need to control the first, cyan-colored stable motif. Second, not all the nodes in the stable motif need to be controlled in order to stabilize the stable motif. For example, forcing S1P in the OFF state is enough to stabilize the cyan stable motif in Fig. 2.10. Thus after these two levels of reduction in the control strategy, one would get a smaller set of nodes to drive the system to the desired state; however, the intervention does not guarantee to be minimum in size. The readers interested in more mathematical or practical details can refer to [69, 70]. All these stable motif analysis and control strategies have also been applied to the EMT network, yielding strategies to prevent the system's convergence to the mesenchymal state and to return the system to the epithelial state [69, 70].

Another interesting scenario is when there is network damage (malfunction of a specific node) that could potentially lead to a cascading effect. We need to design a strategy to prevent the damage from propagating [13, 68]. There are various settings for modeling network damage; here, we consider the situation that the damage is permanent and can be modeled as forcing the node state to be ON (constitutive expression/activity) or OFF (node knockout). We assume that the damage happens after the system is in one of its attractors. The goal is to design immediate solutions to stabilize the system to be as close as possible to the original attractor except the damaged node. This can be accomplished by adding edges between nodes and modifying the corresponding Boolean regulatory functions [13, 68]. To determine the best placement of the repairs, one needs to first identify sensitive nodes, whose node state will be affected by the network damage in the first time step. Then we identify candidate nodes for the stabilization of each sensitive node by adding an edge from the candidate node to the sensitive node. This strategy can be applied to stabilize damage to multiple nodes and to stabilize multiple steady states under single node damage. These strategies have been demonstrated in two biological networks, namely T-LGL leukemia network and EMT network. More details can be found in [68].

Another aspect of dynamical information about the network is to determine the contribution of each node into the system's outcomes. Consider a network with a single source node (signal) and a single sink node that reflects the network's output. The expanded network serves as an important tool to characterize the importance of intermediary (non-signal, non-output) nodes. One useful concept is called elementary signaling mode (ESM), which is defined as the minimal set of components able to perform signal transduction (i.e., able to functionally connect the signal to the output node) regardless of the rest of the network [66]. The elementary signaling mode will be a path or a subgraph connecting from the signal to the output node, which can be identified in the expanded network. For example, there are two ESMs between node I and node O: the path I, A, O and the subgraph consisting of I, A, the composite node, $\sim C$, B, and O. One can show that they are both minimal, as taking a node from the ESM will obstruct the signal from propagating. The elementary signaling modes can be used to rank the importance of the nodes in mediating the signal through studying the reduction in the number of ESMs due to the loss of the node (and of any other nodes that are lost as a consequence) [66]. For example, node A appears in both ESMs found in Fig. 2.9, and its loss eliminated both; however, node B only appears in one of the ESMs and its loss does not affect the other ESM. This suggests that node A is essential in the signal transduction process from node I to node O, while node B is not. In several examples of biological networks it was shown that the ESM-based analysis can identify essential nodes as effectively as a full dynamical analysis of the corresponding perturbed system [66]. ESMs, or more specifically, the number of node-independent ESMs, can also be used to quantify the system's functional redundancy [59].

2.5 Conclusions

The improvements in experimental technology and the large amounts of generated data have brought us into an era where different types of dynamical models are needed to provide system-wide insights in biological systems. Although Boolean models are based on a series of assumptions and are limited in describing the quantitative features of dynamic systems, we have shown that they can capture emergent characteristics of real biological systems, demonstrate considerable dynamic richness, and can predict successful intervention strategies in biological systems. Boolean network models do not require detailed knowledge of the kinetic parameters (as continuous models do), striking a balance between scale and realism. Their parsimonious nature makes them a preferred choice for systems where detailed quantitative experimental data is not available. Qualitative dynamical models, including Boolean network models, exist as a complement to quantitative dynamical models and will be often needed as we gradually develop our understanding of biological systems. The success of Boolean networks also indicates that in certain systems the behavior of the system is largely determined by the organization of the network structure rather than the kinetic details of individual interactions, which highlight the theoretical value of Boolean network models. In summary, Boolean networks serve as a useful foundation for modeling molecular systems; they can identify the network features (e.g., stable motifs) that are key determinants of the dynamics and whose detailed modeling would be most fruitful.

References

1. O.E. Akman, S. Watterson, A. Parton, N. Binns, A.J. Millar, P. Ghazal, Digital clocks: simple Boolean models can quantitatively describe circadian systems. J. R. Soc. Interface **9**(74), 2365–2382 (2012)
2. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47–97 (2002)
3. R. Albert, H.G. Othmer, The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. J. Theor. Biol. **223**(1), 1–18 (July 2003)
4. R. Albert, R. Robeva, Chapter 4-signaling networks: asynchronous Boolean models, in *Algebraic and Discrete Mathematical Methods for Modern Biology*, ed. by R.S. Robeva (Academic Press, Boston, 2015), pp. 65–91
5. R. Albert, B. DasGupta, R. Dondi, S. Kachalo, E. Sontag, A. Zelikovsky, K. Westbrooks, A novel method for signal transduction network inference from indirect experimental evidence. J. Comput. Biol. **14**(7), 927–949 (2007)
6. I. Albert, J. Thakar, S. Li, R. Zhang, R. Albert, Boolean network simulations for life scientists. Source Code Biol. Med. **3**, 16 (2008)
7. M. Aldana, S. Coppersmith, L.P. Kadanoff, Perspectives and problems in nonlinear science: a celebratory volume in honor of Lawrence Sirovich, in *Chapter Boolean Dynamics with Random Couplings* (2003), pp. 23–89
8. B.B. Aldridge, J. Saez-Rodriguez, J.L. Muhlich, P.K. Sorger, D.A. Lauffenburger, Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/Insulin-induced signaling. PLoS Comput. Biol. **5**(4), 1–13 (2009)
9. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 1st edn. (Chapman and Hall/CRC, Boca Raton, July 2006)
10. A.-L. Barabási, M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, 2016)
11. V. Batagelj, A. Mrvar, *Pajek—Analysis and Visualization of Large Networks* (Springer, Berlin, 2002), pp. 477–478
12. S. Bornholdt, Boolean network models of cellular regulation: prospects and limitations. J. R. Soc. Interface **5**(Suppl 1), S85–S94 (2008)
13. C. Campbell, R. Albert, Stabilization of perturbed Boolean network attractors through compensatory interactions. BMC Syst. Biol. **8**(1), 53 (2014)
14. B.J. Campbell, L. Yu, J.F. Heidelberg, D.L. Kirchman, Activity of abundant and rare bacteria in a coastal ocean. Proc. Natl. Acad. Sci. **108**(31), 12776–12781 (2011)
15. C. Chaouiya, Petri net modelling of biological networks. Brief. Bioinform. **8**(4), 210 (2007)
16. C. Chaouiya, A. Naldi, D. Thieffry, Logical modelling of gene regulatory networks with GINsim. Methods Mol. Biol. **804**, 463–79 (2012)
17. C. Chaouiya, D. Bérenguier, S.M. Keating, A. Naldi, M.P. van Iersel, N. Rodriguez, A. Dräger, F. Büchel, T. Cokelaer, B. Kowal, B. Wicks, E. Gonçalves, J. Dorier, M. Page, P.T. Monteiro, A. von Kamp, I. Xenarios, H. de Jong, M. Hucka, S. Klamt, D. Thieffry, N. Le Novère, J. Saez-Rodriguez, T. Helikar, SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. BMC Syst. Biol. **7**(1), 135 (2013)
18. E.M. Clarke, O. Grumberg, D. Peled, *Model-Checking* (MIT Press, Cambridge, 1999)
19. A. Di Cara, A. Garg, G. De Micheli, I. Xenarios, L. Mendoza, Dynamic simulation of regulatory networks using SQUAD. BMC Bioinf. **8**, 462 (2007)
20. L.C. Freeman, A set of measures of centrality based on betweenness. Sociometry **40**(1), 35–41 (Mar 1977)
21. T.D. Gilmore, Introduction to NF-κB: players, pathways, perspectives. Oncogene **25**(51), 6680–6684 (2006)
22. B.D. Gomperts, P.E.R. Tatham, I.M. Kramer, *Signal Transduction* (Elsevier/Academic Press, Amsterdam, 2009)

23. K. Guruharsha, J.-F. Rual, B. Zhai, J. Mintseris, P. Vaidya, N. Vaidya, C. Beekman, C. Wong, D.Y. Rhee, O. Cenaj, E. McKillip, S. Shah, M. Stapleton, K.H. Wan, C. Yu, B. Parsa, J.W. Carlson, X. Chen, B. Kapadia, K. VijayRaghavan, S.P. Gygi, S.E. Celniker, R.A. Obar, S. Artavanis-Tsakonas, A protein complex network of drosophila melanogaster. Cell **147**(3), 690–703 (2011)
24. A.A. Hagberg, D.A. Schult, P.J. Swart, Exploring network structure, dynamics, and function using NetworkX. in *Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena* (2008), pp. 11–15
25. T. Helikar, J.A. Rogers, ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists. BMC Syst. Biol. **3**, 58 (2009)
26. T. Helikar, B. Kowal, J.A. Rogers, A cell simulator platform: the cell collective. Clin. Pharmacol. Ther. **93**, 393–395 (2013)
27. F. Hinkelmann, M. Brandon, B. Guang, R. McNeill, G. Blekherman, A. Veliz-Cuba, R. Laubenbacher, ADAM: analysis of discrete models of biological systems using computer algebra. BMC Bioinf. **12**, 295 (2011)
28. H. Ikushima K. Miyazono, TGFβ signalling: a complex web in cancer progression. Nat. Rev. Cancer **10**(6), 415–424 (2010)
29. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks. Nature **407**(6804), 651–654 (2000)
30. H. Jeong, S.P. Mason, A.-L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks. Nature **411**(6833), 41–42 (2001)
31. S. Kachalo, R. Zhang, E. Sontag, R. Albert, B. DasGupta, Net-synthesis: a software for synthesis, inference and simplification of signal transduction networks. Bioinformatics **24**(2), 293 (2008)
32. G. Karlebach, R. Shamir, Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. **9**(10), 770–780 (2008)
33. S.A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, 1st edn. (Oxford University Press, Oxford, 1993)
34. T.K. Kerppola, Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. Nat. Protocols **1**(3), 1278–1286 (2006)
35. H.A. Kestler, C. Wawra, B. Kracher, M. Kühl, Network modeling of signal transduction: establishing the global view. BioEssays **30**(11–12), 1110–1125 (2008)
36. S. Klamt, J. Saez-Rodriguez, E.D. Gilles, Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. **1**, 2 (2007)
37. K. Klemm, S. Bornholdt, Stable and unstable attractors in Boolean networks. Phys. Rev. E **72**, 055101 (Nov 2005)
38. J. Krumsiek, S. Pölsterl, D.M. Wittmann, F.J. Theis, Odefy–from discrete to continuous models. BMC Bioinf. **11**, 233 (2010)
39. F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. **101**(14), 4781–4786 (2004)
40. S. Li, S.M. Assmann, R. Albert, Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biol. **4**(10), 1–17 (Sept 2006)
41. J.O. Liu, Everything you need to know about the yeast two-hybrid system. Nat. Struct. Mol. Biol. **5**(7), 535–536 (Jul 1998)
42. Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks. Nature **473**(7346), 167–173 (2011)
43. A. Ma'ayan, S.L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N.J. Eungdamrong, G. Weng, P.T. Ram, J.J. Rice, A. Kershenbaum, G.A. Stolovitzky, R.D. Blitzer, R. Iyengar, Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science **309**(5737), 1078–1083 (2005)

44. K. Markham, Y. Bai, G. Schmitt-Ulms, Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins. Anal. Bioanal. Chem. **389**(2), 461–473 (2007)
45. E.J. McCluskey, Minimization of Boolean functions. Bell Syst. Tech. J. **35**(6), 1417–1444 (1956)
46. C. Müssel, M. Hopfensitz, H.A. Kestler, BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. Bioinformatics **26**, 1378–1380 (2010)
47. A. Naldi, D. Berenguier, A. Fauré, F. Lopez, D. Thieffry, C. Chaouiya, Logical modelling of regulatory networks with GINsim 2.3. Biosystems **97**(2), 134–139 (2009)
48. M.E.J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010)
49. G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society. Nature **435**(7043), 814–818 (2005)
50. B. Palsson, *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, Cambridge, 2006)
51. J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D.S. Goldberg, L.V. Zhang, S.L. Wong, G. Franklin, S. Li, J.S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R.S. Sikorski, J. Vandenhaute, H.Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M.E. Cusick, D.E. Hill, F.P. Roth, M. Vidal, Towards a proteome-scale map of the human protein-protein interaction network. Nature **437**(7062), 1173–1178 (2005)
52. A. Saadatpour, R. Albert, Boolean modeling of biological regulatory networks: a methodology tutorial. Methods **62**(1), 3–12 (2013)
53. A. Saadatpour, R.-S. Wang, A. Liao, X. Liu, T.P. Loughran, I. Albert, R. Albert, Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. PLoS Comput. Biol. **7**(11), e1002267 (2011)
54. A. Saadatpour, R. Albert, T.C. Reluga, A reduction method for Boolean network models proven to conserve attractors. SIAM J. Appl. Dyn. Syst. **12**(4), 1997–2011 (2013)
55. R. Samaga, J. Saez-Rodriguez, L.G. Alexopoulos, P.K. Sorger, S. Klamt, The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput. Biol. **5**(8), 1–19 (2009)
56. R. Schlatter, K. Schmich, I. Avalos Vizcarra, P. Scheurich, T. Sauter, C. Borner, M. Ederer, I. Merfort, O. Sawodny, ON/OFF and beyond–A Boolean model of apoptosis. PLoS Comput. Biol. **5**(12), 1–13 (Dec 2009)
57. M.E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics **27**(3), 431 (2011)
58. S.N. Steinway, J.G. Zañudo, W. Ding, C.B. Rountree, D.J. Feith, T.P. Loughran, R. Albert, Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic Hedgehog and Wnt pathway activation. Cancer Res. **74**(21), 5963–5977 (2014)
59. Z. Sun, R. Albert, Node-independent elementary signaling modes: a measure of redundancy in Boolean signaling transduction networks. Netw. Sci. **4**(3), 273–292 (2016)
60. C.D.A. Terfve, T. Cokelaer, D. Henriques, A. Macnamara, E. Gonçalves, M.K. Morris, M. van Iersel, D.A. Lauffenburger, J. Saez-Rodriguez, CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. BMC Syst. Biol. **6**, 133 (2012)
61. J. Thakar, A.K. Pathak, L. Murphy, R. Albert, I.M. Cattadori, Network model of immune responses reveals key effectors to single and co-infection dynamics by a respiratory bacterium and a gastrointestinal helminth. PLoS Comput. Biol. **8**(1), 1–19 (Jan 2012)
62. R. Thomas, R. d'Ari, *Biological Feedback* (CRC Press, Boca Raton, 1990)
63. J.J. Tyson, K.C. Chen, B. Novak, Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell Biol. **15**(2), 221–231 (2003)

64. P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J.M. Rothberg. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403**(6770), 623–627 (Feb 2000)
65. G. von Dassow, E. Meir, E.M. Munro, G.M. Odell, The segment polarity network is a robust developmental module. Nature **406**(6792), 188–192 (July 2000)
66. R.-S. Wang, R. Albert, Elementary signaling modes predict the essentiality of signal transduction network components. BMC Syst. Biol. **5**(1), 44 (2011)
67. R.-S. Wang, A. Saadatpour, R. Albert, Boolean modeling in systems biology: an overview of methodology and applications. Phys. Biol. **9**(5), 055001 (2012)
68. G. Yang, C. Campbell, R. Albert, Compensatory interactions to stabilize multiple steady states or mitigate the effects of multiple deregulations in biological networks. Phys. Rev. E **94**, 062316 (Dec 2016)
69. J.G.T. Zañudo, R. Albert, An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. Chaos Interdiscip. J. Nonlinear Sci. **23**(2), 025111 (2013)
70. J.G.T. Zañudo, R. Albert, Cell fate reprogramming by control of intracellular network dynamics. PLoS Comput. Biol. **11**(4) (2015)
71. J.G.T. Zañudo, G. Yang, R. Albert, *Structure-based Control of Complex Networks with Nonlinear Dynamics*, Proc. Natl. Acad. Sci. USA **14**(28), 7234–7239 (2017)
72. R. Zhang, M.V. Shah, J. Yang, S.B. Nyland, X. Liu, J.K. Yun, R. Albert, T.P. Loughran, Network model of survival signaling in large granular lymphocyte leukemia. Proc. Natl. Acad. Sci. USA **105**(42), 16308–16313 (2008)
73. J. Zheng, D. Zhang, P.F. Przytycki, R. Zielinski, J. Capala, T.M. Przytycka, SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks. Bioinformatics **26**(1), 141–142 (2010)

Chapter 3
Large-Scale Epidemic Models and a Graph-Theoretic Method for Constructing Lyapunov Functions

Michael Y. Li

Abstract The dynamics of the transmission and spread of infectious diseases are known to be highly complex largely due to the heterogeneity of the host population and the ecology of the pathogens that causes the disease. Factors contributing to the heterogeneity of the host population include age distributions, social and ethnical groups, and spatial distributions, all of which can create complex contact patterns among hosts. Ecological factors for disease pathogens include life cycles, disease vectors, multiple hosts, and environmental influences due to local seasonal changes and large-scale climate changes. Mathematical models that incorporate these factors of heterogeneity often result in a large-scale system of nonlinear differential or difference equations that has a high dimension, multi-components and multi-parameters. While these type of models are more realistic than the classical SIR or SEIR models, its mathematical analysis is highly nontrivial because of the high-dimensionality and their validation from data for reliable predictions is often problematic because of the large number of model parameters. In this chapter, I present a graph-theoretic approach to the construction of Lyapunov functions for establishing the global dynamics of large-scale epidemic models. I will start in Sect. 3.1 with an introduction to epidemic modeling and give two examples of large-scale epidemic models. I will also show two methods for computing the basic reproduction number $\mathscr{R}_0$: the method of van den Driessche and Watmough and the method of using Lyapunov functions. Both methods are based on local stability analysis of the disease-free equilibrium P_0. In Sect. 3.2, I will introduce the notion of dynamical systems on networks as a mathematical framework for large-scale epidemic models and explain the graph-theoretic approach to constructing Lyapunov functions in this general framework. In Sect. 3.3, I will present applications of the graph-theoretic approach to various large-scale models in epidemiology and ecology.

M. Y. Li (✉)
Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: myli@ualberta.ca

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_3

3.1 Large-Scale Epidemic Models for Heterogeneous Populations

We begin by a brief introduction of a simple SIR epidemic model and its mathematical analysis; we will then show how large-scale epidemic models can be built using the simple SIR model as building blocks for two types of host populations: one with a discrete group structure and another with a discrete spatial distribution.

3.1.1 A Simple SIR Epidemic Model

To model the spread of an infectious disease within a host population, we start with a simplistic approach by assuming that the population is homogeneous in the way how individuals interact. We then partition the population into subpopulations (compartments) of susceptibles (S), infectious (I), and recovered (R). We denote the number of individuals in the respective compartment at time t by $S(t)$, $I(t)$, and $R(t)$. The model is depicted in the transfer diagram depicted in Fig. 3.1 In the diagram, Λ denotes the influx of susceptible population either through birth or immigration, whose unit is number of people per unit time, $\beta I S$ the rate of new infections (incidence), γI the rate of recovery, and dS, dI, and dR are the removal rate from each of the compartments. Parameter β is often called the transmission coefficient, whose dimension is 1/people $\cdot$ time, γ and d are rate constants for recovery and removal, respectively, whose units are percentage per unit time. Based on the transfer diagram, we can derive the following system of differential equations for $S(t)$, $I(t)$, and $R(t)$:

$$\begin{aligned} S' &= \Lambda - \beta I S - d S \qquad (3.1)\\ I' &= \beta I S - (\gamma + d) I \qquad (3.2)\\ R' &= \gamma I - d R. \qquad (3.3) \end{aligned}$$

We consider this system with an initial condition $(S_0, I_0, R_0) \in \mathbb{R}^3_+$. It can be verified that the nonnegative cone $\mathbb{R}^3_+$ is positively invariant for system (3.1)–(3.3). This means that solutions $(S(t), I(t), R(t))$ with nonnegative initial conditions will remain nonnegative, and the model is well posed.

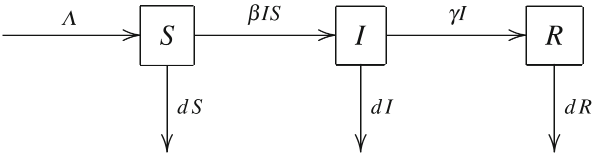


Fig. 3.1 Transfer diagram for a simple SIR model

Adding the three equations and letting $N = S + I + R$, we obtain:

$$N' = \Lambda - dN,$$

which leads to

$$\limsup_{t\to\infty} N(t) \le \frac{\Lambda}{d}.$$

It follows that the bounded region in $\mathbb{R}^3_+$:

$$\{(S, I, R) \in \mathbb{R}^3_+ \mid S + I + R \le \Lambda/d\}$$

is globally attracting and positively invariant. Noticing that the first two equations do not contain the variable R, we can focus on the subsystem consisting of the first two equations:

$$S' = \Lambda - \beta I S - d S \tag{3.4}$$

$$I' = \beta I S - (\gamma + d) I, \tag{3.5}$$

where S, I are taken from the following two-dimensional feasible region:

$$\Gamma = \{(S, I) \in \mathbb{R}^2_+ \mid S + I \le \Lambda/d\}. \tag{3.6}$$

With the dynamics of (3.4)–(3.5) understood, the behaviors of $R(t)$ can be obtained from $R' = \gamma I - dR$.

It can be verified that system (3.4)–(3.5) has two possible equilibria: the disease-free equilibrium $P_0 = (\Lambda/d, 0)$ and a unique endemic (positive) equilibrium $P^* = (S^*, I^*)$, where

$$S^* = \frac{\Lambda}{d}\frac{1}{\mathscr{R}_0}, \qquad I^* = \frac{d}{\beta}(\mathscr{R}_0 - 1),$$

and

$$\mathscr{R}_0 = \frac{\beta}{\gamma + d}\frac{\Lambda}{d}. \tag{3.7}$$

Here, $\mathscr{R}_0$ is the *basic reproduction number*, which measures the average number of direct infections caused by a single infectious individual in an entirely susceptible population (Λ/d) during the average infectious period $(1/(\gamma + d))$ [1, 4, 6, 7, 32, 33].

It is clear from the expression of S^*, I^* that the endemic equilibrium P^* exists in the feasible region Γ if and only if $\mathscr{R}_0 > 1$. In fact, using the standard phase-plane analysis, we can show that an initial outbreak ($I_0 > 0$) can only have two distinct outcomes: either the infection dies out (the epidemic case) or the infection persists

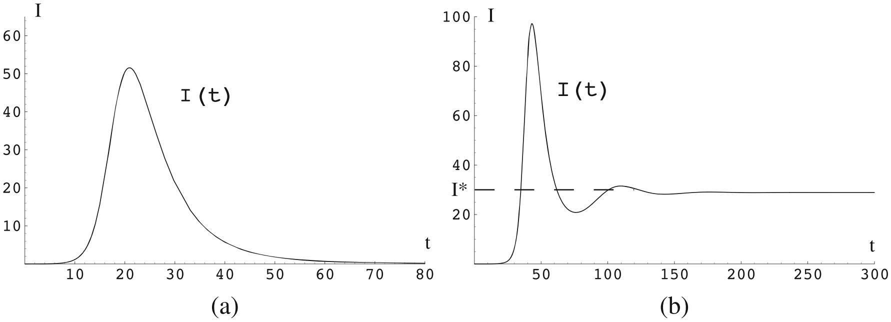


Fig. 3.2 Numerical illustrations of the Threshold Theorem. (**a**) Epidemic case ($\mathscr{R}_0 \leq 1$). (**b**) Endemic case ($\mathscr{R}_0 > 1$)

in the population at a constant level I^* (the endemic case), irrespective of the level of I_0. The threshold parameter that determines the outcome of an disease outbreak is $\mathscr{R}_0$. The precise mathematical statement is given in the following threshold theorem. The two distinct outcomes described in the theorem are illustrated using numerical simulations in Fig. 3.2.

Theorem 3.1 (Threshold Theorem)

1. *If* $\mathscr{R}_0 \leq 1$, *then the disease-free equilibrium* $P_0 = (\Lambda/d\,, 0)$ *is stable and attracts all solutions in* Γ.
2. *If* $\mathscr{R}_0 > 1$, *then* P_0 *is unstable, and the unique endemic (positive) equilibrium* $P^* = (S^*, I^*)$ *is stable and attracts all positive solutions in* Γ.

The threshold theorem can also be interpreted in the context of bifurcation. As the bifurcation parameter $\mathscr{R}_0$ increases across the bifurcation value 1, system (3.4)–(3.5) undergo a transcritical bifurcation in which two branches of equilibria exchange their stability. We note that for $\mathscr{R}_0 < 1$, $I^* < 0$ and the nonzero equilibrium exists but not biological. This bifurcation is illustrated in the bifurcation diagram in Fig. 3.3. In the birfurcation diagram, when the value of $\mathscr{R}_0$ is less than one, system (3.4)–(3.5) has only the disease-free equilibrium P_0 ($I^* = 0$) and it attracts all solutions in Γ. When the value of $\mathscr{R}_0$ is greater than one, system (3.4)–(3.5) has two equilibria: an unstable P_0 (dashed line) and the stable P^*, and P^* attracts all positive solutions. The global attractivity is illustrated numerically using phase portraits in Fig. 3.4, in which orbits from different initial points are shown to converge to P_0 when $\mathscr{R}_0 < 1$ and to P^* when $\mathscr{R}_0 > 1$.

The biological significance of the threshold theorem is that controlling the spread of an infectious disease translates into reducing the value of $\mathscr{R}_0$ to below 1. This can be done for instance by:

- reducing the transmission rate β;
- treating infectious individuals to reduce the mean infectious period $\frac{1}{\gamma+d}$;

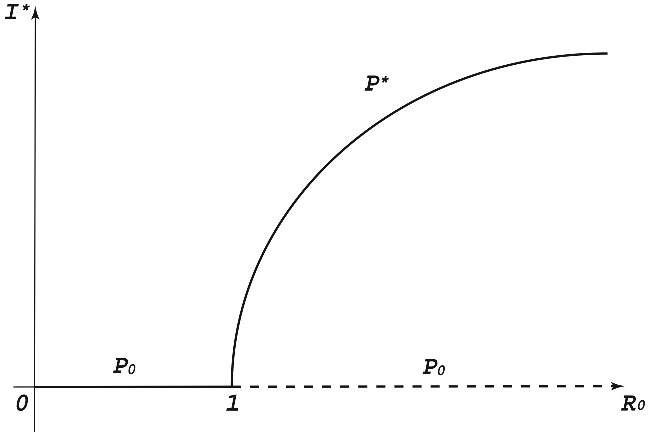


Fig. 3.3 Bifurcation diagram

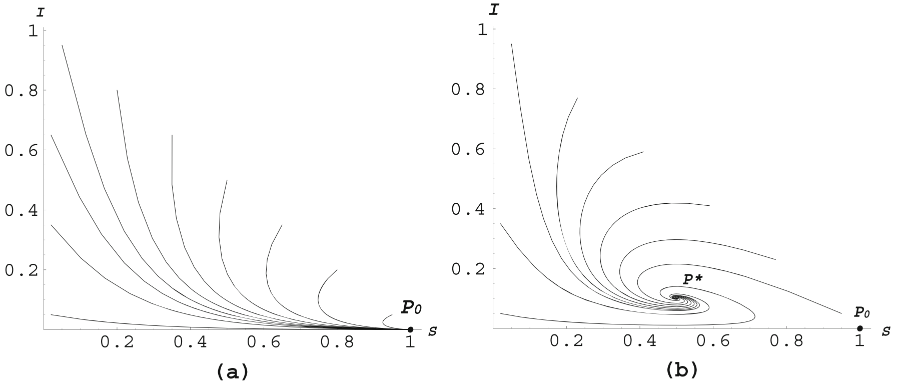


Fig. 3.4 Phase portraits. (**a**) $R_0 < 1$. (**b**) $R_0 > 1$

- vaccinating the population with a vaccine coverage rate $0 < p \leq 1$ to reduce the size of susceptible population to $(1 - p)\Lambda/d$.

3.1.2 Disease Transmission Among Heterogeneous Populations

A major shortcoming of the simple SIR model (3.1)–(3.3) is its assumption of homogeneous mixing: individuals have an equal probability of making a contact with one another. While this assumption can be a first approximation to the mixing of individuals, it is far from adequate as a model for real-world epidemics. A major factor for the complexity of transmission dynamics of infectious diseases is the heterogeneity, which may be caused by many factors:

- structured mixing: different mixing patterns among social, ethnical, and age groups;
- spatial heterogeneity: disease spread across regions, cites, communities, villages;
- epidemiological heterogeneity: differential infectivity or susceptivity, stages of infections, multiple pathogen strains, etc.;
- ecological heterogeneity: intermediate hosts (animals, rodents), disease vectors (mosquitoes), life cycles of pathogens, and environmental influences.

To demonstrate how heterogeneity can be incorporated into an epidemic model, we consider two types of heterogeneity: a group structure and a discrete spatial structure.

3.1.2.1 Disease Transmission in Group-Structured Populations

To overcome the shortcoming of the homogeneous mixing in the simple epidemic model (3.1)–(3.3), we assume that the population can be partitioned into n groups and mixing within each group is homogeneous whereas cross-group mixing can be less frequent than within-group mixing. For each $1 \leq k \leq n$, we further partition the k-th group into compartments S_k, I_k, and R_k of the susceptible, infectious, and recovered individuals, respectively, and let $N_k = S_k + I_k + R_k$ be the total population of the k-th group. Let β_{jk} be the transmission coefficient for transmissions from I_j to S_k, $1 \leq k, j \leq n$. Then the rate of new infections in the k-th group can be written as

$$\sum_{j=1}^{n} \beta_{jk} I_j S_k$$

and the simple SIR model (3.1)–(3.3) can be adapted to the k-th group as

$$S_k' = \Lambda_k - \sum_{j=1}^{n} \beta_{jk} I_j S_k - d_k S_k \tag{3.8}$$

$$I_k' = \sum_{j=1}^{n} \beta_{jk} I_j S_k - (\gamma_k + d_k) I_k \tag{3.9}$$

$$R_k' = \gamma_k I_k - d_k R_k, \tag{3.10}$$

for $k = 1, \cdots, n$. Parameters in (3.8)–(3.10) have the same meaning as those in (3.1)–(3.3) except for the group-specific subindices indicating heterogeneity among groups. We also note that the transmission matrix

$$B = \{\beta_{jk}\} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \cdots & \beta_{nn} \end{bmatrix} \tag{3.11}$$

indicates the within-group and inter-group transmissions. It is nonnegative and may not be strictly positive or symmetric. A reasonable restriction on the transmission matrix B is that it is irreducible. A nonnegative square matrix A is *reducible* if, for some permutation matrix P, PAP^T is block lower triangular, namely,

$$PAP^T = \begin{bmatrix} A_1 & 0 \\ A_2 & A_3 \end{bmatrix},$$

and A_1, A_3 are square matrices. Otherwise, A is *irreducible*. Irreducibility of A can be checked using the associated directed graphs. The weighted *directed graph* $\mathscr{G}(A)$ associated with $A = (a_{jk})_{n\times n}$ has vertices $\{1, 2, \cdots, n\}$, and a directed arc (j, k) from j to k exists if and only if $a_{jk} \neq 0$. Then, $\mathscr{G}(A)$ is *strongly connected* if any two distinct vertices are joined by an oriented path. Then, the matrix A is irreducible if and only if $\mathscr{G}(A)$ is strongly connected [3].

Example 3.1 Consider the matrices A, $\bar{A}$ and their respective digraphs $\mathscr{G}(A)$, $\mathscr{G}(\bar{A})$ to their right in Figs. 3.5 and 3.6. We assume that all the a_{ij} in these two matrices are positive. Matrix $\bar{A}$ is obtained from A by switching the $(1, 2)$ and $(2, 1)$ entries. Correspondingly, in their respective weighted digraphs, the direction of the arrow from vertex 2 to vertex 1 is reversed. By definition, A is a reducible matrix. Correspondingly, $\mathscr{G}(A)$ is not strongly connected since there is no oriented path from vertex 1 to any other vertex. In contrast, the digraph $\mathscr{G}(\bar{A})$ is strongly connected since there is an oriented cycle linking all four vertices. As a result, we know that $\bar{A}$ is irreducible without having to check for all possible permutations.

$$A = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & 0 & a_{44} \end{bmatrix}$$

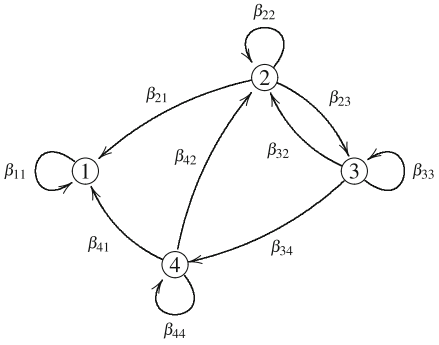


Fig. 3.5 A reducible matrix A and non-strongly connected digraph $\mathscr{G}(A)$

$$\bar{A} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & 0 & a_{44} \end{bmatrix}$$

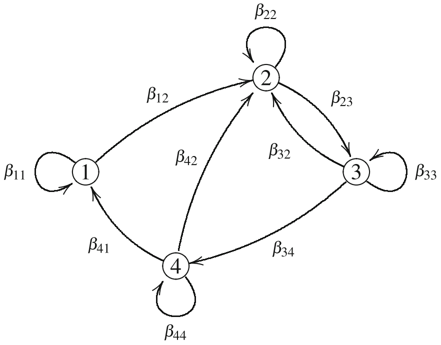


Fig. 3.6 An irreducible matrix $\bar{A}$ and strongly connected digraph $\mathscr{G}(\bar{A})$

For the multi-group model (3.8)–(3.10), the irreducibility of the transmission matrix $B = (\beta_{jk})$ means that infectious individuals in each group can infect the susceptibles in any other group either directly or indirectly through intermediate groups.

For an n-group SIR model (3.8)–(3.10), we can ask similar questions raised for the simple SIR model, as well as questions that are related to heterogeneity.

Mathematical Questions

1. How to compute $\mathscr{R}_0$ for group-structured models?
2. Does the threshold theorem still hold for group-structured models? In particular:
   - (a) If $\mathscr{R}_0 < 1$, is P_0 locally stable or globally stable?
   - (b) If $\mathscr{R}_0 > 1$, is P^* unique? is it locally stable?
   - (c) When P^* is unique, is it globally stable?
   - (d) What methods can we use to prove these results?
3. Can multiple endemic equilibria coexist? If so, what type of bifurcation can occur?

Epidemiological Questions

1. What control strategies does $\mathscr{R}_0$ suggest?
2. When $P^* = (S_1^*, I_1^*, R_n^*, \cdots, S_n^*, I_n^*, R_n^*)$ is globally stable,
   - (a) are there any patterns among I_k^*, $k = 1, 2, \cdots, n$?
   - (b) which I_k^* will be the largest?
   - (c) for which k will the corresponding group have the highest disease prevalence $\frac{I_k^*}{S_k^*+I_k^*+R_k^*}$ and why?

3. If we want to target certain group for disease intervention such as antiviral treatment or disease prevention such as vaccination, what will be the most effective approach?

   We will examine some of these questions in later sections.

3.1.2.2 Disease Spread in Spatially Heterogeneous Populations

In this section, we formulate a mathematical model for the spatial spread of Dengue fever due to travel of infectious individuals. We first consider a spatially homogeneous model for Dengue fever and then incorporate spatial heterogeneity and travel between different regions into the simple model.

A Simple Model for Dengue Fever Dengue diseases are mosquito-borne viral diseases that are prevalent in tropical regions of the world. Dengue virus is transmitted by female mosquitoes of the species *Aedes aegypti*. A female mosquito is infected by the virus from the blood meal drawn from an infectious human or other mammals and can transmit the virus to a susceptible human by biting. Infected mosquitoes are known not to recover from infection of Dengue viruses. Infected humans can develop Dengue hemorrhagic fever which can lead to death. There are four known strains of Dengue viruses. Recovery from infection of one strain provides life-long immunity against the strain, but cross-immunity to other strains after recovery is only partial and temporary.

To model the transmission of Dengue viruses within a homogeneous human population, we need also to consider the mosquito population that serves as the disease vector for Dengue transmission. We partition the human population into the susceptible (S_h), infectious (I_h), and recovered and immune (R_h) compartments, and let $N_h = S_h + I_h + R_h$ be the total human population. For the mosquito population, we consider only the susceptible (S_v) and infectious (I_v) compartment, since infectious mosquitoes do not recover. Let $N_v = S_v + I_v$ be the total mosquito population. A simple Dengue model is depicted in the transfer diagram in Fig. 3.7.

In the model, it is assumed that the influx of susceptible humans is only through birth and is given by bN_h, and that the influx of susceptible mosquitoes is constant and equal to A. The recovery and removal terms in the transfer diagram are similar to those for the simple SIR model (see Fig. 3.2). The key differences between the Dengue model and the SIR model (3.1)–(3.3) lie in the transmission terms, which we explain in the next two paragraphs.

Let b be the average number of bites per day for a typical (female) mosquito among all mammals, β_h the probability of transmission from mosquito bites for humans, and β_v the probability of transmission for mosquitoes from biting an infected mammal. Then the number of new Dengue infections per day in the human population is given by:

$$\beta_h \cdot b \cdot N_v \cdot \frac{N_h}{N_h + a} \cdot \frac{I_v}{N_v} \cdot \frac{S_h}{N_h} = \frac{\beta_h b I_v S_h}{N_h + a}.$$

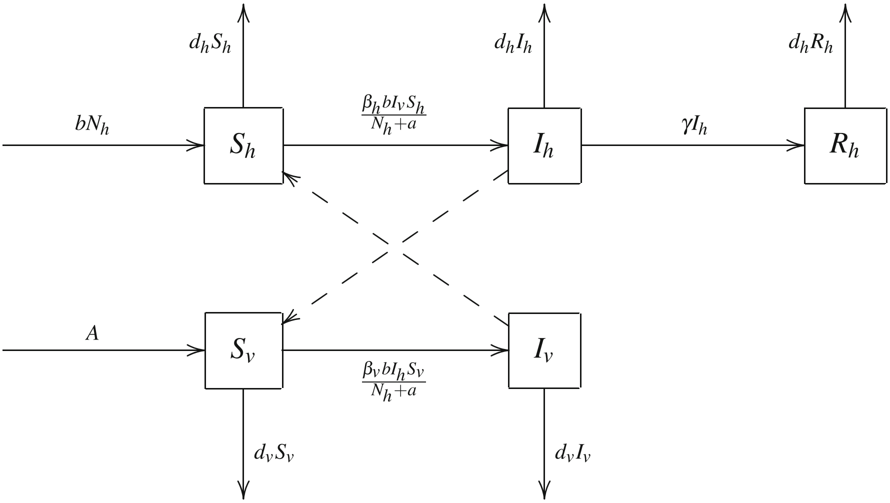


Fig. 3.7 Transfer diagram for a simple Dengue model. The dashed lines indicate cross infections between mosquitoes and human hosts. Parameters are described in the text

In fact, bN_v denotes the total number of mosquito bites per day from all mammals, $\frac{N_h}{N_h+a}$ is the fraction of humans (N_h) among all mammals ($N_h + a$), $\frac{I_v}{N_v}$ is the probability that a bite is from an infectious mosquito, and $\frac{S_h}{N_h}$ is the probability that an infectious mosquito bite is on a susceptible human.

Similarly, the number of new Dengue infections per day in the mosquito population is given by:

$$\beta_v \cdot b \cdot S_v \cdot \frac{N_h}{N_h+a} \cdot \frac{I_h}{N_h} = \frac{\beta_v b I_h S_v}{N_h+a}.$$

Using the transfer diagram in Fig. 3.7, we derive the following model for Dengue fever:

$$S_h' = bN_h - \frac{\beta_h b I_v S_h}{N_h+a} - d_h S_h \tag{3.12}$$

$$I_h' = \frac{\beta_h b I_v S_h}{N_h+a} - (d_h - \gamma) I_h \tag{3.13}$$

$$S_v' = A - \frac{\beta_v b I_h S_v}{N_h+a} - d_v S_v \tag{3.14}$$

$$I_v' = \frac{\beta_v b I_h S_v}{N_h+a} - d_v I_v. \tag{3.15}$$

Here, γI_h represents the number of humans recovered from Dengue. Once recovered, they are immune to reinfection and thus removed from the transmission process. For more details on the model derivation and its mathematical analysis, we refer the reader to Esteva and Vargas [8]. Note that we have neglected the R_h equation since R_h does not appear in the equations for S_h and I_h as in our treatment of the simple SIR model.

A Multi-Region Model for the Spread of Dengue Fever The spread of Dengue diseases over a large geographical area such as a city, province, or country can be caused by the movements of infectious humans and mosquitoes among different geographical regions. We focus on the impact of human movement on the spread of Dengue diseases.

Consider a travel network of n geographical regions represented by a digraph, for instance that as shown in Fig. 3.8. Here, each vertex represents a region, an arrow indicates movement of humans between two regions/vertices, each weight m_{ij} represents rate of movement, i.e. the fraction of humans moving from region i to region j per unit time. The movement (arrow) between i and j exists if and only if $m_{ij}, m_{ji} > 0$, but rates in opposite directions may not be the same. We also note that, because arcs representing movements are bidirectional, the digraph $\mathscr{G}(M)$, $M = (m_{ij})$ is strongly connected if each pair of vertices are connected by a path.

Let $S_{hi}, I_{hi}, R_{hi}, S_{vi}, I_{vi}$ denote the compartments of our simple Dengue model (3.12)–(3.15) for the i-th region. Our interest is to investigate the impact of regional movement of infected humans on the transmission dynamics of Dengue. The net change in the infectious population in the i-th region due to movement is given by

$$\text{inflow rate} - \text{outflow rate} = \sum_{j\neq i} m_{ji} I_{hj} - (\sum_{j\neq i} m_{ij}) I_{hi}.$$

Adding this change to the equation for the infectious human population to the simple Dengue model (3.12)–(3.15), we arrive at the following n-region model for Dengue:

$$S'_{hi} = b_i N_{hi} - \frac{\beta_{hi} b I_{vi} S_{hi}}{N_{hi} + a_i} - d_{hi} S_{hi} \tag{3.16}$$

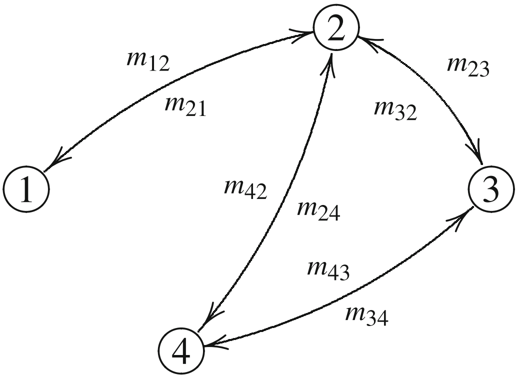


Fig. 3.8 A travel network represented by a digraph $\mathscr{G}(M)$, $M = (m_{ij})$

$$I'_{hi} = \frac{\beta_{hi} b_i I_{vi} S_{hi}}{N_{hi} + a} - (d_{hi} - \gamma_i) I_{hi} + \sum_{j \neq i} m_{ji} I_{hj} - \Big(\sum_{j \neq i} m_{ij}\Big) I_{hi} \tag{3.17}$$

$$S'_{vi} = A - \frac{\beta_{vi} b_i I_{hi} S_{vi}}{N_{hi} + a_i} - d_{vi} S_{vi} \tag{3.18}$$

$$I'_{vi} = \frac{\beta_{vi} b_i I_{hi} S_{vi}}{N_{hi} + a_i} - d_{vi} I_{vi}, \tag{3.19}$$

for $i = 1, 2, \cdots, n$. Note that the whole system has dimension $4n$.

Mathematical questions we can investigate for the n-region Dengue model (3.16)–(3.19) are similar to those for the n-group SIR model (3.8)–(3.10), and they include the computation of the basic reproduction number $\mathscr{R}_0$, number and local stability of equilibria, and validity of the threshold theorem.

Biological questions we can investigate include how to use $\mathscr{R}_0$ to design effective Dengue intervention measures, and how to use a complicated model of large dimension to interpret disease data and make reliable predictions.

However, the complexity and high-dimensionality of these two large-scale models pose serious challenges for our investigations.

3.1.3 Computation of $\mathscr{R}_0$ and Local Stability of P_0

In this section, we discuss the computation of the basic reproduction number $\mathscr{R}_0$ for complex epidemic models as done in [6, 33], and the relation between $\mathscr{R}_0$ and the local stability of the disease-free equilibrium P_0. The presentation in Sect. 3.1.3.1 follows that of van den Driessche and Watmough [33].

3.1.3.1 Computation of $\mathscr{R}_0$ Using the Next Generation Matrix

Let $x = (x_1, \cdots, x_p, x_{p+1}, \cdots, x_q)$ denote the state variable in an epidemic model, where $x_1, \cdots, x_p$ denote the variables for all the infected compartments. Then a disease-free equilibrium P_0 can be expressed as $\bar{x} = (0, \cdots, 0, \bar{x}_{p+1}, \cdots, \bar{x}_q)$. We rearrange the p equations for variables $x_1, \cdots, x_p$ in the model as:

$$x'_i = \mathscr{F}_i(x) - \mathscr{V}_i(x), \quad i = 1, 2, \cdots, p, \tag{3.20}$$

where, for each i, $\mathscr{F}_i$ and $\mathscr{V}_i$ are functions such that $\mathscr{F}_i(x)$ includes all the terms representing new infections and $\mathscr{V}_i(x)$ are made of other terms which typically represent transfers among compartments. Since P_0 is an equilibrium of the epidemic model, $\mathscr{F}_i$ and $\mathscr{V}_i$ satisfy the obvious conditions $\mathscr{F}_i(\bar{x}) = 0$ and $\mathscr{V}_i(\bar{x}) = 0$. We also assume that $\mathscr{F}_i$ and $\mathscr{V}_i$ are smooth functions.

We define the matrices:

$$F = \Big(\frac{\partial \mathscr{F}_i}{\partial x_j}(\bar{x})\Big)_{1\le i,\, j\le p}, \quad V = \Big(\frac{\partial \mathscr{V}_i}{\partial x_j}(\bar{x})\Big)_{1\le i,\, j\le p}.$$

Linearizing the model around the disease-free equilibrium $\bar{x} = (0, \cdots, 0, \bar{x}_{p+1}, \cdots, \bar{x}_q)$ and restricting the linearized system to its first p equations, we obtain the following p-dimensional linear system:

$$y' = (F - V)y, \quad y \in \mathbb{R}^p, \tag{3.21}$$

which is the linearization of system (3.21). The disease-free equilibrium $\bar{x}$ is locally asymptotically stable if the linear system (3.21) is asymptotically stable. The following proposition, established in [33], gives a hint on how to determine the stability of $\bar{x}$.

Proposition 3.1 (van den Driessche and Watmough) *The following statements are equivalent:*

1. *the disease-free equilibrium $\bar{x}$ is locally asymptotically stable;*
2. *all the eigenvalues of $F - V$ have negative real parts;*
3. *all the eigenvalues of $FV^{-1} - I_{m\times m}$ have negative real parts;*
4. *$\rho(FV^{-1}) < 1$, where $\rho(FV^{-1})$ denotes the spectral radius of FV^{-1}.*

The matrix FV^{-1} is called the *second generation matrix*, and its spectral radius $\rho(FV^{-1})$ is the largest modulus of all its eigenvalues. It turns out that $\rho(FV^{-1})$ is the basic reproduction number $\mathscr{R}_0$. For details, see [6, 33].

Example 3.2 (Computing $\mathscr{R}_0$ for the n-group SIR model) As an exercise, we apply the method of van den Driessche and Watmough to compute the basic reproduction number for the n-group SIR model (3.8)–(3.10).

Let $x = (I_1, \cdots, I_n, S_1, \cdots, S_n, R_1, \cdots, R_n)$ represent the rearranged state variables of system (3.8)–(3.10). Note that $p = n$ and $q = 3n$. Then the disease-free equilibrium P_0 can be written as $\bar{x} = (0, \cdots, 0, \bar{S}_1, \cdots, \bar{S}_n, \bar{R}_1, \cdots, \bar{R}_n)$, with $\bar{S}_i = \Lambda_i/d_i$, $i = 1, \ldots, n$. The equations for the infected variables $I_1, \cdots, I_n$ are:

$$I_i' = \sum_{j=1}^{n} \beta_{ji} I_j S_i - (d_i + \gamma_i) I_i, \quad i = 1, \cdots, n.$$

They can be rewritten as,

$$x_i' = \mathscr{F}_i(x) - \mathscr{V}_i(x), \quad i = 1, \cdots, n,$$

where $x_i = I_i$ and

$$\mathscr{F}_i(x) = \sum_{j=1}^{n} \beta_{ji} I_j S_i, \quad \mathscr{V}_i(x) = (d_i + \gamma_i) I_i, \quad i = 1, 2, \cdots, n.$$

Computing the partial derivatives of $\mathscr{F}_i$ and $\mathscr{V}_i$ with respect to $x_1, \cdots, x_n$ at $\bar{x}$, we obtain:

$$F = (\beta_{ji} \bar{S}_i), \quad V = \mathrm{diag}(d_1 + \gamma_1, \cdots, d_n + \gamma_n), \tag{3.22}$$

and the basic reproduction number is given by:

$$\mathscr{R}_0 = \rho(FV^{-1}) = \rho\Big(\frac{\beta_{ji} \bar{S}_i}{d_j + \gamma_j}\Big)_{1 \le i, j \le n}.$$

When parameters in the model are known, the spectral radius $\rho(FV^{-1})$ can be computed using a computer software package to verify if $\mathscr{R}_0 < 1$. For more examples of computation of $\mathscr{R}_0$ for various epidemic models, we refer the reader to [33].

3.1.3.2 Local Stability Analysis of P_0 Using Lyapunov Functions

Another method for establishing local stability of an equilibrium is the method of Lyapunov functions. An advantage of the method of Lyapunov functions is that it is applicable even if real parts of some eigenvalues of the Jacobian matrix at the equilibrium are zero, in which case the method of linearization is not applicable. This is often the case for P_0 when the basic reproduction number $\mathscr{R}_0 = 1$. Another advantage is that Lyapunov functions can be constructed to show global stability, whereas linearization can only determine properties of solutions near the equilibrium.

We use the n-group SIR model (3.8)–(3.10) as an example to show how to use a Lyapunov function and the LaSalle's invariance principle to prove global stability of the disease-free equilibrium P_0 when $\mathscr{R}_0 \le 1$.

Let $S = (S_1, S_2, \cdots, S_n)$, $I = (I_1, I_2, \cdots, I_n)$, and $\bar{S} = (\bar{S}_1, \cdots, \bar{S}_n)$, $\bar{S}_i = \Lambda_i / d_i$, $1 \le i \le n$. Define the matrix:

$$M(S) = \Big(\frac{\beta_{ji} S_i}{d_i + \gamma_i}\Big)_{1 \le i, j \le n}.$$

Then the equations for $I_1, \cdots, I_n$ in system (3.8)–(3.10) can be written in matrix form:

$$I' = M(S) I - \mathrm{diag}(d_1 + \gamma_1, \cdots, d_n + \gamma_n) I.$$

Let $\bar{M} = M(\bar{S})$. If the transmission matrix $B = (\beta_{ji})$ is irreducible, then the nonnegative matrix $\bar{M}$ is also irreducible, since $\frac{\bar{S}_i}{d_i+\gamma_i} > 0$ for all i. The following Perron–Frobenius theorem is well known for nonnegative matrices [3].

Theorem 3.2 (Perron–Frobenius Theorem) *Let $A \geq 0$ be an irreducible square matrix. Then, the spectral radius $\rho(A)$ is a simple eigenvalue with a positive left eigenvector $(w_1, w_2, \cdots, w_n)$, i.e.*

$$(w_1, w_2, \cdots, w_n)A = \rho(A)(w_1, w_2, \cdots, w_n).$$

We note that the matrix $\bar{M}$ can be expressed using F, V in (3.22) as

$$\bar{M} = \left(\frac{\beta_{ji}\bar{S}_i}{d_i + \gamma_i}\right) = V^{-1}F = V^{-1}(FV^{-1})V.$$

Therefore, $\bar{M}$ has the same eigenvalues as the next generation matrix FV^{-1} and thus the same spectral radius:

$$\mathscr{R}_0 = \rho(M(\bar{S})) = \rho(\bar{M}) = \rho(FV^{-1}).$$

We prove the following result.

Proposition 3.2

1. *If $\mathscr{R}_0 \leq 1$, then the disease-free equilibrium P_0 of the n-group SIR model (3.8)–(3.10) is globally stable in Γ.*
2. *If $\mathscr{R}_0 > 1$, then P_0 is unstable.*

Proof We use the following Lyapunov function:

$$V = \sum_{i=1}^{n} \frac{w_i}{d_i + \gamma_i} I_i.$$

Differentiating V along solutions of (3.8)–(3.10), we obtain:

$$\begin{aligned}\frac{dV}{dt} &= \sum_{i=1}^{n} \frac{w_i}{d_i + \gamma_i}\frac{dI_i}{dt} = (w_1, w_2, \cdots, w_n)[M(S)\,I \,-\, I] \\ &\leq (w_1, w_2, \cdots, w_n)[M(\bar{S})\,I \,-\, I] \quad (\text{since } S \leq \bar{S}) \\ &= (\rho(M(\bar{S})) - 1)\sum_{j=1}^{n} w_j I_j \leq 0, \quad \text{if } \rho(M(\bar{S})) \leq 1.\end{aligned}$$

Furthermore:

$$\frac{dV}{dt} = 0 \quad \Longleftrightarrow \quad I = 0,\ S = \bar{S}.$$

LaSalle's invariance principle [22] implies that P_0 is globally stable if $\rho(M(\bar{S})) \leq 1$. If $\rho(M(\bar{S})) > 1$, then for $I > 0$,

$$(w_1, w_2, \cdots, w_n)[M(\bar{S})\, I - I] = [\rho(M(\bar{S})) - 1] \sum_{j=1}^{n} w_j I_j > 0,$$

and thus

$$\frac{dV}{dt} = (w_1, w_2, \cdots, w_n)[M(S)\, I - I] > 0, \quad \text{for } I > 0,$$

in a neighborhood of P_0 by continuity. This implies that P_0 is unstable. □

3.1.3.3 Significance of the Basic Reproduction Number

The basic reproduction number $\mathscr{R}_0$ is an important concept in epidemiology of infectious diseases, and we have seen that it plays a crucial role in determining the dynamics of simple and complex epidemic models.

Mathematical Significance of $\mathscr{R}_0$ For many epidemic models, the basic reproduction number is a sharp threshold parameter that determines the model dynamics. More specifically:

1. If $\mathscr{R}_0 < 1$, then the disease-free equilibrium P_0 is locally stable [33], and the disease dies out when the initial number of infected individuals is small.
2. If $\mathscr{R}_0 > 1$, then P_0 is unstable, and the system is uniformly persistent, namely, there exist positive constants $c_1, \cdots, c_n$ such that $\liminf_{t\to\infty} x_i(t) > c_i$ for all positive solutions [13]; the disease persists in the population.
3. Uniform persistence and boundedness of solutions imply the existence of an endemic (positive) equilibrium P^* [5, 34].

Important mathematical questions to be investigated include:

1. When $\mathscr{R}_0 < 1$, is P_0 globally stable? Or will the disease die out even when there are many infected individuals initially?
2. When $\mathscr{R}_0 > 1$, is P^* unique? Interesting bifurcations may occur when multiple endemic equilibria coexist.
3. If P^* exists, is it unique and locally stable? Is it globally stable? This is often quite challenging to establish.

We comment that there are important classes of epidemic models where $\mathscr{R}_0$ is not a sharp threshold parameter. Endemic equilibria may exist even when $\mathscr{R}_0 < 1$. This is the case when a *backward bifurcation* occurs. For discussions on backward bifurcations in epidemic models, we refer the reader to [33] and references therein.

Epidemiological Significance of $\mathscr{R}_0$ Lessons from epidemic modeling suggest that a key objective of infectious disease control is to reduce the value of basic reproduction number $\mathscr{R}_0$ to below 1. Expressions of $\mathscr{R}_0$ derived from epidemic models can often help design effective control strategies, as we have seen in the case of simple SIR models.

For large-scale epidemic models, the basic reproduction number is given as the spectral radius of a high-dimensional matrix, and an explicit formula for $\mathscr{R}_0$ is often not readily available. As a result, many interesting questions remain regarding disease control and intervention:

1. How should we change parameter values in order to lower the spectral radius $\rho(FV^{-1})$?
2. Can we identify key vertices on the network for targeted control and intervention? For multi-group modeling of sexually transmitted diseases (STD), such a question is related to the idea of "core groups" of Yorke, Hethcote, and Nold [35].
3. Using multi-region models, how do we design an effective way of travel restriction to stop the spread of local outbreaks to other regions?
4. For vector controls using genetically modified (GM) mosquitoes, how can we identify key regions in the network for the release of GM mosquitoes to effectively lower $\mathscr{R}_0$ while minimizing environmental risks?

These epidemiological questions can inspire many interesting mathematical studies on complex modeling and complex systems.

3.2 Dynamical Systems on Networks

What common features do the multi-group SIR model (3.8)–(3.10) and the multi-region Dengue model (3.16)–(3.19) share? First of all, they are both large-scale systems with a large number of variables, which make them difficult to analyze mathematically, and a large number of parameters, which make parameter estimation and model identification from disease data difficult. However, a closer examination shows that both systems share an important structure:

1. the equations for each group or region contain many terms from those of the simple homogeneous model, together with interaction terms among groups or regions;
2. if we remove the interactions among groups (cross infections) or regions (movements), then each isolated group or region has a simple homogeneous model, whose dynamics are well-understood;
3. the interactions can be encoded on a directed graph.

In fact, we will show later that such a structure is shared by many large-scale systems in biology and epidemiology, as well as in physical sciences and engineering. It makes mathematical sense to have a common framework in which all such large-scale systems can be investigated.

In this section, we describe the framework of dynamical systems on networks under which we investigate large-scale epidemic models. We begin by introducing some terminologies and results in graph theory including the Kirchhoff's matrix-tree theorem and a new Tree Cycle Identity. We then present a general graph-theoretic approach to the construction of Lyapunov functions for dynamical systems on networks that was developed in a series of papers by Guo et al. [11–14, 26, 27]. This approach makes it possible to systematically construct global Lyapunov functions for large-scale models. We also present applications of the approach to several models in ecology, epidemiology, and engineering. The development of the materials in this section follows that in [26].

3.2.1 Preliminaries of Graph Theory

A directed graph or a *digraph* $\mathscr{G} = (V, E)$ consists of a set $V = \{1, 2, \ldots, n\}$ of vertices and a set E of directed arcs connecting pairs of vertices. We denote the directed arc from vertex i to vertex j by (i, j). A digraph $\mathscr{G}$ is *weighted* if each arc (i, j) is assigned a weight $a_{ij} \geq 0$. When an arc (i, j) does not exist, we set the corresponding weight $a_{ij} = 0$. The weight matrix $A = (a_{ij})$ is a nonnegative matrix. In general, matrix A may not be strictly positive or symmetric.

Conversely, each nonnegative $n \times n$ matrix A defines a weighted digraph $\mathscr{G}(A)$ with n vertices, and a directed arc (i, j) exists if and only if $a_{ij} > 0$.

A digraph $\mathscr{G}$ is *strongly connected* if, for each pair of vertices $i \neq j$, there exists an oriented path from i to j. A nonnegative matrix A is *reducible* if there exists a permutation matrix P such that $P^T AP$ is block lower triangular, and otherwise A is irreducible. We have seen in Sect. 3.1.2.1 that a nonnegative matrix A is irreducible if and only if the weighted digraph $\mathscr{G}(A)$ is strongly connected [3].

Let $\mathscr{G}$ be a weighted digraph with weight matrix $A = (a_{ij})_{n\times n}$. A directed *tree* is a connected subgraph containing no cycles, directed or undirected. The weight $w(\mathscr{T}$ of a tree $\mathscr{T}$ is the product of the weights of its arcs. A tree $\mathscr{T}$ is *rooted* at a vertex i if the remaining vertices of $\mathscr{T}$ are connected by directed paths from the *root* i. A tree $\mathscr{T}$ is *spanning* if it contains all vertices of $\mathscr{G}$. A subgraph $\mathscr{H}$ of $\mathscr{G}$ is *unicyclic* if it contains a unique directed cycle. An illustration of a rooted spanning tree is given in Fig. 3.9.

The *Laplacian matrix* $L(A)$ of matrix A is defined as

$$L(A) = \mathrm{diag}(d_1, \cdots, d_n) - A$$

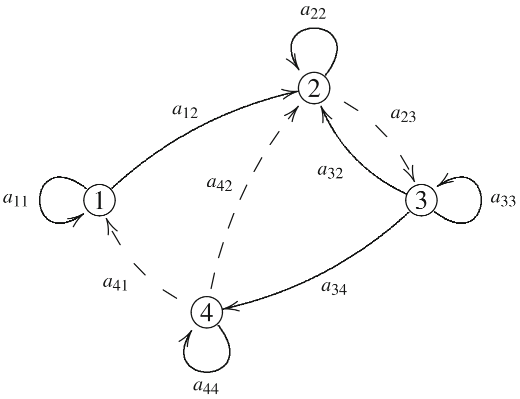


Fig. 3.9 A spanning tree rooted at vertex 4 is marked using dashed arrows

where $d_i = \sum_{j=1}^{n} a_{ij}$, the sum of elements of the i-th row of A. Thus $L(A)$ looks like:

$$L(A) = \begin{bmatrix} \sum_{k\neq 1} a_{1k} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \sum_{k\neq 2} a_{2k} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \sum_{k\neq n} a_{nk} \end{bmatrix}.$$

The following result is known as Kirchhoff's matrix-tree theorem, which was first published in 1847 [17]. An English translation appeared in 1958 [18]. For a modern treatment, we refer the reader to [29].

Theorem 3.3 (Kirchhoff's Matrix-Tree Theorem) *Let C_{ii} be the co-factor of the i-th diagonal element of $L(A)$, namely, the determinant of the submatrix obtained by removing the i-th row and i-th column of $L(A)$. Then*

$$C_{ii} = \sum_{\mathscr{T}\in\mathbb{T}_i} w(\mathscr{T}), \quad i = 1, 2, \cdots, n,$$

where $\mathbb{T}_i$ is the set of all spanning trees of $\mathscr{G}(A)$ rooted at vertex i.

Example 3.3 As an illustration of the matrix-tree theorem, we consider the case $n = 3$. For a 3×3 matrix $A = (a_{ij})$, direct calculation shows that

$$C_{11} = a_{32}a_{21} + a_{21}a_{31} + a_{23}a_{31}. \tag{3.23}$$

All possible spanning trees rooted at vertex 1 are shown in Fig. 3.10. Adding the sum of weights of the three digraphs in Fig. 3.10, we obtain the same expression for C_{11} as in (3.23).

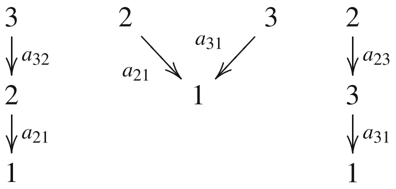


Fig. 3.10 All possible spanning trees of a 3-digraph that are rooted at vertex 1

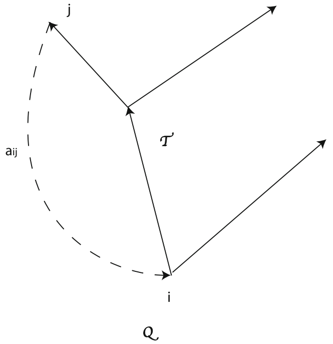


Fig. 3.11 A unicycle graph $\mathscr{Q}$ is formed by adding an arc (dashed) to a rooted tree $\mathscr{T}$ (shown in solid lines)

Using the matrix-tree theorem, Li and Shuai proved the following Tree-Cycle Identity [26], which is a regrouping of double sums according to unicyclic spanning graphs.

Theorem 3.4 (Tree-Cycle Identity) *Let $c_i = C_{ii}$ be that given in the matrix-tree theorem. Then the following identity holds:*

$$\sum_{i,j=1}^{n} c_i\, a_{ij}\, F_{ij}(x) = \sum_{\mathscr{Q}\in\mathbb{Q}} w(\mathscr{Q}) \sum_{(r,s)\in E(\mathscr{C}_{\mathscr{Q}})} F_{rs}(x)$$

where $F_{ij}(x)$, $1 \le i, j \le n$, are arbitrary functions, $\mathbb{Q}$ is the set of all spanning unicyclic subgraphs $\mathscr{Q}$ of $\mathscr{G}(A)$, $w(\mathscr{Q})$ is the weight of $\mathscr{Q}$, and $\mathscr{C}_{\mathscr{Q}}$ denotes the oriented cycle of $\mathscr{Q}$.

Proof For a detailed proof, we refer the reader to [26]. Here we note that $w(\mathscr{T})\, a_{ij} = w(\mathscr{Q})$, where $\mathscr{Q}$ is the unicyclic graph obtained by adding an arc (j, i) to $\mathscr{T}$. See the illustration in Fig. 3.11. □

3.2.2 Dynamical Systems on Networks

A network is defined as a digraph $\mathscr{G}$ such that at each vertex i, a system of differential equations is defined:

$$x_i' = f_i(x_i), \quad x_i \in \mathbb{R}^{d_i}, \quad f_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i}. \tag{3.24}$$

Such a vertex system (3.24) describes the dynamics on an isolated vertex. Influence of vertex i on a vertex j is described by a function $g_{ij} : \mathbb{R}^{d_i} \times \mathbb{R}^{d_j} \to \mathbb{R}^{d_i}$. An arc from j to i in the digraph $\mathcal{G}$ is absent if and only if $g_{ij} \equiv 0$.

A *dynamical system on a network* $\mathcal{G}$ is defined as a coupled system of differential equations:

$$x_i' = f_i(x_i) + \sum_{j=1}^{n} g_{ij}(x_i, x_j), \qquad i = 1, 2, \ldots, n. \tag{3.25}$$

The state variable of (3.25) is $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \mathbb{R}^{d_n} \times$, with a total dimension $N = d_1 + d_2 + \cdots + d_n$.

The multi-group SIR model (3.8)–(3.10) can be regarded as a dynamical system on the transmission network $\mathcal{G}(B)$ defined by the transmission matrix $B = (\beta_{ij})$. Each vertex represents a single group with vertex dynamics described by

$$\begin{aligned} S_k' &= \Lambda_k - \beta_{kk} I_k S_k - d_k\, S_k \\ I_k' &= \beta_{kk} I_k S_k - (\gamma_k + d_k)\, I_k \\ R_k' &= \gamma_k\, I_k - d_k\, R_k, \end{aligned}$$

for $k = 1, \cdots, n$. The influence from vertex j to vertex i is the cross infection between I_j and S_i:

$$g_{ij}(S_i, I_i, S_j, I_j) = (-\beta_{ij} S_i I_j,\ \beta_{ij} S_i I_j, 0)^T.$$

Similarly, for the multi-region Dengue model (3.16)–(3.19), the network is the travel network $\mathcal{G}(M)$ among regions defined by the matrix $M = (m_{ij})$. Each vertex is a region with vertex dynamics described a single region Dengue model:

$$S_{hi}' = b_i N_{hi} - \frac{\beta_{hi} b I_{vi} S_{hi}}{N_{hi} + a_i} - d_{hi} S_{hi} \tag{3.26}$$

$$I_{hi}' = \frac{\beta_{hi} b_i I_{vi} S_{hi}}{N_{hi} + a} - (d_{hi} - \gamma_i) I_{hi} \tag{3.27}$$

$$S_{vi}' = A - \frac{\beta_{vi} b_i I_{hi} S_{vi}}{N_{hi} + a_i} - d_{vi} S_{vi} \tag{3.28}$$

$$I_{vi}' = \frac{\beta_{vi} b_i I_{hi} S_{vi}}{N_{hi} + a_i} - d_{vi} I_{vi}, \tag{3.29}$$

for $i = 1, 2, \cdots, n$. The influence from vertex j to vertex i is the net movement of infected humans from region j to region i:

$$g_{ij}(S_{hi}, I_{hi}, S_{vi}, I_{vi}) = (0, m_{ji} I_{hj} - m_{ij} I_{hi}, 0, 0)^T.$$

In Sect. 3.3, we will describe more examples of complex models as dynamical systems defined on a network.

3.2.3 Global Stability of Equilibria

Local stability of an equilibrium describes properties of solutions in a small neighborhood of the equilibrium. Local stability can be established either by the method of linearization (if the equilibrium is hyperbolic) or by the method of local Lyapunov functions. In contrast, global stability of an equilibrium $\bar{x}$ describes the property that all solutions in a large region converge to the same equilibrium $\bar{x}$.

For a system of differential equations defined in an open set $D \subset \mathbb{R}^N$:

$$x' = f(x), \quad x \in D \subset \mathbb{R}^N, \tag{3.30}$$

where $f : D \to \mathbb{R}^N$ is sufficiently smooth so that a solution to the initial value problem exists and is unique.

An equilibrium $\bar{x}$ of (3.30) is *stable* if for each ϵ-neighborhood $N(\bar{x}, \epsilon)$ of $\bar{x}$, there exists a δ-neighborhood $N(\bar{x}, \delta)$ of $\bar{x}$ such that $x_0 \in N(\bar{x}, \delta)$ implies $x(t, x_0) \in N(\bar{x}, \epsilon)$ for all $t \geq 0$. The equilibrium $\bar{x}$ is *asymptotically stable* if it is stable and if there exists b-neighborhood $N(\bar{x}, b)$ such that $x_0 \in N(\bar{x}, b)$ implies $x(t, x_0) \to \bar{x}$ as $t \to \infty$. In this case, the equilibrium $\bar{x}$ is said to "attract points" in the neighborhood $N(\bar{x}, b)$.

An equilibrium $\bar{x}$ is *globally stable* with respect to a set $G \subset \mathbb{R}^N$ if (a) it is locally stable and (b) it attracts points in G, namely, $x_0 \in G$ implies $x(t, x_0) \to \bar{x}$ as $t \to \infty$. The property (b) is often called *global attractivity* of $\bar{x}$. We note that in general global attractivity may not imply local stability.

The proof of global stability of an equilibrium is more challenging than that of the local stability. General methods for proving global stability include:

1. constructing of global Lyapunov functions;
2. applying the theory of monotone dynamical systems (see [15, 31]);
3. applying the theory of autonomous convergence (see [23–25, 30]).

For complex and large-scale mathematical models, the most practical and effective method is the construction of global Lyapunov functions.

Let $U \subset \mathbb{R}^N$ be a neighborhood of the equilibrium $\bar{x}$ and $V : U \mapsto \mathbb{R}$ a C^1 real-valued function. The gradient vector of $V(x)$ is

$$\text{grad } V(x) = \left(\frac{\partial V}{\partial x_1}, \cdots, \frac{\partial V}{\partial x_n}\right).$$

The derivative of V in the direction of the vector field f of system (3.30) is defined as

$$\overset{*}{V}(x) = \text{grad } V(x) \cdot f(x).$$

This is also called the Lyapunov derivative with respect to system (3.30). The function $V(x)$ is called a *Lyapunov function* of system (3.30) near an equilibrium $\bar{x}$ if

$$\overset{*}{V}(x) \leq 0, \quad \text{for } x \text{ in a neighborhood } U \text{ of } \bar{x}. \tag{3.31}$$

If the inequality (3.31) holds in an open set $G \subset \mathbb{R}^N$, then V is said to be a *global* Lyapunov function*Lyapunov function!global Lyapunov function* with respect to G.

Let $x(t, x_0)$ be a solution of system (3.30) that stays in G, then

$$\frac{d}{dt}V(x(t)) = \text{grad } V(x(t)) \cdot x'(t) = \text{grad } V(x(t)) \cdot f(x(t)) = \overset{*}{V}(x(t)) \leq 0.$$

Therefore, $V(x(t))$ decreases along a solution in G. As a consequence, the omega-limit set of the solution, namely,

$$\omega(x_0) = \{x \in G \ : \ \text{there exists } t_n \to \infty \text{ such that } x(t_n, x_0) \to x_1 \text{ as } n \to \infty\},$$

is contained in the set where $\overset{*}{V}(x) = 0$. Since omega-limit sets are invariant, we know that $\omega(x_0)$ must be contained in the largest invariant subset K of $\{x \in G \ : \ \overset{*}{V}(x) = 0\}$. This is the well-known LaSalle's invariance principle [22], which is often used with global Lyapunov functions to prove global stability.

Theorem 3.5 (LaSalle's Invariance Principle) *Let V be a Lyapunov function for system* (3.30) *with respect to G. If a solution $x(t, x_0)$ stays entirely in G for $t \geq 0$, then $\omega(x_0) \cap G \subset K$, where $\omega(x_0)$ is the omega-limit set of $x(t, x_0)$ and K is the largest invariant subset of $\{x \in G \ : \ \overset{*}{V}(x) = 0\}$.*

3.2.4 Constructing Global Lyapunov Functions for Heterogeneous Models

In this section, we describe a graph-theoretic method for constructing global Lyapunov functions for dynamical systems on networks, which is a mathematical framework for many large-scale heterogeneous models in biology.

The idea is quite simple: since each vertex system (3.24) is lower dimensional and usually well studied, we assume that it has a global Lyapunov function $V_i(x_i)$ with respect to $x_i \in D_i \subset \mathbb{R}^{d_i}$, for $i = 1, \cdots, n$. We consider a Lyapunov function for the coupled system (3.25) in the form:

$$V(x) = \sum_{i=1}^{n} c_i V_i(x_i), \quad x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^N, \quad N = d_1 + \cdots + d_n. \tag{3.32}$$

Key Question How to choose suitable constants $c_i \geq 0$ such that V is a global Lyapunov function for the coupled system (3.25) with respect to $D = D_1 \times \cdots \times D_n$?

The following theorem of Li and Shuai [26] provides an answer to this question and a general approach to construct Lyapunov functions of large-scale heterogeneous models.

Theorem 3.6 (Li and Shuai) *Assume that:*

1. *there exists a family* $\{F_{ij}(x)\}$ *such that*

$$\overset{*}{V_i}(x_i) \leq \sum_{j=1}^{n} a_{ij} F_{ij}(x), \quad x \in D = D_1 \times \cdots \times D_n, \quad i = 1, \cdots, n;$$

2. *the family* $\{F_{ij}(x)\}$ *satisfies the Cycle Conditions, i.e. along each directed cycle* $\mathscr{C}$ *of* $\mathscr{G}(A)$*,* $A = (a_{ij})$*,*

$$\sum_{(r,s) \in E(\mathscr{C})} F_{rs}(x) \leq 0, \quad t > 0, \ x \in D.$$

Let $c_i = C_{ii}$ *as in the matrix-tree theorem for* $\mathscr{G}(A)$*. Then* $V(x) = \sum_{i=1}^{n} c_i V_i(x)$ *satisfies*

$$\overset{*}{V}(x) \leq 0, \quad x \in D.$$

Proof Let $c_i = C_{ii}$ be given in the matrix-tree theorem for $\mathscr{G}(A)$. Then, for $x \in D$,

$$\begin{aligned}
\overset{*}{V}(x) = \sum_{i=1}^{n} c_i \overset{*}{V_i} &\leq \sum_{i,j=1}^{n} c_i a_{ij} F_{ij}(x) \quad \text{(Assumption 1)} \\
&= \sum_{\mathscr{Q} \in \mathbb{Q}} w(\mathscr{Q}) \sum_{(r,s) \in E(\mathscr{C}_{\mathscr{Q}})} F_{rs}(x) \quad \text{(Tree − Cycle Identity)} \\
&\leq 0. \quad \text{(Cycle Conditions)}
\end{aligned}$$

□

We note that, from the matrix-tree theorem, if the matrix $A = (a_{ij})$ is strongly connected, then $c_i > 0$ for all i.

Theorem 3.6 offers a systematic approach to the construction of global Lyapunov functions for a coupled system, using individual Lyapunov functions for its vertex systems. To demonstrate the applicability of the approach, we apply it to prove the global stability of the endemic equilibrium of the multi-group SIR model (3.8)–(3.10).

Example 3.4 Consider the multi-group SIR model (3.8)–(3.10). We ignore the equation for R_k since R_k does not appear in the equations of S_k and I_k, and obtain the following reduced system:

$$S_k' = \Lambda_k - \sum_{j=1}^{n} \beta_{jk} I_j S_k - d_k S_k \tag{3.33}$$

$$I_k' = \sum_{j=1}^{n} \beta_{jk} I_j S_k - (\gamma_k + d_k) I_k, \tag{3.34}$$

for $k = 1, \cdots, n$. We study solutions to (3.33)–(3.34) in the following feasible region in $\mathbb{R}_+^{2n}$:

$$\Gamma_1 = \{(S_1, I_1, \cdots, S_n, I_n) \in \mathbb{E}^{2n} \mid 0 \le S_k + I_k \le \frac{\Lambda_k}{d_k},\ k = 1, 2, \cdots, n\}.$$

From our discussions on the computation of $\mathscr{R}_0$ in Sect. 3.1.3.1, we know that $\mathscr{R}_0$ of model (3.8)–(3.10) only depends on the I_k equations, which are preserved in model (3.33)–(3.34). Therefore, the basic reproduction number of the reduced model (3.33)–(3.34) is the same as that of model (3.33)–(3.34). Furthermore, from Proposition 3.2 in Sect. 3.1.3.2, we know that if $\mathscr{R}_0 > 1$, the disease-free equilibrium P_0 is unstable, system (3.33)–(3.34) is uniformly persistent in Γ_1, and there exists an endemic equilibrium P^* in the interior $\overset{\circ}{\Gamma}_3$. The next result of Guo et al. [12, 13] establishes the uniqueness and global stability of P^*.

Theorem 3.7 *Assume that $B = (\beta_{ij})$ is irreducible. If $\mathscr{R}_0 > 1$, then the endemic equilibrium P^* is unique for system (3.33)–(3.34) and globally stable in the interior $\overset{\circ}{\Gamma}_1$ of Γ_1.*

Proof We use the Lyapunov function for a single-group SIR model discovered by Korobeinikov [19, 20]:

$$V_i(S_i, I_i) = (S_i - S_i^* + S_i^* \log \frac{S_i}{S_i^*}) + (I_i - I_i^* - I_i^* \log \frac{I_i}{I_i^*}),$$

and consider a Lyapunov function for the n-group model (3.33)–(3.34) of the form:

$$V(S_1, I_1, \cdots, S_n, I_n) = \sum_{i=1}^{n} c_i V_i(S_i, I_i).$$

Differentiating V_i along solutions of (3.33)–(3.34) and simplifying, we obtain

$$\overset{*}{V_i} = -\frac{d_i^S}{S_i}(S_i - S_i^*)^2 + \sum_{j=1}^{n} \beta_{ij} S_i^* I_j^* \Big(2 - \frac{S_i^*}{S_i} - \frac{I_i}{I_i^*} + \frac{I_j}{I_j^*} - \frac{S_i I_j I_i^*}{S_i^* I_j^* I_i}\Big).$$

Let $a_{ij} = \beta_{ij} S_i^* I_j^*$, $G_i(I_i) = -\frac{I_i}{I_i^*} + \log \frac{I_i}{I_i^*}$, $\phi(a) = 1 - a + \log a$, and

$$F_{ij}(S_i, I_i, I_j) = 2 - \frac{S_i^*}{S_i} - \frac{I_i}{I_i^*} + \frac{I_j}{I_j^*} - \frac{S_i I_j I_i^*}{S_i^* I_j^* I_i}.$$

Note that $\phi(a) = 1 - a + \log a \leq 0$ for all $a \in \mathbb{R}$ and $\phi(1) = 0$, then

$$\overset{*}{V_i} \leq \sum_{ij} a_{ij} F_{ij}(S_i, I_i, I_j).$$

This shows that Assumption 1 of Theorem 3.6 is satisfied. Furthermore,

$$F_{ij} = G_i(I_i) - G_j(I_j) + \phi\Big(\frac{S_i^*}{S_i}\Big) + \phi\Big(\frac{S_i I_j I_i^*}{S_i^* I_j^* I_i}\Big) \leq G_i(I_i) - G_j(I_j),$$

and thus F_{ij} satisfies the Cycle Conditions (Assumption 2 of Theorem 3.6).

Therefore, by Theorem 3.6, if we choose $c_i = C_{ii}$ as in the matrix-tree theorem, V satisfies $\overset{*}{V} \leq 0$ for $S_k > 0, I_k > 0$, $k = 1, \cdots, n$. By the LaSalle's invariance principle, the omega-limit set ω of each positive solution $(S_1(t), I_1(t), \cdots, S_n(t), I_n(t))$ to (3.33)–(3.34) belongs to the largest invariant subset K of

$$G = \{(S_1, I_1, \cdots, S_n, I_n) \in \overset{\circ}{\Gamma}_3 \mid \overset{*}{V} = 0\}.$$

To characterize the largest invariant set K, we first observe that irreducibility of matrix $B = (\beta_{ij})$ implies that $A = (\beta_{ij} S_i^* I_j^*)$ is irreducible, and thus $c_i = C_{ii} > 0$ for $i = 1, \cdots, n$, by the matrix-tree theorem. Therefore, $\overset{*}{V} = 0$ implies $F_{ij}(S_i, I_i, I_j) = 0$ for $S_i > 0, I_i > -0$, $i = 1, \cdots, n$. As a result, we know that $S_i = S_i^*, I_i = aI_i^*$, $i = 1, 2, \ldots, n$ for some constant $a > 0$ independent i. Substituting these relations into the first equation of system (3.33)–(3.34), we obtain

$$0 = \Lambda_k - d_i^S S_k^* - a \sum_{j=1}^{n} \beta_{kj} S_k^* I_j^*. \tag{3.35}$$

Since the right-hand side of (3.35) is strictly decreasing in a, we know (3.35) holds if and only if $a = 1$. Therefore, the largest invariant subset K of the set G is the singleton $\{P^*\}$, and thus all omega-limit sets are the same as $\{P^*\}$. This establishes the global stability of P^*. □

3.3 Further Applications

In this section, we provide further examples of large-scale mathematical models whose analysis can be done using the approach described in Sect. 3.2. We provide a detailed analysis of a multi-group SEIR epidemic model and give a brief description on the global-stability analysis for a model for a network of coupled oscillators, a multi-patch predator-prey model, and a multi-group SEIR model with time delays.

3.3.1 Application I: A Multi-Group SEIR Model with Bilinear Incidence

Many infectious diseases have a latent period during which an infected individual is not contagious. For instance, measles has a latent period of 14–20 days. To model diseases with latency, we divide the infected subpopulation into two compartments: the latent compartment E for those who are not infectious and an infectious compartment I. We continue to use S to denote the susceptible compartment and R for the recovered and immune compartment. We also assume that the recovery from the infection causes a permanent immunity against reinfection, as in the case of measles.

For a heterogeneous population with group structure, we partition the k-th group into S_k, E_k, I_k, and R_k compartments, and the transmission of the disease and transfer of individuals among compartments are illustrated in the transfer diagram in Fig. 3.12, where only two groups are shown. Based on the transfer diagram, we can derive the following system of differential equations for an n-group SEIR model:

$$S_k' = \Lambda_k - d_k^S S_k - \sum_{j=1}^{n} \beta_{kj} S_k I_j \tag{3.36}$$

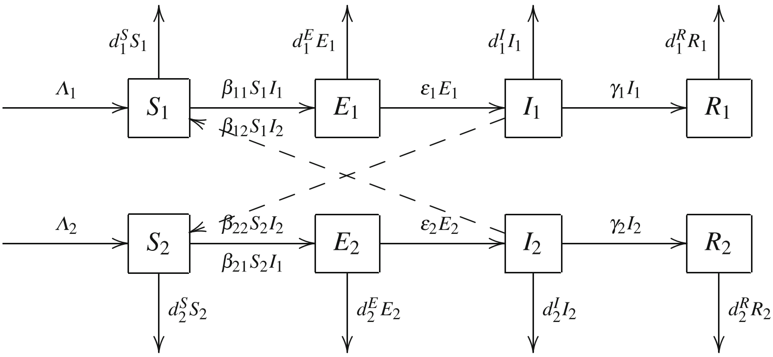


Fig. 3.12 Transfer diagram for a 2-group SEIR model. Dashed arrows indicate cross-group transmissions

$$E_k' = \sum_{j=1}^{n} \beta_{kj} S_k I_j - (d_k^E + \epsilon_k) E_k \tag{3.37}$$

$$I_k' = \epsilon_k E_k - (d_k^I + \gamma_k) I_k, \qquad k = 1, \cdots, n. \tag{3.38}$$

Here, we have neglected the equation for R_k since R_k does not appear in the equations for S_k, E_k, and I_k. We investigate system (3.36)–(3.38) in the following feasible region:

$$\Gamma_2 = \Big\{(S_1, E_1, I_1, \cdots, S_n, E_n, I_n) \in \mathbb{R}_+^{3n} \quad | \quad S_k \leq \frac{\Lambda_k}{d_k^S}, \\ S_k + E_k + I_k \leq \frac{\Lambda_k}{d_k^*},\ k = 1, 2, \cdots, n\Big\},$$

with $d_k^* = \min\{d_k^S, d_k^E, d_k^I + \gamma_k\} > 0$. Here, for each k, Λ_k is the influx of susceptible individuals in the k-th group.

3.3.1.1 Equilibria and the Basic Reproduction Number

System (3.36)–(3.38) always has a unique disease-free equilibrium $P_0 = (S_1^0, 0, 0, \cdots, S_n^0, 0, 0)$, with $S_k^0 = \frac{\Lambda_k}{d_k^S}$, $1 \leq k \leq n$. Using the method of van den Driessche and Watmough, we can derive that the basic reproduction number is

$$\mathscr{R}_0 = \rho \begin{bmatrix} \frac{\beta_{11}\epsilon_1 S_1^0}{(d_1^E+\epsilon_1)(d_1^I+\gamma_1)} & \cdots & \frac{\beta_{1n}\epsilon_n S_1^0}{(d_n^E+\epsilon_n)(d_n^I+\gamma_n)} \\ \vdots & \ddots & \vdots \\ \frac{\beta_{n1}\epsilon_1 S_n^0}{(d_1^E+\epsilon_1)(d_1^I+\gamma_1)} & \cdots & \frac{\beta_{nn}\epsilon_n S_n^0}{(d_n^E+\epsilon_n)(d_n^I+\gamma_n)} \end{bmatrix},$$

where $\rho(A)$ denotes the spectral radius of a matrix A.

As in the case of the n-group SIR model (3.8)–(3.10) in Sect. 3.2, we can prove the following result. A detailed proof can be found in [12, 13].

Proposition 3.3 *1. If $\mathscr{R}_0 \leq 1$, then P_0 is globally asymptotically stable in Γ.*
2. If $\mathscr{R}_0 > 1$, then P_0 is unstable and the system is uniformly persistent.
3. There exists $P^* \in \overset{\circ}{\Gamma}_2$ when $\mathscr{R}_0 > 1$.

In the following, we show how the graph-theoretic approach to constructing Lyapunov functions can be applied to model (3.36)–(3.38) to prove the uniqueness and global stability of the endemic equilibrium. Let $B = (\beta_{kj})$ be the transmission matrix.

Theorem 3.8 (Guo, Li, and Shuai) *Assume that B is irreducible. If $\mathscr{R}_0 > 1$, then there is a unique endemic equilibrium P^* and it is globally asymptotically stable in $\overset{\circ}{\Gamma}_4$.*

We present a sketch of the proof in several steps. We first prove that any endemic equilibrium $P^* = (S_1^*, I_1^*, \dots, S_n^*, I_n^*)$ is globally stable using a Lyapunov function, then the uniqueness of P^* follows from the global stability.

Lyapunov Functions We use the class of Lyapunov functions:

$$V = \sum_{k=1}^{n} v_k \Big[(S_k - S_k^* \ln S_k) + (E_k - E_k^* \ln E_k) + \frac{d_k^E + \epsilon_k}{\epsilon_k} (I_k - I_k^* \ln I_k) \Big]$$

and choose a suitable set of constants $v_k > 0$ so that $\frac{dV}{dt} \leq 0$ in $\overset{\circ}{\Gamma}_4$ with the help of graph theory.

Differentiating V along solutions to (3.36)–(3.38), we obtain

$$\begin{aligned} V' &= \sum_{k=1}^{n} v_k \Big[(S_k' - \frac{S_k^*}{S_k} S_k') + (E_k' - \frac{E_k^*}{E_k} E_k') + \frac{d_k^E + \epsilon_k}{\epsilon_k} (I_k' - \frac{I_k^*}{I_k} I_k') \Big] \\ &= \sum_{k=1}^{n} v_k \Big[d_k^S S_k^* \Big(2 - \frac{S_k^*}{S_k} - \frac{S_k}{S_k^*} \Big) \Big] \\ &\quad + \sum_{k=1}^{n} v_k \Big[\sum_{j=1}^{n} \beta_{kj} S_k^* I_j - \frac{(d_k^E + \epsilon_k)(d_k^I + \alpha_k + \gamma_k)}{\epsilon_k} I_k \Big] \\ &\quad + \sum_{j,k=1}^{n} v_k \, \beta_{kj} S_k^* I_j^* \Big(3 - \frac{S_k^*}{S_k} - \frac{S_k}{S_k^*} \frac{I_j}{I_j^*} \frac{E_k^*}{E_k} - \frac{I_k^*}{I_k} \frac{E_k}{E_k^*} \Big). \end{aligned}$$

We set

$$H_n := \sum_{j,k=1}^{n} v_k \, \bar{\beta}_{kj} \Big(3 - \frac{S_k^*}{S_k} - \frac{S_k}{S_k^*} \frac{I_j}{I_j^*} \frac{E_k^*}{E_k} - \frac{I_k^*}{I_k} \frac{E_k}{E_k^*} \Big).$$

Choosing Constants v_k We choose v_k so that

$$\sum_{k=1}^{n} v_k \Big[\sum_{j=1}^{n} \beta_{kj} S_k^* I_j - \frac{(d_k^E + \epsilon_k)(d_k^I + \alpha_k + \gamma_k)}{\epsilon_k} I_k \Big] \equiv 0$$

for all $I_1, \cdots, I_n > 0$. This is equivalent to

$$\begin{bmatrix} \beta_{11} S_1^* I_1^* & \cdots & \beta_{n1} S_n^* I_1^* \\ \vdots & \ddots & \vdots \\ \beta_{1n} S_1^* I_n^* & \cdots & \beta_{nn} S_n^* I_n^* \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} \beta_{1j} S_1^* I_j^* v_1 \\ \vdots \\ \sum_{j=1}^{n} \beta_{nj} S_n^* I_j^* v_n, \end{bmatrix},$$

since, at P^*:

$$\frac{(d_k^E + \epsilon_k)(d_k^I + \alpha_k + \gamma_k)}{\epsilon_k} = \sum_{j=1}^{n} \beta_{kj} S_k^* I_j^*.$$

Set $\bar{\beta}_{kj} = \beta_{kj} S_k^* I_j^*$ and

$$\overline{B} = \begin{bmatrix} \sum_{l\neq 1} \bar{\beta}_{1l} & -\bar{\beta}_{21} & \cdots & -\bar{\beta}_{n1} \\ -\bar{\beta}_{12} & \sum_{l\neq 2} \bar{\beta}_{2l} & \cdots & -\bar{\beta}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -\bar{\beta}_{1n} & -\bar{\beta}_{2n} & \cdots & \sum_{l\neq n} \bar{\beta}_{nl} \end{bmatrix}.$$

Then $v = (v_1, \dots, v_k)^T$ satisfies the linear system

$$\overline{B}\, v = 0.$$

Matrix $\overline{B}$ is the Laplacian matrix of $(\bar{\beta}_{ij})$, and the column sums of $\bar{B}$ are zero, so that nontrivial solutions of $v_1, \cdots, v_k$ exist. The matrix-tree theorem and the irreducibility of $\bar{B}$ imply that

$$v_k = \sum_{T\in\mathbb{T}_k} \prod_{(j,h)\in E(T)} \bar{\beta}_{jh}.$$

Re-Grouping Terms in H_n According to Unicyclic Cycles Using the Tree-Cycle Identity

We have

$$\begin{aligned} H_n &= \sum_{j,k=1}^{n} v_k\, \bar{\beta}_{kj} \Big(3 - \frac{S_k^*}{S_k} - \frac{S_k}{S_k^*}\frac{I_j}{I_j^*}\frac{E_k^*}{E_k} - \frac{I_k^*}{I_k}\frac{E_k}{E_k^*}\Big) \\ &= \sum_{Q} w(Q) \sum_{(p,q)\in E(CQ)} \Big[3 - \frac{S_p^*}{S_p} - \frac{S_p I_q E_p^*}{S_p^* I_q^* E_p} - \frac{E_p I_p^*}{E_p^* I_p}\Big] \quad \text{(Tree} - \text{Cycle Identity)} \\ &= \sum_{Q} w(Q) \cdot \Big[3r - \sum_{(p,q)\in E(CQ)} \Big(\frac{S_p^*}{S_p} + \frac{S_p I_q E_p^*}{S_p^* I_q^* E_p} + \frac{E_p I_p^*}{E_p^* I_p}\Big)\Big] \end{aligned}$$

Finally, because CQ is a cycle:

$$\prod_{(p,q)\in E(CQ)} \frac{S_p^*}{S_p} \cdot \frac{S_p I_q E_p^*}{S_p^* I_q^* E_p} \cdot \frac{E_p I_p^*}{E_p^* I_p} = \prod_{(p,q)\in E(CQ)} \frac{I_q I_p^*}{I_q^* I_p} = 1.$$

This implies that:

$$3r - \sum_{(p,q)\in E(CQ)} \Big(\frac{S_p^*}{S_p} + \frac{S_p I_q E_p^*}{S_p^* I_q^* E_p} + \frac{E_p I_p^*}{E_p^* I_p}\Big) \leq 0,$$

and thus $H_n \leq 0$. We have proved that $V'(t) \leq 0$.

Proving the Global Stability of P^* Using LaSalle's Invariance Principle We can use the same arguments as in the proof of Theorem 3.7 in Sect. 3.2.4 to show that P^* is globally stable in $\overset{\circ}{\Gamma}_4$, and thus is also unique. This completes the proof.

3.3.2 Application II: A Network of Coupled Oscillators

Consider a network of coupled oscillators described by the following system of second-order differential equations:

$$x_i'' + \alpha x_i' + f_i(x_i) + \sum_{j=1}^{n} \epsilon_{ij}(x_i' - x_j') = 0 \tag{3.39}$$

where x_i is the displacement of the i-th oscillator. The parameter α is the damping coefficient, and $f_i(x_i)$ is the restoring force for the i-th oscillator. Let $y_i = x_i'$ be the velocity, then we obtain the following equivalent coupled system of first-order equations:

$$x_i' = y_i, \tag{3.40}$$

$$y_i' = -\alpha_i y_i - f_i(x_i) - \sum_{j=1}^{n} \epsilon_{ij}(y_i - y_j). \tag{3.41}$$

System (3.40)–(3.41) can be naturally regarded as a dynamical system on a network. The network is given by the digraph $\mathscr{G}(E)$, $E = (\epsilon_{ij})$. Each vertex is an oscillator, and the vertex dynamics are given by:

$$\begin{aligned} x_i' &= y_i, \\ y_i' &= -\alpha_i y_i - f_i(x_i). \end{aligned}$$

Assume that $\alpha_i \geq 0$ and that the potential energy $F_i(x_i) = \int^{x_i} f_i(s)ds$ has a strictly global minimum at $x_i = x_i^*$. Then $E^* = (x^*, 0, \cdots, x_n^*, 0)$ is an equilibrium for system (3.40)–(3.41). The total energy of each isolated oscillator is a natural Lyapunov function for the vertex system:

$$V_i(x_i, y_i) = F_i(x_i) + \frac{y_i^2}{2}.$$

Direct calculation yields

$$\begin{aligned}
\overset{*}{V_i} &= (x_i - x_i^*)\Big[f_i(x_i) + \sum_{j=1}^{n} d_{ij}\Big(\frac{x_j}{x_i} - \alpha_{ij}\Big)\Big] \\
&= (x_i - x_i^*)\Big[-\sum_{j=1}^{n} d_{ij}\Big(\frac{x_j^*}{x_i^*} - \alpha_{ij}\Big) + (f(x_i) - f(x_i^*)) + \sum_{j=1}^{n} d_{ij}\Big(\frac{x_j}{x_i} - \alpha_{ij}\Big)\Big] \\
&= (x_i - x_i^*)(f(x_i) - f(x_i^*)) + \sum_{j=1}^{n} d_{ij}x_j^*\Big(\frac{x_j}{x_j^*} - \frac{x_i}{x_i^*} + 1 - \frac{x_i^* x_j}{x_i x_j^*}\Big).
\end{aligned}$$

Let $a_{ij} = d_{ij}x_j^*$, $F_{ij}(x_i, x_j) = \frac{x_j}{x_j^*} - \frac{x_i}{x_i^*} + 1 - \frac{x_i^* x_j}{x_i x_j^*}$, and $G_i(x_i) = -\frac{x_i}{x_i^*} + \ln\frac{x_i}{x_i^*}$. Then we have:

$$\overset{*}{V_i} \le \sum_{j=1}^{n} a_{ij} F_{ij}(x_i, x_j)$$

and

$$F_{ij}(x_i, x_j) = G_i(x_i) - G_j(x_j) + 1 - \frac{x_i^* x_j}{x_i x_j^*} + \ln\frac{x_i^* x_j}{x_i x_j^*} \le G_i(x_i) - G_j(x_j).$$

This shows that V_i and F_{ij} satisfy the assumptions of Theorem 3.6, and thus

$$V(x_1, y_1, \cdots, x_n, y_n) = \sum_{i=1}^{n} c_i V_i(x_i, y_i)$$

is a global Lyapunov function for the coupled system (3.40)–(3.41) if we choose $c_i = C_{ii}$ using the matrix-tree theorem with $A = (d_{ij}x_j^*)$. The global stability of $E^* = (x^*, 0, \cdots, x_n^*, 0)$ follows from an application of the LaSalle's invariance principle. This establishes the following result in [26].

Theorem 3.9 (Li and Shuai) *Assume that the digraph $\mathscr{G}(E)$ is strongly connected. Suppose that there exists k such that $\alpha_k > 0$. Then E^* is globally asymptotically stable in $\mathbb{R}^{2n}$.*

3.3.3 Application III: An n-Patch Predator-Prey Model

Consider a predator-prey model in which preys disperse among n patches ($n \geq 2$):

$$x_i' = x_i(r_i - b_i x_i - e_i y_i) + \sum_{j=1}^{n} d_{ij}(x_j - x_i), \tag{3.42}$$

$$y_i' = y_i(-\gamma_i - \delta_i y_i + \epsilon_i x_i), \quad i = 1, 2, \ldots, n. \tag{3.43}$$

Here, x_i, y_i denote the densities of preys and predators on the i-th patch, respectively. All parameters in the model are nonnegative constants, and e_i, ϵ_i are positive. The dispersal constants d_{ij} are nonnegative, and $D = (d_{ij})$ is the dispersal matrix, which defines the dispersal network $\mathscr{G}(D)$. We refer the reader to [9, 21] for interpretations of predator-prey models and parameters.

System (3.42)–(3.43) is a dynamical system on the dispersal network defined by the digraph $\mathscr{G}(D)$. Each vertex is a patch, and vertex dynamics is given by a single-patch predator-prey model:

$$x_i' = x_i(r_i - b_i x_i - e_i y_i) \tag{3.44}$$

$$y_i' = y_i(-\gamma_i - \delta_i y_i + \epsilon_i x_i), \tag{3.45}$$

$i = 1, 2, \ldots, n$. We leave it to the reader to verify that the vertex Lyapunov function

$$V_i(x_i, y_i) = \epsilon_i(x_i - x_i^* \ln x_i) + e_i(y_i - y_i^* \ln y_i)$$

satisfies the assumptions of Theorem 3.6. Details can be found in [26]. We comment that this form of Lyapunov function was used in the ecological modeling literature since the 1970s, see [10, 16].

Theorem 3.10 (Li and Shuai) *Assume that the dispersal matrix $D = (d_{ij})$ is irreducible and that there exists k such that $b_k \delta_k > 0$. Then the positive equilibrium E^*, whenever it exists, is unique and globally asymptotically stable in $\mathbb{R}_+^{2n}$.*

3.3.4 Application IV: A Multi-Group Epidemic Model with Time Delays

In this section, we give an example to demonstrate that the graph-theoretic approach is also applicable to large-scale models with time delays.

Consider a multi-group SIR model with discrete time delays:

$$S_i' = \Lambda_i - d_i^S S_i - \sum_{j=1}^{n} \beta_{ij} S_i I_j(t - \tau_j), \tag{3.46}$$

$$I_i' = \sum_{j=1}^{n} \beta_{ij} S_i I_j(t-\tau_j) - (d_i^I + \gamma_i) I_i, \qquad i = 1, 2, \cdots, n. \tag{3.47}$$

Model parameters have similar interpretations as in the multi-group SIR model (3.8)–(3.10). The time delays τ_i are the result of disease latency. We customarily omitted the equation for R_i since R_i does not appear in the equations of S_i and I_i.

Similar to the ODE case, delayed model (3.46)–(3.47) can be regarded as a coupled system of differential equations on the transmission network $\mathscr{G}(B)$ defined by the transmission matrix $B = (\beta_{ij})$. The vertex dynamics at each vertex (group) are defined by a system of delay differential equations describing a single-group SIR model with latency (see [2]),

$$S_i' = \Lambda_i - d_i^S S_i - \beta_{ii} S_i I_i(t-\tau_i), \tag{3.48}$$

$$I_i' = \beta_{ii} S_i I_i(t-\tau_i) - (d_i^I + \gamma_i) I_i, \tag{3.49}$$

$i = 1, 2, \cdots, n$. The coupling between vertices i and j is provided by cross infections $\beta_{ij} S_i I_j(t-\tau_j)$ and $\beta_{ji} S_j I_i(t-\tau_i)$. For each vertex system (3.48–3.49), we use a Lyapunov functional first constructed by McCluskey [28]:

$$\begin{aligned} V_i(S_i, I_i(\cdot)) =& S_i - S_i^* + S_i^* \ln \frac{S_i}{S_i^*} + I_i - I_i^* - I_i^* \ln \frac{I_i}{I_i^*} \\ &+ \sum_{j=1}^{n} \beta_{ij} S_i^* \int_0^{\tau_j} \Big(I_j(t-r) - I_j^* - I_j^* \ln \frac{I_j(t-r)}{I_j^*} \Big)\, dr. \end{aligned}$$

The reader can verify that V_i satisfies the assumptions of Theorem 3.6.

Theorem 3.11 (Li and Shuai) *Assume that $B = (\beta_{ij})$ is irreducible. If $\mathscr{R}_0 > 1$, then the unique endemic equilibrium P^* for system (3.46)–(3.47) is globally asymptotically stable.*

Acknowledgements Much of the work presented in this lecture note is based on joint work with Drs. Hongbin Guo and Zhisheng Shuai. I acknowledge the support of Natural Science and Engineering Research Council of Canada (NSERC) and Canada Foundation for Innovation (CFI) for my research.

References

1. R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1992)
2. E. Beretta, Y. Takeuchi, Global stability of an SIR epidemic model with time delays. J. Math. Biol. **33**, 250–260 (1995)

3. A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Academic Press, New York, 1979)
4. F. Brauer, C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology, Texts in Applied Mathematics*, vol. 40 (Springer, Berlin, 2001)
5. G.J. Butler, H.I. Freedman, P. Waltman, Uniformly persistent systems. Proc. Amer. Math. Soc. **96**, 425–430 (1986)
6. O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**, 365–382 (1990)
7. O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics* (Princeton University Press, Princeton, 2012)
8. L. Esteva, C. Vargas, Analysis of a dengue disease transmission model. Math Biosci. **150**, 131–51 (1998)
9. H.I. Freedman, *Deterministic Mathematical Models in Population Ecology* (Marcel Dekker, New York, 1980)
10. B.S. Goh, Global stability in many-species systems. Am. Nat. **111**, 135–143 (1977)
11. H. Guo, M.Y. Li, Global dynamics of a staged-progression model with amelioration for infectious diseases. J. Biol. Dynam. **2**, 154–168 (2008)
12. H. Guo, M.Y. Li, Z. Shuai, Global stability of the endemic equilibrium of multigroup SIR epidemic models. Can. Appl. Math. Q. **14**, 259–284 (2006)
13. H. Guo, M.Y. Li, Z. Shuai, A graph-theoretic approach to the method of global Lyapunov functions. Proc. Amer. Math. Soc. **136**, 2793–2802 (2008)
14. H. Guo, M.Y. Li, Z. Shuai, Global dynamics of a general class of multistage models for infectious diseases. SIAM J. Appl. Math. **72**, 261–270 (2012)
15. M.W. Hirsch, Systems of differential equations that are competitive or cooperative: I. limit sets. SIAM J. Math. Anal. **13**, 167–179 (1982)
16. S.B. Hsu, On global stability of a predator-prey systems. Math. Biosci. **39**, 1–10 (1978)
17. G. Kirchhoff, Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung Galvanischer Ströme geführt wird. Annalen der Physik und Chemie **72**, 497–508 (1847)
18. G. Kirchhoff, On the solution of the equations obtained from the investigation of the linear distribution of Galvanic currents. IRE Trans. Circuit Theory **5**, 4–7 (1958)
19. A. Korobeinikov, Global properties of infectious disease models with nonlinear incidence. Bull. Math. Biol. **69**, 1871–1886 (2007)
20. A. Korobeinikov, P.K. Maini, A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence. Math. Biosci. Eng. **1**, 57–60 (2004)
21. Y. Kuang, Y. Takeuchi, Predator-prey dynamics in models of prey dispersal in two-patch environments. Math. Biosci. **120**, 77–98 (1994)
22. J.P. LaSalle, *The Stability of Dynamical Systems, Regional Conference Series in Applied Mathematics* (SIAM, Philadelphia, 1976)
23. M.Y. Li, J.S. Muldowney, On Bendixson's criterion. J. Differ. Equ. **106**, 27–39 (1993)
24. M.Y. Li, J.S. Muldowney, On RA Smith's autonomous convergence theorem. Rocky Mount. J. Math. **25**, 365–379 (1995)
25. M.Y. Li, J.S. Muldowney, A geometric approach to global-stability problems. SIAM J. Math. Anal. **27**, 1070–1083 (1996)
26. M.Y. Li, Z. Shuai, Global-stability problem for coupled systems of differential equations on networks. J. Differ. Equ. **248**, 1–20 (2010)
27. M.Y. Li, Z. Shuai, C. Wang, Global stability of multi-group epidemic models with distributed delays. J. Math. Anal. Appl. **361**, 38–47 (2010)
28. C.C. McCluskey, Complete global stability for an SIR epidemic model with delay – distributed or discrete. Nonlinear Anal. Real World Appl. **11**, 55–59 (2010)
29. J.W. Moon, *Counting Labelled Trees* (Canadian Mathematical Congress, Montreal, 1970)
30. R.A. Smith, Some applications of Hausdorff dimension inequalities for ordinary differential equations. Proc. Roy. Soc. Edinburgh Sect. A **104**, 235–259 (1986)

31. H.L. Smith, *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems* (American Mathematical Society, Providence, 1995)
32. H.R. Thieme, *Mathematics in Population Biology* (Princeton University Press, Princeton, 2003)
33. P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
34. P. Waltman, A brief survey of persistence in dynamical systems, in: *Delay Differential Equations and Dynamical Systems, Lecture Notes in Mathematics 1475*, ed. by S. Busenberg, M. Martelli (Springer, Berlin, 1991), pp. 31–40
35. J.A. Yorke, H.W. Hethcote, A. Nold, Dynamics and control of the transmission of gonorrhea. Sex. Transm. Dis. **5**, 51–56 (1978)

Chapter 4
Mixing in Meta-Population Models

Zhilan Feng and John W. Glasser

Abstract Among the means by which heterogeneity can be modeled, Levins' (Bull Entomol Soc Am 15:237–240, 1969) meta-population approach preserves the most analytical tractability. When model populations are stratified, contacts among their respective sub-populations must be described. Using a simple meta-population model, Feng et al. (J Theor Biol 386:177–187, 2015) showed that mixing among sub-populations, as well as heterogeneity in characteristics affecting sub-population reproduction numbers, must be considered when evaluating public health interventions to prevent or control infectious disease outbreaks. We employed the convex combination of preferential within- and proportional among-group contacts devised by Nold (Math Biosci 52:227–240, 1980) and generalized by Jacquez et al. (Math Biosci 92:119–199, 1988). As the utility of meta-population modeling in support of public policymaking depends on more realistic mixing functions, Glasser et al. (Math Biosci 235:1–7, 2012) included preferential contacts between parents and children and among co-workers as well as contemporaries. Feng et al. (Math Biosci 287:93–104, 2017) omitted workplace contacts, but added those between grandparents and grandchildren. We also devised a general scheme for multi-level mixing that meets the conditions for mixing functions specified by Busenberg and Castillo-Chavez (IMA J Math Appl Med Biol 8:1–29, 1991) and provided several two-level examples.

Z. Feng (✉)
Department of Mathematics, Purdue University, West Lafayette, IN, USA
e-mail: zfeng@math.purude.edu

J. W. Glasser
National Center for Immunization and Respiratory Diseases, CDC, Atlanta, GA, USA
e-mail: jglasser@cdc.gov

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics
of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_4

4.1 Introduction

Analysis of effective reproduction numbers from models of stratified populations, to which Levins [1] referred as meta-populations (i.e. populations composed of sub-populations), can identify targets for effective outbreak prevention or control measures. Such models must specify how infectious members of one sub-population contact susceptible members of others, to which we refer as *mixing*. Heterogeneity in characteristics affecting sub-population reproduction numbers affects the magnitude of meta-population reproduction numbers, especially if mixing is non-random [2]. In this chapter, we present several examples of mixing functions for meta-population models appropriate for heterogeneous contacts in age, spatial location, gender, etc.

4.2 Forces of Infection

If immunity following recovery from infection is lifelong, the probability of remaining susceptible at age α is:

$$P_S(\alpha) = e^{-\int_0^\alpha \lambda(u)du}$$

where $\lambda(u)$ is the *force* or *hazard rate of infection* at age u. One can estimate the $\lambda(u)$ by fitting $P_I(\alpha) = 1 - P_S(\alpha)$, the cumulative probability of infection at age α, to histories of infection. From results from a cross-sectional (i.e. including all ages) serological survey, for example, one can calculate the proportions infected by age:

$$-P_S'(\alpha) = \lambda(\alpha)\, e^{-\int_0^\alpha \lambda(u)du}.$$

To estimate the infection rates β_{ij} required for age-structured transmission modeling from such information when parameters are constant within age groups, Anderson and Grenfell [3] defined

$$\lambda_i := \sum_{j=1}^{n} \beta_{ij} \frac{I_j}{N_j}, \quad i = 1, 2, \ldots, n, \tag{4.1}$$

where I_j/N_j is the proportion of infected/infectious individuals in age group j. They coined the phrase *Who Acquires Infection from Whom* for the matrix consisting of as many unique β_{ij} values as age groups, and explored sensible alternative arrangements. This approach was described as recently as 2012 by Hens et al. [4], who also recount other essentially descriptive methods.

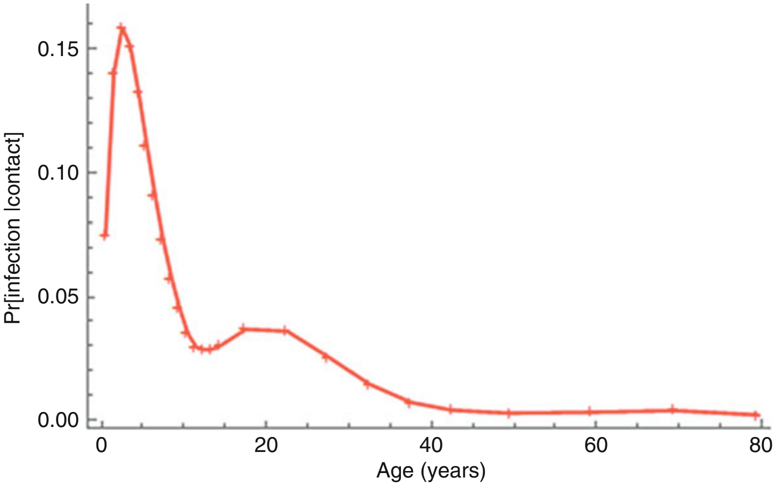


Fig. 4.1 Age-specific susceptibility to infection on contact β_i, estimated from (**a**) hazard rates of infection λ_i and proportions infectious I_i/N_i, in turn from proportions with antibodies to pertussis toxin above 150 Elisa Units per ml from a cross-sectional serological survey in Sweden, together with (**b**) per capita contact rates a_i and proportions with all age groups c_{ij} in Finland, in turn from PolyMod as described under parameter estimation below. Given this information, we can solve for the remaining unknown in the n Eq. (4.2). Source: Feng et al. [5]

We advocate a more mechanistic approach that requires information about the intensity, a_i, and pattern of inter-personal contacts, c_{ij}. Re-defining λ_i as

$$\lambda_i := a_i \beta_i \sum_{j=1}^{n} c_{ij} \frac{I_j}{N_j}, \quad i = 1, 2, \ldots, n, \tag{4.2}$$

we are able to estimate the probability of infection on contact with an infectious person, β_i. Evidently, the equation $\beta_{ij} = a_i \beta_i c_{ij}$ describes the relationship between parameters β in Eqs. (4.1) and (4.2). The requisite information about mixing may be empirical, hypothetical (i.e. a model), or hybrid (i.e. a model whose parameters have been estimated from observations).

The probabilities of infection on contact with infectious people are informative. Fig. 4.1, for example, indicates increased susceptibility to pertussis among older adolescents and young adults as well as children in Sweden during the 17-year hiatus in vaccination.[1] Some of those young adults were caring for infants, for

[1] Production problems, together with widespread concern about the safety of the whole-cell pertussis vaccine, led Swedish health authorities to discontinue vaccination from 1979 to 1995. Pertussis became endemic again in Sweden, permitting evaluation of several acellular vaccines in clinical trials, upon whose successful conclusion those vaccines were licensed and vaccination resumed.

whom pertussis may be fatal. Previously infected or vaccinated adults are likely to experience mild, immunity-modified disease, and may not realize that they are infected or, in fact, be particularly infectious. Nonetheless, caregiver contacts with infants may be sufficiently intimate and prolonged for transmission.

4.3 A Simple Mixing Model

In 1991, Busenberg and Castillo-Chavez described three conditions that mixing functions should meet [6]:

$$\begin{aligned} &1)\ c_{ij} \geq 0, \\ &2)\ \textstyle\sum_{j=1}^{n} c_{ij} = 1, \quad i = 1, \ldots, n, \\ &3)\ a_i N_i c_{ij} = a_j N_j c_{ji}, \end{aligned} \tag{4.3}$$

where the a_i*'s* are per capita contact rates (termed activities), c_{ij} is the proportion of their contacts that members of group *i* have with members of group *j*, and the N_i*'s* are group sizes. The first model meeting these conditions of which we are aware was described by Jacquez et al. [7], who modified the model of Nold [8]. They defined c_{ij} as:

$$c_{ij} := \varepsilon_i \delta_{ij} + (1 - \varepsilon_i) f_j, \quad f_j = \frac{(1 - \varepsilon_j) a_j N_j}{\sum_k (1 - \varepsilon_k) a_k N_k}, \tag{4.4}$$

where the ε_i*'s* are fractions of contacts reserved for one's own group (termed preferences), a constant in Nold's [8] model, δ_{ij} is the Kronecker delta (equals 1 when $i = j$ and 0 otherwise), and a_j and N_j are as previously defined. The function f_j describes mixing that is random (i.e. proportional to contacts not reserved for one's own group $(1 - \varepsilon_j) a_j N_j$).

When $n = 2$ sub-populations, the mixing matrix C is

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 + (1 - \varepsilon_1) f_1 & (1 - \varepsilon_1) f_2 \\ (1 - \varepsilon_2) f_1 & \varepsilon_2 + (1 - \varepsilon_2) f_2 \end{bmatrix}.$$

Because $0 \leq \varepsilon_i \leq 1$, this model is very flexible. The limiting conditions (i.e. all $\varepsilon_i = 0$ or $0 < \varepsilon_i \leq 1$) are termed *proportional* and *preferential* mixing, respectively, and preferential mixing may be *heterogeneous* (i.e. all ε_i need not be the same).

4.3.1 Parameter Estimation

Empirical contact matrices have been described from proxies such as face-to-face conversations or periods sharing spaces. Observed mean per capita numbers of contacts C_{ij} are typically summed over all groups j to yield a_i and then the c_{ij} are calculated by dividing the C_{ij} by a_i. While data may be collected by single year of age, generally they are aggregated into 5-year or larger groups. Because such matrices rarely meet Busenberg and Castillo-Chavez' third condition in (4.3), possibly because study populations are not closed, contacts may be averaged in some way (see, e.g., [9]).

To illustrate these calculations, we collapse observations from the PolyMod study—a survey of face-to-face conversations in eight European countries (that Professor John Edmunds of the London School of Hygiene and Tropical Medicine kindly shared)—into two groups, aged <20 and ≥20 years. There were 7221 participants, $N_1 = 2719$ children and $N_2 = 4502$ adults. They recorded

$$\begin{bmatrix} 24,284 & 17,168 \\ 9,166 & 46,097 \end{bmatrix}$$

face-to-face conversations on an average day with members of their own and the other group (i,j = 1,2). Dividing these daily numbers of conversations with children and adults (the first row) by the number of child participants and those in the second row by the number of adult participants, we obtain average daily per capita contacts,

$$\left(C_{ij}\right) = \begin{bmatrix} 8.93122 & 6.31409 \\ 2.03598 & 10.2392 \end{bmatrix}$$

activities (row sums), $a_1 = 15.2453$, $a_2 = 12.2752$, and mixing matrix (quotients of elements and row sums)

$$C = \left(c_{ij}\right) = \begin{bmatrix} 0.585834 & 0.414166 \\ 0.165861 & 0.834139 \end{bmatrix}.$$

Finally, solving the equations $c_{11} = \varepsilon_1 + (1 - \varepsilon_1) f_1$ and $c_{22} = \varepsilon_2 + (1 - \varepsilon_2) f_2$ simultaneously, we obtain the preferences, $\varepsilon_1 = 0.28321$ and $\varepsilon_2 = 0.607144$.

4.4 Effect on Reproduction Numbers

Incorporating the mixing function (4.4) in the simplest transmission model capable of informing vaccination policy, Feng et al. [2] illustrated the impact of heterogeneity in factors affecting sub-population reproduction numbers and non-random mixing on meta-population reproduction numbers.

Our model population comprises n sub-populations in which people in population i $(0 \le i \le n)$ are divided in susceptible (S_i), infected and infectious (I_i), or removed (R_i) from the infection process by virtue of immunization or immunity following infection. In this model, μ is both the birth and death rate (introducing susceptible people without changing population size), the p_i's are proportions immunized at birth (i.e. products of proportions vaccinated and vaccine efficacy, defined as the conditional probability of being immune given vaccination), the λ_i's are per capita forces (or hazard rates) of infection among susceptible people, and γ is the recovery rate. (See Appendix 3 for a model that is better suited for age groups.) The rates μ and γ are reciprocals of the mean age at death (or life expectancy at birth) and infectious period, respectively. The full model is then:

$$\begin{aligned}
\tfrac{dS_i}{dt} &= \mu N_i\,(1-p_i) - (\lambda_i+\mu)\,S_i\\
\tfrac{dI_i}{dt} &= \lambda_i S_i - (\gamma+\mu)\,I_i,\\
\tfrac{dR_i}{dt} &= \mu N_i p_i + \gamma I_i - \mu R_i,\\
N_i &= S_i + I_i + R_i, \quad i = 1,\ldots,n,
\end{aligned}$$

where λ_i is defined in (4.2). The basic and effective reproduction numbers for sub-population i, denoted by $\Re_{0i}$ and $\Re_{vi}$, respectively, are:

$$\Re_{0i} = \frac{\beta a_i}{\gamma+\mu}, \qquad \Re_{vi} = \Re_{0i}\,(1-p_i)\,.$$

We used the approach of van den Driessche and Watmough [10] to derive the next-generation matrix K for $n=2$ (see Appendix 1 for details). The larger eigenvalue of $K = \begin{bmatrix} \Re_{v1}c_{11} & \Re_{v1}c_{12} \\ \Re_{v2}c_{21} & \Re_{v2}c_{22} \end{bmatrix}$ is $\Re_v = \frac{1}{2}\left[A + D + \sqrt{(A-D)^2 + 4BC}\right]$, where

$$A = \Re_{v1}c_{11}, B = \Re_{v1}c_{12}, C = \Re_{v2}c_{21}, D = \Re_{v2}c_{22}.$$

Dietz [11] was the first of many to show that, when mixing is proportional, $\Re_0$ can be written as a function of the ratio of the variance and mean activity. Table 4.1 illustrates $\Re_0$ from two meta-populations with the same mean activity, but different variances. For the parameter values described, heterogeneity in sub-population activities increases $\Re_0$ from 3.5 to 3.72 for proportional mixing ($\varepsilon_1 = \varepsilon_2 = 0$).

Barbour [12] showed that $\Re_0$ attains its maximum when individuals having high average per capita contact rates mix exclusively with each other. Thus, preferential mixing magnifies the effect of heterogeneity in activity. Moreover, while homogeneous preferential mixing (e.g. $\varepsilon_1 = \varepsilon_2 = 0.5$) further increases $\Re_0$ to 3.88, heterogeneous preferential mixing (e.g. $\varepsilon_1 = 0.25$, $\varepsilon_2 = 0.75$ or vice versa) increases it to 3.92 or 3.98 (Fig. 4.2). Colizza and Vespignani [13] reported a similar result, namely that heterogeneity in connectivity, by which they represent individual movement in a network model, increases the reproduction number.

Table 4.1 Preferential mixing magnifies the impact of heterogeneity in person-to-person contact rates (activities) on $\Re_0$

	Scenario A		Scenario B	
Parameter	$a_1 = 10$	$a_2 = 10$	$a_1 = 7.5$	$a_2 = 12.5$
$\Re_{0i}$	3.5	3.5	2.62	4.37
$\Re_0$ ($\varepsilon_1 = \varepsilon_2 = 0$)	3.5		3.72	
$\Re_0$ ($\varepsilon_1 = \varepsilon_2 = 0.5$)	3.5		3.88	
$\Re_0$ ($\varepsilon_1 = 0.25$, $\varepsilon_2 = 0.75$)	3.5		3.92	
$\Re_0$ ($\varepsilon_1 = 0.75$, $\varepsilon_2 = 0.25$)	3.5		3.98	

The number of sub-populations and their sizes also affect these results, but here $n = 2$, $N_1 = N_2 = 500$, so scenarios A and B have the same mean activity, but differ in the variance. Other parameters: $\beta = 0.05$, $\gamma = 1/7$. Source: Feng et al. [2]

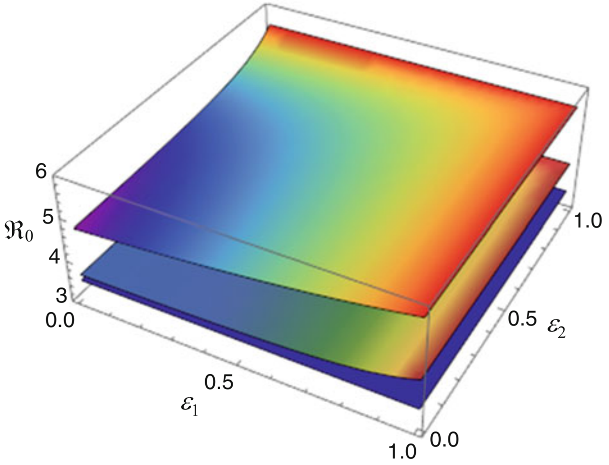


Fig. 4.2 The meta-population $\Re_0$ as a function of fractions of the contacts that members of two sub-populations reserve for others within their own sub-populations (ε_1, ε_2) when their activities (average contact rates) are more or less heterogeneous. $\Re_0$ decreases from the top surface ($a_1 = 4$, $a_2 = 16$), through the middle ($a_1 = 8$, $a_2 = 12$), to the bottom ($a_1 = a_2 = 10$). See Table 4.1 for other parameter values. As heterogeneity in ε increases away from the line $\varepsilon_1 = \varepsilon_2$, heterogeneous preferential mixing also increases $\Re_0$. Source: Feng et al. [2]

Writing $\Re_0$ as a function of (ε_1, ε_2), Feng et al. [2] showed that $\Re_0(\varepsilon_1,\varepsilon_2)$ is an increasing function of ε_1 and ε_2. They also showed that, for $\varepsilon_1 = \varepsilon_2 = 0$ and other parameters the same for both populations, $\Re_0(\varepsilon_1, \varepsilon_2)$ is minimized when $a_1 = a_2 = T/2$, where $T = a_1 + a_2$, and is a monotonically increasing function of the difference $|a_2 - a_1|$ (i.e. $\Re_0$ is maximized when heterogeneity in activity is greatest). A similar result holds when either $\varepsilon_1 = 1$ or $\varepsilon_2 = 1$. That is, $\Re_0(\varepsilon_1, 1) = \Re_0(1, \varepsilon_2)$ is minimized when $a_1 = a_2 = T/2$ and is a monotonically increasing function of the difference $|a_2 - a_1|$. Therefore, $\Re_0(\varepsilon_1, 1) = \Re_0(1, \varepsilon_2)$ is maximized when heterogeneity in activity is greatest.

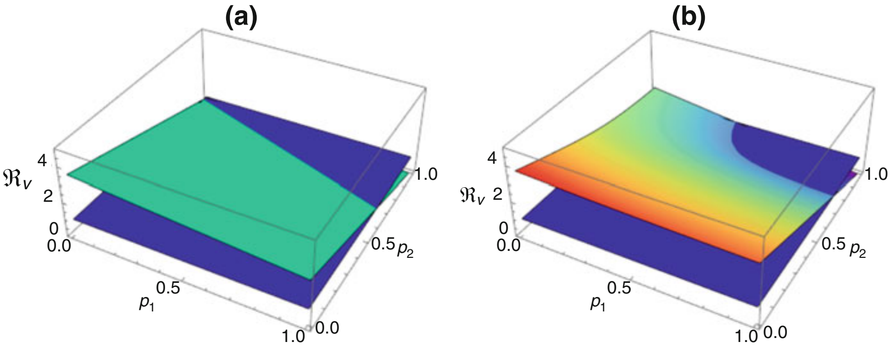


Fig. 4.3 The function $\Re_v$ for scenario B of Table 4.1 with (**a**) proportional and (**b**) preferential mixing. The dark blue planes represent $\Re_v = 1$, and the lighter blue plane and curved (rainbow) surface represent $\Re_v$ at all possible (p_1, p_2) pairs when a) $\varepsilon_1 = \varepsilon_2 = 0$ and b) $\varepsilon_1 = \varepsilon_2 = 0.5$, respectively. $\Re_v \leq 1$ for all combinations of p_i $(i = 1, 2)$ at or below the dark blue plane. See the legend to Table 4.1 for other parameter values. Source: Feng et al. [2]

Figure 4.3 illustrates the impact of heterogeneity in p_i on $\Re_v$ in meta-populations whose sub-populations differ in mean activity (scenario B in Table 4.1), mixing proportionally and preferentially on the left and right, respectively. Heterogeneity increases away from the line connecting the points at which $(p_1, p_2) = 0$ and $(p_1, p_2) = 1$. Values of $\Re_v$ for all combinations of p_i $(i = 1, 2)$ form a plane when mixing is proportional, but curve upward about the above-mentioned line when mixing is preferential. Values at or below their intersection with the dark blue plane, $\Re_v = 1$, are combinations of p_i $(i = 1, 2)$ at which population immunity attains or exceeds this threshold.

These figures explain May and Anderson's [14] observation that "under a uniformly applied immunization programme (i.e. $p_1 = p_2$), the overall fraction that must be immunized is larger than would be estimated by (incorrectly) assuming the population to be homogeneously mixed." While the values of (p_1, p_2) yielding any overall fraction form a straight line, the values of (p_1, p_2) at which $\Re_v = 1$ are curved when mixing is non-random.

A real-world application may be our study of the 2008 measles outbreak in San Diego County [15], where preferential mixing $(\varepsilon_i > 0)$ had a significant effect on the basic reproduction number. When homogeneous mixing was assumed, measles' $\Re_0 \approx 10.7$, whereas when proximity-preferential mixing was considered, the meta-population $\Re_0$ became 18.1 (i.e. 70% greater).

4.4.1 *Vaccination Strategies*

To demonstrate the influence of preferential mixing on the impact of vaccination more explicitly, Chow et al. [16] considered $\Re_v = \Re_v(\varepsilon_1, \varepsilon_2)$ as a function of

ε_1 and ε_2. They defined $\Omega = \{(p_1, p_2) \mid 0 \leq p_1 \leq 1,\ 0 \leq p_2 \leq 1\}$, whereupon each point $(p_1, p_2) \in \Omega$ represents a vaccination strategy. Denoting by Δ_2 the set of all values of ε_1 and ε_2 in [0,1] except $\varepsilon_1 = \varepsilon_2 = 1$ (the case where members of the two sub-populations do not interact), (i.e. $\Delta_2 = \{(\varepsilon_1, \varepsilon_2) \mid 0 \leq \varepsilon_I \leq 1, i = 1,2\} \backslash \{(\varepsilon_1, \varepsilon_2) \mid \varepsilon_1 = \varepsilon_2 = 1\}$), they showed that

$$\frac{\partial \Re_v}{\partial \varepsilon_1} > 0, \quad \frac{\partial \Re_v}{\partial \varepsilon_2} > 0 \text{ for all } (\varepsilon_1, \varepsilon_2) \in \Delta_2.$$

For ease of presentation, they first considered the case where $\varepsilon_1 = \varepsilon_2 = \varepsilon$ and $\Re_v = \Re_{\mathrm{v}}(\varepsilon)$ is a function of ε. Then, for each fixed $\varepsilon \in [0,1)$, the curve $\Re_v(\varepsilon) = 1$ divides the region Ω into $\Omega_\varepsilon = \{(p_1, p_2) \mid 0 \leq \Re_v(\varepsilon) < 1, (p_1, p_2) \in \Omega, 0 \leq \varepsilon < 1\}$, which includes all points above the curve corresponding to $\Re_{\mathrm{v}}(\varepsilon) = 1$, and $\mathrm{D}_\varepsilon = \{(p_1, p_2) \mid \Re_v(\varepsilon) > 1, (p_1, p_2) \in \Omega, 0 \leq \varepsilon < 1\}$, which includes all points below the curve (Fig. 4.4). It can be shown that

$$\Omega_{\tilde{\varepsilon}} \supseteq \Omega_{\hat{\varepsilon}}, \quad D_{\tilde{\varepsilon}} \subseteq D_{\hat{\varepsilon}}, \quad \text{if } 0 < \tilde{\varepsilon} < \hat{\varepsilon} < 1,$$

which implies that, if $\tilde{\varepsilon} < \hat{\varepsilon}$, the curve corresponding to $\Re_v(\tilde{\varepsilon}) = 1$ is below that corresponding to $\Re_v\left(\hat{\varepsilon}\right) = 1$. All such curves intersect at a single point (p_{1c}, p_{2c}) with $p_{1c} = 1 - \frac{1}{\Re_{01}}$, $p_{2c} = 1 - \frac{1}{\Re_{02}}$. Letting $\Omega^* \subseteq \bigcap_{0 \leq \varepsilon < 1} \Omega_\varepsilon$, $D^* \subseteq \bigcap_{0 \leq \varepsilon < 1} D_\varepsilon$, we observe that the region Ω_* (lighter shaded area in Fig. 4.4) is determined by the two inequalities $p_{1c} < p_1 < 1$, $p_{2c} < p_2 < 1$. For region D_* (darker shaded area in Fig. 4.4), the upper bound is determined by the line $p_2 = -Ap_1 + B$, where

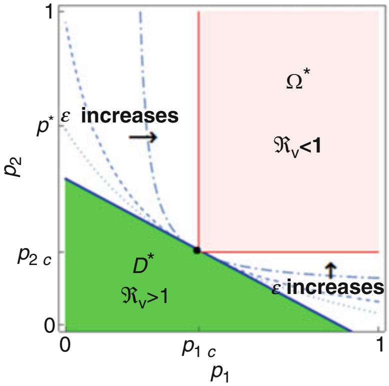


Fig. 4.4 Plot of $\Re_v$ as a function of sub-population immunities, p_1 and p_2. Several curves of $\Re_v(\varepsilon) = 1$ for different ε values are also shown, with the dashed curves corresponding to $0 < \varepsilon < 1$, the thin solid red lines (interior boundary of the region Ω_*) corresponding to $\varepsilon = 1$, and the thick blue line corresponding to $\varepsilon = 0$ (interior boundary of the region D_*). The arrows indicate the direction of change of the curve $\Re_v(\varepsilon) = 1$ as ε increases from 0 to 1. All of the $\Re_v(\varepsilon) = 1$ curves intersect at the single point (p_{1c}, p_{2c}). Source: Chow et al. [16]

$A = \frac{\Re_{01} a_1 N_1}{\Re_{02} a_2 N_2}$ and $B = \frac{(\Re_{01}-1)a_1 N_1 + (\Re_{02}-1)a_2 N_2}{\Re_{02} a_2 N_2}$. The two regions intersect at the point (p_{1c}, p_{2c}).

Chow et al. [16] extended this analysis for $\varepsilon_1 = \varepsilon$ to the case where $\varepsilon_1 \neq \varepsilon_2$ and proved that: (1) if $(p_1, p_2) \in \Omega_*$, then $\Re_v < 1$ for all $(\varepsilon_1, \varepsilon_2) \in \Delta_2$, (2) if $(p_1, p_2) \in D_*$, then $\Re_v > 1$ for all $(\varepsilon_1, \varepsilon_2) \in \Delta_2$, and (3) for every point $(\varepsilon_1, \varepsilon_2) \in \Delta_2$, the curve determined by $\Re_v = 1$ lies in the region $\Omega \backslash (\Omega_* \cup D_*)$. All such curves intersect at a single point, (p_{1c}, p_{2c}). Moreover, these curves have the property that the one corresponding to $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2)$ is lower than that corresponding to $\left(\hat{\varepsilon}_1, \hat{\varepsilon}_2\right)$ if $\tilde{\varepsilon}_1 < \hat{\varepsilon}$ and $\tilde{\varepsilon}_2 < \hat{\varepsilon}_2$.

The first of these results indicates that there is a lower bound for vaccination efforts (p_1, p_2) above which a pathogen can be eliminated regardless of mixing pattern. Similarly, the second result indicates that there is an upper bound for vaccination efforts (p_1, p_2) below which a pathogen cannot be eliminated regardless of the mixing pattern. And the third result indicates that mixing patterns can influence the effect of vaccination on $\Re_v$. Thus, in the design of vaccination programs, one must consider mixing within and between sub-populations.

4.5 Elaborations of the Mixing Model

Given the influence of mixing among heterogeneous sub-populations on meta-population reproduction numbers and inspired by empirical observations [17–20], Glasser et al. [21] further elaborated Nold's model to include preferential contacts between parents and children and among co-workers as well as contemporaries (i.e. people similar in age) by defining:

$$c_{ij} = \phi_{ij} + \left(1 - \sum_{l=1}^{4} \varepsilon_{li}\right) f_j, \quad f_j = \frac{\left(1 - \sum_{l=1}^{4} \varepsilon_{lj}\right) a_j N_j}{\sum_{k=1}^{n} \left(1 - \sum_{l=1}^{4} \varepsilon_{lk}\right) a_k N_k}.$$

Because the sub- and super-diagonals extend over ages $i > G$ and $i < L - G$, respectively, where G is the generation time (i.e. average age at which women bear daughters), L is longevity (i.e. average age at death or expectation of life at birth), and $L > G$, they define ϕ_{ij} as

$$\phi_{ij} := \begin{cases} \delta_{ij}\varepsilon_{1i} + \delta_{i(j+G)}\varepsilon_{2i} + I_{W_{\min} \le i, j \le W_{\max}} \frac{\varepsilon_{4i}}{W_{\max} - W_{\min}}, & i > G \\ \delta_{ij}\varepsilon_{1i} + \delta_{i(j-G)}\varepsilon_{3i} + I_{W_{\min} \le i, j \le W_{\max}} \frac{\varepsilon_{4i}}{W_{\max} - W_{\min}}, & i < L - G \end{cases}.$$

If age classes are 0–4, 5–9, ... and $G = 25$ years, $i > G$ means $i >$ class 5. $W_{\min}$ and $W_{\max}$ ($W_{\min} < W_{\max}$) are the average ages at entry to and exit from the workforce, $\varepsilon_{1i}, \ldots, \varepsilon_{4i}$ are the fractions of contacts reserved for contemporaries, children ($j - G$), parents ($j + G$), and co-workers (if $W_{\min} \le i, j \le W_{\max}$), respectively,

$$\delta_{i(j\pm G)} = \begin{cases} 1 & \text{if } i = j \pm G \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad I_{W_{\min}\le i,j\le W_{\max}} = \begin{cases} 1 & \text{if } W_{\min} \le i, j \le W_{\max} \\ 0 & \text{otherwise.} \end{cases}$$

Because of the third condition in (4.3), the non-zero elements of ε_2 and ε_3 are related. If $G = 25$ years, for example, then $a_i\, N_i\, \varepsilon_{2i} = a_j\, N_j\, \varepsilon_{3j}$, for $i = 6, 7, \ldots, j = i - 5$. Accordingly, we can estimate ε_{3i} by assuming that $\varepsilon_{2i} = a_j N_j \varepsilon_{3j} / a_i N_i$. Notice also that $0 \le \sum_{l=1}^{4} \varepsilon_{li} < 1$ and that mixing among co-workers does not depend on age provided that $i \ge W_{\min}$ and $j \le W_{\max}$.

4.5.1 Gaussian Kernels

While delta formulations are undeniably heuristic, contemporaries need not be exactly the same age [22], nor need the ages of parents and children differ by exactly the generation time. Accordingly, we reformulated ϕ_{ij} to incorporate this more realistic feature. Let α and α' denote the ages of susceptible and infected individuals, respectively. Further, let $a(\alpha)$ denote the average number of contacts per capita aged α per unit of time, $N(\alpha)$ denote the number of people aged α, and $I_{[W_{\min}, W_{\max}]}\left(\alpha, \alpha'\right)$ denote the function

$$I_{[W_{\min}, W_{\max}]}\left(\alpha, \alpha'\right) = \begin{cases} 1, & \text{if } W_{\min} \le \alpha, \alpha' \le W_{\max} \\ 0, & \text{otherwise.} \end{cases}$$

Then the continuous analogue of c_{ij} can be formulated as:

$$c\left(\alpha, \alpha'\right) = \phi\left(\alpha, \alpha'\right) + \left[1 - \sum_{l=1}^{4} \varepsilon_l\left(\alpha\right)\right] f\left(\alpha'\right)$$
$$f\left(\alpha'\right) = \frac{\left[1 - \sum_{l=1}^{4} \varepsilon_l\left(\alpha'\right)\right] a\left(\alpha'\right) N\left(\alpha'\right)}{\int_0^\infty \left[1 - \sum_{l=1}^{4} \varepsilon_l(u)\right] a(u) N(u)\, du},$$

where

$$\phi\left(\alpha, \alpha'\right) = \begin{cases} g_1\left(\alpha, \alpha'\right) \varepsilon_1\left(\alpha\right) + g_2\left(\alpha, \alpha'\right) \varepsilon_2\left(\alpha\right) \\ \qquad + I_{[W_{\min}, W_{\max}]}\left(\alpha, \alpha'\right) \frac{\varepsilon_4(\alpha)}{W_{\max} - W_{\min}}, & \alpha > G \\ g_1\left(\alpha, \alpha'\right) \varepsilon_1\left(\alpha\right) + g_3\left(\alpha, \alpha'\right) \varepsilon_3\left(\alpha\right) \\ \qquad + I_{[W_{\min}, W_{\max}]}\left(\alpha, \alpha'\right) \frac{\varepsilon_4(\alpha)}{W_{\max} - W_{\min}}, & \alpha < L - G \end{cases},$$

with

$$g_1\left(\alpha,\alpha'\right)=\frac{1}{\sqrt{2\pi}\sigma_1(\alpha)}e^{-\frac{\left(\alpha'-\alpha\right)^2}{2[\sigma_1(\alpha)]^2}},$$
$$g_2\left(\alpha,\alpha'\right)=\frac{1}{\sqrt{2\pi}\sigma_2(\alpha)}e^{-\frac{\left[\alpha'-(\alpha-G)\right]^2}{2[\sigma_2(\alpha)]^2}},$$
$$g_3\left(\alpha,\alpha'\right)=\frac{1}{\sqrt{2\pi}\sigma_3(\alpha)}e^{-\frac{\left[\alpha'-(\alpha+G)\right]^2}{2[\sigma_3(\alpha)]^2}}.$$

Here the $g_k(\alpha,\alpha^{'})$ ($k = 1, 2, 3$) are Gaussian kernels with standard deviations $\sigma_k(\alpha)$. Besides the above-mentioned relationship between $\varepsilon_2(\alpha)$ and $\varepsilon_3(\alpha - G)$, for each α, we have that $0 \le \sum_{i=1}^{4}\varepsilon_l(\alpha) < 1$.

We fit a hybrid of these formulations (i.e. the discrete formulation with Gaussian kernels instead of deltas) to observations from the above-mentioned empirical studies, which are discrete, using the FindMinimum function in *Mathematica*™. This amounts to choosing ε_{li} and σ_{ki}, as well as G, L, and the W's, that minimize an objective function, here the mean squared error. With one starting value for each variable, FindMinimum uses BFGS quasi-Newton methods. When there are constraints, FindMinimum uses interior point methods. It was necessary to constrain the parameter vectors $\overrightarrow{\varepsilon}_l$ and $\overrightarrow{\sigma}_k$ so that the main diagonal does not dominate, and to fix G, L, $W_{\min}$, and $W_{\max}$ after convergence.

Figure 4.5 illustrates fits of this model with 5-year age classes (0–4, 5–9, ..., 70+ years) to weighted averages of casual and physical contacts from the eight European countries involved in the PolyMod study. Feng et al. [23] added preferential contacts between grandparents and grandchildren, but—to facilitate parameter estimation—omitted workplace contacts (Fig. 4.6). Motivated by concern about the net benefit of a proposed pandemic mitigation measure, prolonged school closures, they fitted this model to gender-stratified observations from the same study.

4.5.2 A Spatial Model

To explore the impact of heterogeneity in vaccine coverage due to personal-belief exemptions to vaccination on the potential for outbreaks of vaccine-preventable childhood diseases, Glasser et al. [15] developed a spatial model.

Reasoning that, in spatially stratified populations, proximity must affect contacts, we defined the average per capita contact rate or activity of children attending elementary school i as a negative exponential function of inter-school distances; that is:

$$a_i := \sum_j \exp\left(-bd_{ij}\right), \tag{4.5}$$

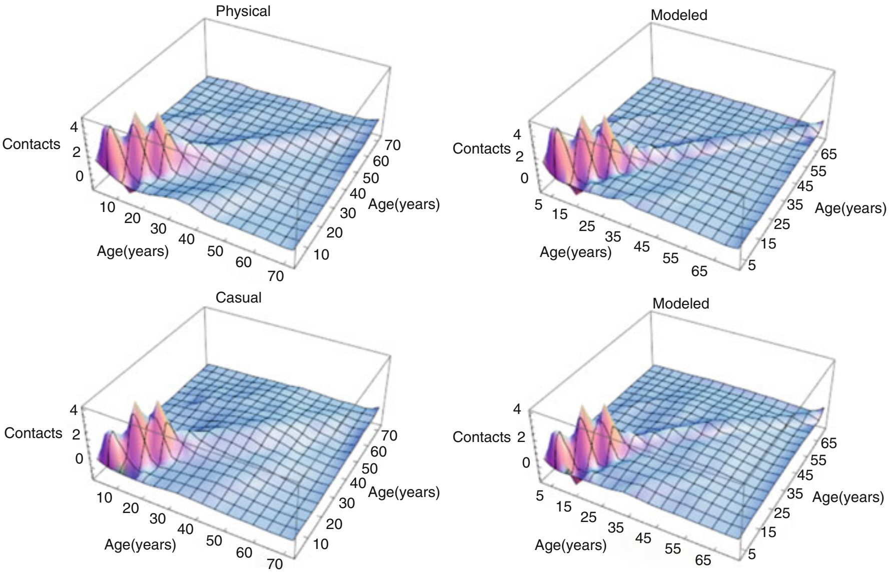


Fig. 4.5 Left: Average daily per capita numbers of physical and casual (all contacts less physical ones) contacts (top and bottom, respectively) from Mossong et al. [18]. Right: The mixing function introduced by Glasser et al. [21] fitted to the observations on the left. Values of the fitted parameters ($n = 15$) are in Appendix 2. Source: Glasser et al. [21]

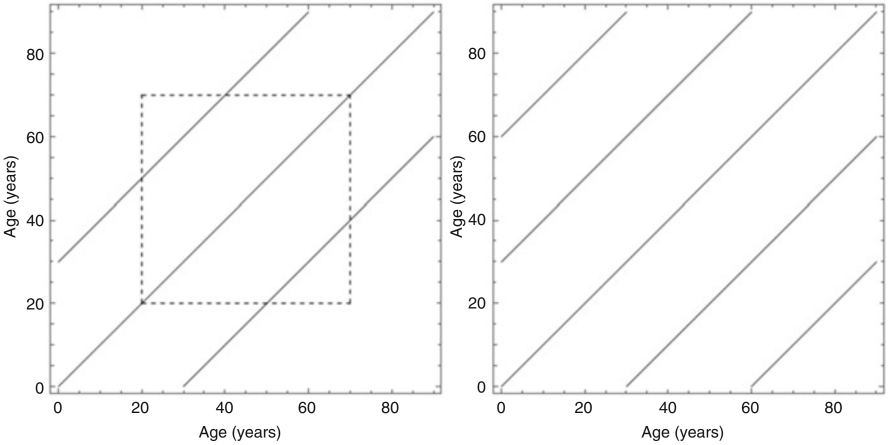


Fig. 4.6 Generalizations of the function of Nold [8] and Jacquez et al. [7], which allows fractions of contacts to be reserved for one's own group and complements to be distributed proportionally among groups. The age-specific function on the left [21] includes preferential contacts between parents and children (sub- and super-diagonals) and among co-workers (dashed box) as well as contemporaries (main diagonal) while that on the right includes preferential contacts with grandparents and grandchildren (sub-sub- and super-super-diagonals) as well as parents, children, and contemporaries. Source: Feng et al. [23]

where b is the rate at which contacts diminish with distance and the d_{ij}'s are distances between school i and all others.

Because most children attend elementary school in their own neighborhood, elementary schools are proxies for neighborhoods. To obtain total contacts, we multiplied these rates by school enrollments, N, and to obtain proportions of contacts with children in any school, we divided by the sum of contacts over all schools:

$$c_{ij} := \frac{N_j \exp\left(-bd_{ij}\right)}{\sum_k N_k \exp\left(-bd_{ik}\right)}, \tag{4.6}$$

whereupon the c_{ij} are proportions of contacts that children in school i have with children in all schools including their own. Consequently, complements of the proportions of contacts that are intra-school (or neighborhood), 1–c_{ii}, are interpretable as *connectedness* (sensu strength of connections with other locations).

Using this model (Fig. 4.7), together with information about vaccination from school-entry surveys during 2008 in San Diego County, when and where a measles outbreak occurred in a school having ~30% of children with personal-belief exemptions to vaccination, we calculated the meta-population reproduction numbers of measles, mumps, and rubella. As mentioned, these are the dominant eigenvalues associated with next-generation matrices, in turn products of diagonal matrices, whose elements are basic or effective sub-population reproduction numbers, and the contact matrix.

For each of these eigenvalues $\Re$, the corresponding next-generation matrix K has associated nonzero right and left eigenvectors $\overrightarrow{v}_R$ and $\overrightarrow{v}_L$, respectively. That is, $K\overrightarrow{v}_R = \lambda_R\overrightarrow{v}_R$ and $\overrightarrow{v}_L K = \lambda_L\overrightarrow{v}_L$, where λ_R and λ_L are constants with $\lambda_R = \Re$. Both $\overrightarrow{v}_R$ and $\overrightarrow{v}_L$ have biological interpretations: $\overrightarrow{v}_R$ is the prevalence

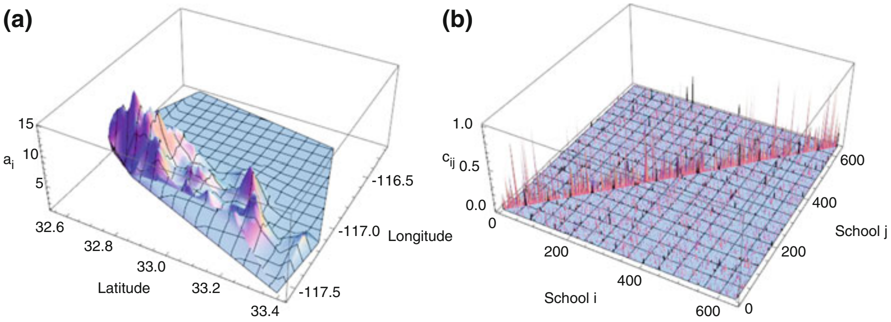


Fig. 4.7 Mixing among elementary schoolchildren in San Diego County ($n = 638$ schools). The peaks and valleys of the activity surface (**a**) indicate the nearby or isolated schools or neighborhoods characterizing the more and less densely populated coastal and eastern regions. Because rows of the mixing matrix (**b**) sum to one, children in schools or neighborhoods with larger diagonal elements (i.e. more of their contacts within schools) have smaller enrollments or are more isolated while ones with smaller diagonal elements (i.e. greater proportions of their contacts between schools) are larger or more highly interconnected. Source: Glasser et al. [15]

of infection by sub-population and $\overrightarrow{v}_L$ is their respective contributions to the reproduction number [24]. Figures 4.8 and 4.9 plot neighborhood contributions to measles reproduction numbers in space.

With appropriate parameter values for measles, but an arbitrary value of *b,* Figs. 4.8a, b and 4.9 illustrate the left eigenvectors of the next-generation matrices corresponding to reproduction numbers of 25.9, 2.8, and 1.5, respectively. Although measles' $\Re_0$ in San Diego County is unknown, evidently routine vaccination reduces

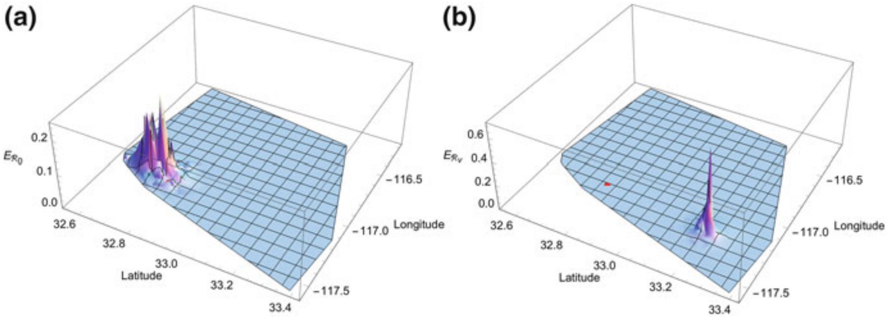


Fig. 4.8 Spatial plots of left eigenvectors associated with the dominant eigenvalues of next-generation matrices whose elements are sub-population (**a**) basic and (**b**) effective reproduction numbers. The school where the outbreak occurred, indicated by a red dot in (**b**), is surrounded by schools with more highly vaccinated populations. Source: Glasser et al. [15]

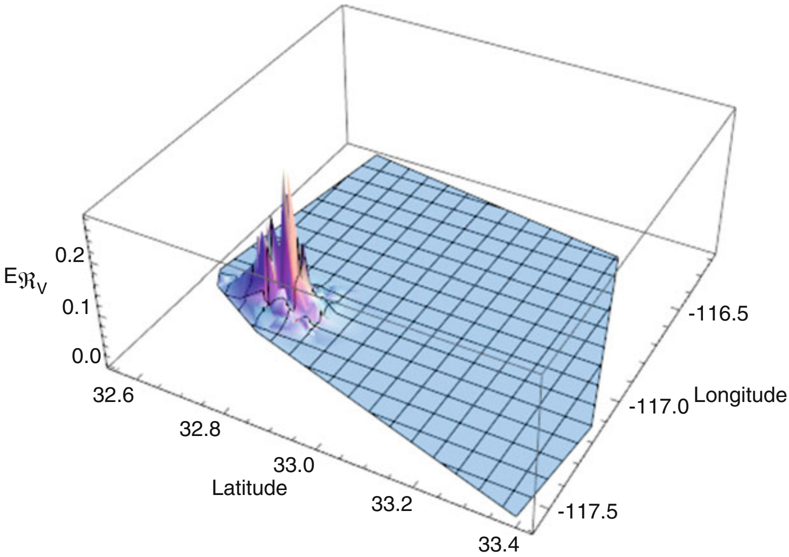


Fig. 4.9 The impact of vaccinating the same proportion of children with personal-belief exemptions as others in each school ($n = 638$) on the residual outbreak potential illustrated in Fig. 4.8b. Source: Glasser et al. [15]

$\Re_v$ by an order of magnitude (cf. Fig. 4.8a, b). And eliminating personal-belief exemptions, as has since been done in California, reduces $\Re_v$ even further (cf. Figs. 4.8b and 4.9).

4.6 Two-Level Mixing

Mixing may be a multi-dimensional phenomenon, but so far we have considered only single dimensions (e.g. age, gender, or location). Based on fitting their inter-generational mixing function to gender-stratified PolyMod observations, for example, Feng et al. [23] argued that, in the event of prolonged school closures, some working parents with young children would involve grandmothers in child-care. As hospitalizations and mortality increase with age, they believe that this proposed pandemic mitigation measure warrants more careful scrutiny to ensure that it is consistent with public health policy goals.

Other applications also require models with multiple strata. Beginning with two, consider a meta-population with m spatial locations (or other characteristics such as gender) and n classes (e.g., age or activity groups). Let l_i denote the ith location (l for location) and a_j denote the jth age group (a for age), $1 \le i \le m$ and $1 \le j \le n$. We use this compound notation whenever indices might otherwise be confused.

In our first multi-level model, we combine the models of Glasser et al. [15] and Nold [8], as modified by Jacquez et al. [7], defining

$$c_{l_i a_j l_p a_q} := \frac{c^{(p)}_{a_j a_q} e^{-bd_{l_i l_p}}}{\sum_{r=1}^{n} \sum_{s=1}^{m} c^{(p)}_{a_j a_r} e^{-bd_{l_i l_s}}}, \quad 1 \le i, p \le m, \quad \text{and} \quad 1 \le j, q \le n,$$

$$\text{where} \quad c^{(p)}_{a_j a_q} = \varepsilon_{a_j} \delta_{a_j a_q} + \left(1 - \varepsilon_{a_j}\right) F_{l_p a_q} \quad \text{and} \quad F_{l_p a_q} = \frac{\left(1-\varepsilon_{a_q}\right) A_{l_p a_q} N_{l_p a_q}}{\sum_{k=1}^{n} \left(1-\varepsilon_{a_k}\right) A_{l_p a_k} N_{l_p a_k}}.$$

In these expressions, b and $d_{l_i l_p}$ are as defined earlier, $F_{l_p a_q}$ corresponds to proportional mixing (with respect to age) among persons in group q at location p, ε_{a_q} denotes the fraction of contacts that individuals aged q (at any location) reserve for others in the same group (preference), and $c^{(p)}_{a_j a_q}$ represents the fraction of their contacts that individuals aged j have with individuals aged a_q at location l_p.

This two-level mixing function was employed by Feng et al. [23], who used a meta-population model with spatial- and age-structure to compare the actual monthly and optimal vaccination strategies (i.e. allocations of available vaccine among seven age groups) in 51 locations in the United States (i.e. 50 states plus the District of Columbia) during the 2009 H1N1 pandemic. Another example of using meta-population model with the two-level mixing can be found in Hao et al. [25], in which the model is used to identify optimal vaccination strategies for measles elimination in China.

4.6.1 Parameter Estimation

Given information about face-to-face conversations or other proxies for daily contacts, we first calculate the activities $A_{a_q} = \sum_{a_j} C_{a_q a_j}$. Insofar as people are mobile, however, these A_{a_q} average over m locations, i.e. $\overline{A}_{a_q} = \frac{1}{m}\sum_{i=1}^{m} A_{l_i a_q}$. Assuming that the activity of an individual aged a_j at location l_i, $A_{l_i a_j}$, depends not only on his/her age, but also on the distance, ease of travel, ... to other locations, $A_{l_i a_j} = \sum_{k=1}^{m} e^{-b_{a_j} d_{l_i l_k}}$, whereupon $\overline{A}_{a_q} = \frac{1}{m}\sum_{r=1}^{m}\sum_{k=1}^{m} e^{-b_{a_q} d_{l_r l_k}}$. Thus, if $\overline{A}_{a_q}$ and $d_{l_i l_j}$are known, the b_{a_q} can be estimated for $q = 1, 2, \ldots, n$. Figure 4.10 shows that the rate at which contacts decline with distance is indeed age-dependent. Mobility increases to a maximum during adolescence, plateaus at a lower level during the reproductive years and declines increasingly afterwards. Given those rates, we can obtain the $A_{l_i a_j}$, from which we can obtain $F_{l_i a_j}$, $c^{(p)}_{a_j a_q}$, and $c_{l_i a_j l_p a_q}$.

4.6.2 A General Scheme

In an effort to develop a template for multi-level mixing, Feng et al. [23] described the probability of contact between persons in location l_i, age a_j and location l_p, age a_q by a matrix with entries

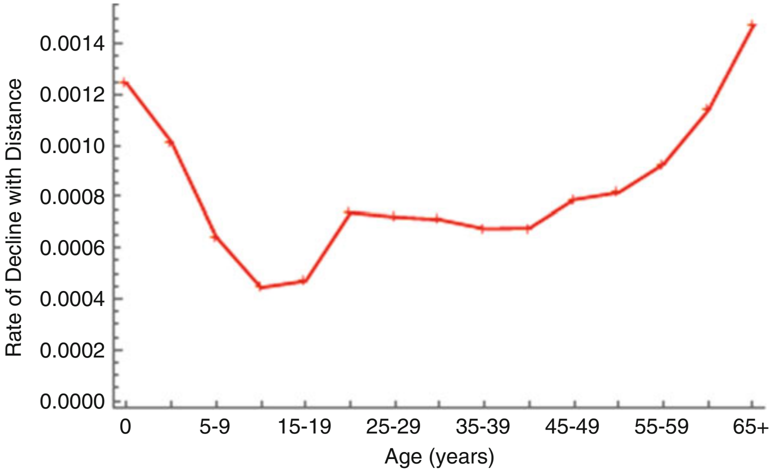


Fig. 4.10 Initially, the spatial range of contacts is small, but it increases throughout childhood to a maximum during adolescence, declines to a plateau during the childbearing and working years and finally decreases further during old age. Source: Hao et al. [25]

$$c_{l_ia_jl_pa_q} := \varepsilon_{l_ia_j}\delta_{l_il_p}\delta_{a_ja_q} + \left(1-\varepsilon_{l_ia_j}\right) f_{l_pa_q}, \quad 1 \le i, p \le m, \quad 1 \le j, q \le n,$$

where

$$f_{l_pa_q} := \frac{\left(1-\varepsilon_{l_pa_q}\right) A_{l_pa_q} N_{l_pa_q}}{\sum_{j=1}^{n}\sum_{i=1}^{m}\left(1-\varepsilon_{l_ia_j}\right) A_{l_ia_j} N_{l_ia_j}} \ .$$

In these expressions, $\varepsilon_{l_ia_j}$ represents preference for one's own age/location group, δ_{rs} is the Kronecker delta function, taking values of 1 (if $r = s$) or 0 (if $r \neq s$), and $f_{l_pa_q}$ is random mixing (i.e. proportional to contacts not reserved for others in one's own group, $\left(1-\varepsilon_{l_pa_q}\right) A_{l_pa_q} N_{l_pa_q}$). For most applications, however, mixing among ages and locations (or other strata) is independent (e.g. members of an age class may contact others of the same age preferentially regardless of their location, gender, or any other discrete characteristic).

Letting $\varepsilon^{(l)}_{l_ia_j}$ and $\varepsilon^{(a)}_{l_ia_j}$ represent preferences for one's own location and age class, respectively, matrix entries become

$$\begin{aligned} c_{l_ia_jl_pa_q} := {} & \varepsilon^{(l)}_{l_ia_j}\delta_{l_il_p}\left[\varepsilon^{(a)}_{l_ia_j}\delta_{a_ja_q} + \left(1-\varepsilon^{(a)}_{l_ia_j}\right) F_{l_ia_q}\right] \\ & + \left(1-\varepsilon^{(l)}_{l_ia_j}\right)\left[\varepsilon^{(a)}_{l_ia_j}\delta_{a_ja_q}G_{l_pa_q} + \left(1-\varepsilon^{(a)}_{l_ia_j}\right) H_{l_pa_q}\right], \quad 1 \le i, p \le m, \ 1 \le j, q \le n, \end{aligned}$$

where

$$F_{l_ia_q} = \frac{\left[1-\varepsilon^{(a)}_{l_ia_q}\right] A_{l_ia_q}N_{l_ia_q}}{\sum_k \left[1-\varepsilon^{(a)}_{l_ia_k}\right] A_{l_ia_k}N_{l_ia_k}}, \quad G_{l_pa_q} = \frac{\left[1-\varepsilon^{(l)}_{l_pa_q}\right] A_{l_pa_q}N_{l_pa_q}}{\sum_r \left[1-\varepsilon^{(l)}_{l_ra_q}\right] A_{l_ra_q}N_{l_ra_q}},$$

$$\text{and} \quad H_{l_pa_q} = \frac{\left[1-\varepsilon^{(a)}_{l_pa_q}\right]\left[1-\varepsilon^{(l)}_{l_pa_q}\right] A_{l_pa_q}N_{l_pa_q}}{\sum_r\sum_k \left[1-\varepsilon^{(a)}_{l_ra_k}\right]\left[1-\varepsilon^{(l)}_{l_ra_k}\right] A_{l_ra_k}N_{l_ra_k}}.$$

In this expression for $c_{l_ia_jl_pa_q}$, the terms in square brackets represent age-preferential mixing in one's own and other locations, respectively, and $F_{l_pa_q}$, $G_{l_pa_q}$, and $H_{l_pa_q}$ represent proportional mixing with respect to age, location, and both.

Checking to ensure that $\sum\limits_{p=1}^{m}\sum\limits_{q=1}^{n} c_{l_ia_jl_pa_q} = 1$ for any given i and j, we find that

$$\sum_{p=1}^{m}\sum_{q=1}^{n} \varepsilon^{(l)}_{l_ia_j}\delta_{l_il_p}\left[\varepsilon^{(a)}_{l_ia_j}\delta_{a_ja_q} + \left(1-\varepsilon^{(a)}_{l_ia_j}\right) F_{l_ia_q}\right] = \varepsilon^{(l)}_{l_ia_j}\left[\varepsilon^{(a)}_{l_ia_j} + \left(1-\varepsilon^{(a)}_{l_ia_j}\right)\sum_q F_{l_ia_q}\right] = \varepsilon^{(l)}_{l_ia_j},$$

$$\sum_{p=1}^{m}\sum_{q=1}^{n}\left[\varepsilon^{(a)}_{l_ia_j}\delta_{a_ja_q}G_{l_pa_q} + \left(1-\varepsilon^{(a)}_{l_ia_j}\right) H_{l_pa_q}\right] = \sum_{p=1}^{m}\varepsilon^{(a)}_{l_ia_j}G_{l_pa_j} + \sum_{p=1}^{m}\sum_{q=1}^{n}\left(1-\varepsilon^{(a)}_{l_ia_j}\right) H_{l_pa_q} = 1.$$

We can also verify that the balance condition

$$A_{l_ia_j}N_{l_ia_j}c_{l_ia_jl_pa_q} = A_{l_pa_q}N_{l_pa_q}c_{l_pa_ql_ia_j}, \quad i \neq p, \quad j \neq q,$$

is satisfied:

$$\begin{aligned}A_{l_i a_j} N_{l_i a_j} c_{l_i a_j l_p a_q} &= A_{l_i a_j} N_{l_i a_j} \left(1-\varepsilon^{(l)}_{l_i a_j}\right)\left(1-\varepsilon^{(a)}_{l_i a_j}\right) H_{l_p a_q}\\ &= A_{l_i a_j} N_{l_i a_j} \left(1-\varepsilon^{(l)}_{l_i a_j}\right)\left(1-\varepsilon^{(a)}_{l_i a_j}\right) \frac{\left(1-\varepsilon^{(a)}_{l_p a_q}\right)\left(1-\varepsilon^{(l)}_{l_p a_q}\right) A_{l_p a_q} N_{l_p a_q}}{\sum_r \sum_k \left[1-\varepsilon^{(a)}_{l_r a_k}\right]\left[1-\varepsilon^{(l)}_{l_r a_k}\right] A_{l_r a_k} N_{l_r a_k}}\end{aligned}$$

and

$$\begin{aligned}A_{l_p a_q} N_{l_p a_q} c_{l_p a_q l_i a_j} &= A_{l_p a_q} N_{l_p a_q} \left(1-\varepsilon^{(a)}_{l_p a_q}\right)\left(1-\varepsilon^{(l)}_{l_p a_q}\right) H_{l_i a_j}\\ &= A_{l_p a_q} N_{l_p a_q} \left(1-\varepsilon^{(a)}_{l_p a_q}\right)\left(1-\varepsilon^{(l)}_{l_p a_q}\right) \frac{\left(1-\varepsilon^{(l)}_{l_i a_j}\right)\left(1-\varepsilon^{(a)}_{l_i a_j}\right) A_{l_i a_j} N_{l_i a_j}}{\sum_r \sum_k \left[1-\varepsilon^{(a)}_{l_r a_k}\right]\left[1-\varepsilon^{(l)}_{l_r a_k}\right] A_{l_r a_k} N_{l_r a_k}}.\end{aligned}$$

Once we have an expression for $c_{l_i a_j l_p a_q}$ that is suitable for our application, we can formulate the force or hazard rate of infection per susceptible person as

$$\lambda_{l_i a_j} = A_{l_i a_j} \beta_{l_i a_j} \sum_{p=1}^{m} \sum_{q=1}^{n} c_{l_i a_j l_p a_q} \left(\frac{I_{l_p a_q}}{N_{l_p a_q}}\right), \quad 1 \le i \le m, \quad 1 \le j \le n.$$

4.6.3 Two Examples

Feng et al. [23] provide examples to indicate the flexibility of this template. The main advantage of using it, versus developing ad hoc mixing models, is that contacts will balance (i.e. $C_{ij} = C_{ji}$, $i,j = 1, \ldots, n$).

4.6.3.1 Immigrants and Natives

Consider the case of immigrant ($l_1 = 1$) and native ($l_2 = 2$) populations, a distinction that may matter for models designed to evaluate interventions to mitigate diseases whose prevalence differs at home and abroad (e.g. tuberculosis). The preference for population 1 of individuals aged a_j in population 1 is $\varepsilon^{(l)}_{1a_j}$; similarly, the preference for population 2 of individuals in population 2 is $\varepsilon^{(l)}_{2a_j}$.

If there is no age preference $\left(\varepsilon^{(a)}_{l_i a_j} = 0\right)$, the probability that individuals aged a_j in population 1 contact persons aged a_q in population 1 (note that, in this case, $\delta_{l_1 l_1} = \delta_{11} = 1$) is:

$$\begin{aligned}c_{1a_j 1a_q} &= \varepsilon^{(l)}_{1a_j} F_{1a_q} + \left(1-\varepsilon^{(l)}_{1a_j}\right) H_{1a_q}\\ &= \varepsilon^{(l)}_{1a_j} \frac{A_{1a_q} N_{1a_q}}{\sum_k A_{1a_k} N_{1a_k}} + \left(1-\varepsilon^{(l)}_{1a_j}\right) \frac{\left[1-\varepsilon^{(l)}_{1a_q}\right] A_{1a_q} N_{1a_q}}{\sum_r \sum_k \left[1-\varepsilon^{(l)}_{l_r a_k}\right] A_{l_r a_k} N_{l_r a_k}}.\end{aligned}$$

Similarly, the probability that individuals aged a_j in population 1 contact persons aged a_q in population 2 (note that, in this case, $\delta_{l_1 l_2} = \delta_{12} = 0$) is:

$$c_{1a_j 2a_q} = \left(1 - \varepsilon_{1a_j}^{(l)}\right) H_{2a_q} = \left(1 - \varepsilon_{1a_j}^{(l)}\right) \frac{\left[1 - \varepsilon_{2a_q}^{(l)}\right] A_{2a_q} N_{2a_q}}{\sum_r \sum_k \left[1 - \varepsilon_{l_r a_k}^{(l)}\right] A_{l_r a_k} N_{l_r a_k}}.$$

4.6.3.2 Sexual Contacts

Another case with $m = 2$ is the *age- or activity-stratified mixing* between females ($l_1 = 1$) and males ($l_2 = 2$), most of whose contacts are reserved for members of the other gender. (Replacing age with sexual activity, the groups could become sex workers and their clients.)

If contacts are entirely heterosexual, $\varepsilon_{1a_j}^{(l)} = \varepsilon_{2a_q}^{(l)} = 0$, $0 \leq \varepsilon_{1a_j}^{(a)}, \varepsilon_{2a_q}^{(a)} < 1$, j, $q = 1, \ldots, n$. Thus, F is irrelevant. And, if there are no contacts within l_1 and l_2, the denominator in G should not be a sum, whereupon $G = 1$. Similarly, the sum over r in the denominator of H should be omitted. That is,

$$G_{1a_q} = G_{2a_q} = 1, \quad \text{and } H_{2a_q} = \frac{\left[1 - \varepsilon_{2a_q}^{(a)}\right] A_{2a_q} N_{2a_q}}{\sum_k \left[1 - \varepsilon_{2a_k}^{(a)}\right] A_{2a_k} N_{2a_k}},$$

whereupon $c_{1a_j 2a_q} = \varepsilon_{1a_j}^{(a)} \delta_{a_j a_q} + \left(1 - \varepsilon_{1a_j}^{(a)}\right) H_{2a_q}$. Thus, the hazard rate of infection for a female aged j is

$$\lambda_{1a_j} = \beta_{1a_j} A_{1a_j} \sum_{q=1}^{n} c_{1a_j 2a_q} \left(\frac{I_{2a_q}}{N_{2a_q}}\right).$$

4.7 Questions for Future Research

In no empirical dataset is mixing proportional, yet we often assume that it is to reduce continuous formulations (of, e.g., forces of infection) to familiar discrete ones or systems of PDEs to ODEs or to derive explicit formulae for reproduction numbers.

Consider reducing the continuous formulation of the force of infection in the model of Glasser et al. [21] to discrete. When mixing is proportional (i.e. $\varepsilon(\alpha) = 0$), $c(\alpha,\alpha') = f(\alpha')$. In this case, assuming that $a(\alpha)$ is piecewise constant in age group i, then

$$\int_0^\infty a(u)N(u)du = \sum_{k=1}^{n} a_k \int_{\alpha_{k-1}}^{\alpha_k} N(u)du = \sum_{k=1}^{n} a_k N_k,$$

where $N_k = \int_{a_{k-1}}^{a_k} N(u)du$. Thus, the force of infection for age group i becomes:

$$\begin{aligned}
\lambda_i &= a_i \beta_i \int_0^{\alpha_{\max}} f\left(\alpha'\right)\left[I\left(\alpha'\right)/N\left(\alpha'\right)\right] d\alpha' \\
&= a_i \beta_i \int_0^{\alpha_{\max}} \frac{a(\alpha')N(\alpha')}{\int_0^{\infty} a(u)N(u)du}\left[I\left(\alpha'\right)/N\left(\alpha'\right)\right] d\alpha' \\
&= a_i \beta_i \frac{1}{\sum\limits_{k=1}^{n} a_k \int_{\alpha_{k-1}}^{\alpha_k} N(u)du} \sum_{j=1}^{n} \int_{\alpha_{j-1}}^{\alpha_j} a_j N\left(\alpha'\right)\left[I\left(\alpha'\right)/N\left(\alpha'\right)\right] d\alpha' \\
&= a_i \beta_i \frac{1}{\sum\limits_{k=1}^{n} a_k N_k} \sum_{j=1}^{n} a_j \int_{\alpha_{j-1}}^{\alpha_j} I\left(\alpha'\right) d\alpha' = a_i \beta_i \frac{1}{\sum\limits_{k=1}^{n} a_k N_k} \sum_{j=1}^{n} a_j I_j \\
&= a_i \beta_i \sum_{j=1}^{n} \frac{a_j N_j}{\sum\limits_{k=1}^{n} a_k N_k}\left(\frac{I_j}{N_j}\right) = a_i \beta_i \sum_{j=1}^{n} c_{ij}\left(\frac{I_j}{N_j}\right).
\end{aligned}$$

What is the problem when the mixing is not proportional? Note that, when $\sigma(\alpha)$ is constant within age group i, the functions g_k are also constant in those age groups. Then, when $\varepsilon(\alpha)$ is piecewise constant, the function $\phi(\alpha,\alpha')$ will also be piecewise constant. That is, the $\phi(\alpha,\alpha')$ does not include a factor $N(\alpha')$, which cancels the $N(\alpha')$ in the denominator $I(\alpha')/N(\alpha')$ in going from step 3 to step 4 above.

To obtain an explicit expression for the meta-population reproduction number whose partial derivative could be calculated, in another example, Feng et al. [23] assumed that mixing was proportional. Feng et al. [26] incorporate more realistic mixing functions into these calculations. More work of that sort is needed.

In ongoing work on measles and rubella in China, our estimates of $\Re_0$ are greater from models with spatial as well as age structure than from models with age structure alone. Is this a general phenomenon (i.e. the more kinds of heterogeneity, the higher the basic reproduction number)?

Empirical contact matrices rarely balance. Assuming that this reflects open study populations, various ad hoc averaging schemes have been used to adjust them. Is that the explanation for such imbalance? If so, which scheme is best? If not, does contact matrix asymmetry contain useful information? If so, what (e.g. that study participants are more likely to record contacts that they initiated)?

If the negative exponential function of inter-location distances is weighted by sub-population sizes [15], contacts won't balance. Similarly, if the rate at which contacts diminish with distance varies with age (Fig. 4.10), they won't balance in our two-level age/space model either. How can these realistic features be included in spatial mixing models without violating the balance condition? One possible improvement is to keep the same activity expression (4.5) but re-define the contact matrix (4.6) as $c_{ij} = \frac{a_j N_j}{\sum_r a_r N_r}$, i.e.,

$$c_{ij} = \frac{N_j \sum_k \exp\left(-bd_{jk}\right)}{\sum_r N_r \sum_k \exp\left(-bd_{rk}\right)}.$$

Then the balance condition in (4.3) is satisfied. This can be verified by noticing that:

$$a_i N_i c_{ij} = \sum\nolimits_k \exp\left(-bd_{ik}\right) N_i \frac{N_j \sum_k \exp\left(-bd_{jk}\right)}{\sum_r N_r \sum_k \exp\left(-bd_{rk}\right)},$$

$$a_j N_j c_{ji} = \sum\nolimits_k \exp\left(-bd_{jk}\right) N_j \frac{N_i \sum_k \exp\left(-bd_{ik}\right)}{\sum_r N_r \sum_k \exp\left(-bd_{rk}\right)}.$$

Clearly, $a_i N_i c_{ij} = a_j N_j c_{ji}$.

Another modification is to keep c_{ij} the same as in (4.6) but re-define the activity (4.5) as $a_i = \xi \sum_k N_k \exp\left(-bd_{ik}\right)$ where ξ is a scaling constant. Then,

$$a_i N_i c_{ij} = \xi \sum\nolimits_k N_k \exp\left(-bd_{ik}\right) \frac{N_i N_j \exp\left(-bd_{ij}\right)}{\sum_k N_k \exp\left(-bd_{ik}\right)} = \xi N_i N_j \exp\left(-bd_{ij}\right),$$

$$a_j N_j c_{ji} = \xi \sum\nolimits_k N_k \exp\left(-bd_{jk}\right) \frac{N_j N_i \exp\left(-bd_{ji}\right)}{\sum_k N_k \exp\left(-bd_{jk}\right)} = \xi N_i N_j \exp\left(-bd_{ji}\right).$$

Again, we have $a_i N_i c_{ij} = a_j N_j c_{ji}$.

4.8 Summary

In this chapter, we have endeavored to explain why mixing is important in meta-population models and to describe several formulations that meet specified conditions and questions for further research. Meta-population models are used in public health to identify interventions that will reduce the effective reproduction number (number of secondary infections per infectious person) the most.

Appendix 1

Using the approach of van den Driessche and Watmough [10], here we find the next-generation matrix K. Given that $N = S + I + R$, first we eliminate one equation. Letting $x_i = \frac{S_i}{N_i}$, $y_i = \frac{I_i}{N_i}$, $i = 1, 2$, we have the equations for fractions:

$$\begin{aligned} x_1' &= \mu\left(1-p_1\right) - \left(\lambda_1+\mu\right)x_1 \\ x_2' &= \mu\left(1-p_2\right) - \left(\lambda_2+\mu\right)x_2 \\ y_1' &= \lambda_1 x_1 - \left(\gamma+\mu\right)y_1 \\ y_1' &= \lambda_2 x_2 - \left(\gamma+\mu\right)y_2 \\ \lambda_i &= a_i\beta\sum\nolimits_j c_{ij}y_j. \end{aligned}$$

At the disease-free equilibrium, $x_i = 1 - p_i$, $i = 1, 2$. Substituting $1 - p_i$ for x_i in the y_i equation,

$$\begin{aligned} y_1' &= \left(a_1c_{11}\beta y_1 + a_1c_{12}\beta y_2\right)\left(1-p_1\right) - \left(\gamma+\mu\right)y_1 \\ y_2' &= \left(a_2c_{21}\beta y_1 + a_2c_{22}\beta y_2\right)\left(1-p_2\right) - \left(\gamma+\mu\right)y_2. \end{aligned}$$

Denote the functions on the right-hand side of the y_1 and y_2 equations by $f_1(y_1,y_2)$ and $f_1(y_1,y_2)$, respectively. Then the Jacobian matrix at the disease-free equilibrium is:

$$\begin{aligned} J &= \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} \end{pmatrix}_{(y_1=0,y_2=0)} \\ &= \begin{pmatrix} a_1c_{11}\beta\left(1-p_1\right) - \left(\gamma+\mu\right) & a_1c_{12}\beta\left(1-p_1\right) \\ a_2c_{21}\beta\left(1-p_2\right) & a_2c_{22}\beta\left(1-p_2\right) - \left(\gamma+\mu\right) \end{pmatrix}. \end{aligned}$$

We can rewrite J as $F - V$, where F includes infection terms and V other terms:

$$J = \begin{pmatrix} a_1c_{11}\beta\left(1-p_1\right) & a_1c_{12}\beta\left(1-p_1\right) \\ a_2c_{21}\beta\left(1-p_2\right) & a_2c_{22}\beta\left(1-p_2\right) \end{pmatrix} - \begin{pmatrix} \left(\gamma+\mu\right) & 0 \\ 0 & \left(\gamma+\mu\right) \end{pmatrix}.$$

The next-generation matrix is $K{=}FV^{-1}$, i.e.,

$$K = \begin{pmatrix} a_1c_{11}\beta\left(1-p_1\right) & a_1c_{12}\beta\left(1-p_1\right) \\ a_2c_{21}\beta\left(1-p_2\right) & a_2c_{22}\beta\left(1-p_2\right) \end{pmatrix} \begin{pmatrix} \left(\gamma+\mu\right)^{-1} & 0 \\ 0 & \left(\gamma+\mu\right)^{-1} \end{pmatrix} =$$

$$\begin{pmatrix} \frac{a_1c_{11}\beta(1-p_1)}{(\gamma+\mu)} & \frac{a_1c_{12}\beta(1-p_1)}{(\gamma+\mu)} \\ \frac{a_2c_{21}\beta(1-p_2)}{(\gamma+\mu)} & \frac{a_2c_{22}\beta(1-p_2)}{(\gamma+\mu)} \end{pmatrix} = \begin{pmatrix} \Re_{01}\left(1-p_1\right)c_{11} & \Re_{01}\left(1-p_1\right)c_{12} \\ \Re_{02}\left(1-p_2\right)c_{21} & \Re_{02}\left(1-p_2\right)c_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \Re_{v1}c_{11} & \Re_{v1}c_{12} \\ \Re_{v2}c_{21} & \Re_{v2}c_{22} \end{pmatrix}.$$

The reproduction number is the dominant eigenvalue of K.

Appendix 2

Values of parameters of the mixing function proposed by Glasser et al. [21] and illustrated in Fig. 4.5 fitted to observations from the PolyMod study are provided in Table 4.2 (physical contacts) and Table 4.3 (casual contacts).

Table 4.2 Values of parameters and vectors estimated from physical contacts

Fractions reserved for contemporaries, children, parents and co-workers	
$\vec{\varepsilon}_1$	(0.21, 0.4, 0.4, 0.32, 0.2, 0.07, 0.04, 0.05, 0.03, 0.03, 0.04, 0.08, 0.08, 0.11, 0.14)
$\vec{\varepsilon}_2$	(0, 0, 0, 0, 0, 0.129, 0.198, 0.372, 0.175, 0.06, 0.014, 0, 0, 0.063, 0)
$\vec{\varepsilon}_3$	(0.097, 0.093, 0.175, 0.087, 0.056, 0.012, 0, 0, 0.015, 0, 0, 0, 0, 0, 0)
$\vec{\varepsilon}_4$	(0, 0. 0, 0.3, 0.073, 0, 0, 0, 0.003, 0.09, 0.06, 0.015, 0.012, 0, 0)
Variances of age-distributions of contemporaries, parents and children	
$\vec{\sigma}_1$	(0.38, 0.42, 0.33, 0.27, 0.31, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.25, 0.24, 0.27, 0.41)
$\vec{\sigma}_2$	(1, 1, 1, 1, 1, 2.5, 1.54, 3.97, 3.06, 1, 1, 2.46, 2.49, 1.23, 2.68)
$\vec{\sigma}_3$	(1, 1, 3.04, 1, 3.92, 1, 2.44, 2.49, 1, 2.5, 1, 1, 1, 1, 1)
Ages at entry and exit from workforce, generation time and longevity	
$W_{\min} = 25$, $W_{\max} = 55$, $G = 30$, $L = 75$	

Table 4.3 Values of parameters and vectors estimated from casual contacts

Fractions reserved for contemporaries, children, parents and co-workers	
$\vec{\varepsilon}_1$	(0.18, 0.4, 0.4, 0.31, 0.31, 0.15, 0.02, 0.02, 0.03, 0.02, 0.002, 0.03, 0.03, 0.15, 0.2)
$\vec{\varepsilon}_2$	(0, 0, 0, 0, 0, 0.09, 0.14, 0.28, 0.12, 0.06, 0, 0, 0, 0.026, 0)
$\vec{\varepsilon}_3$	(0.3, 0.17, 0.21, 0.08, 0.07, 0, 0, 0, 0.004, 0, 0, 0, 0, 0, 0)
$\vec{\varepsilon}_4$	(0, 0. 0, 0.3, 0.06, 0.04, 0, 0, 0.1, 0.16, 0.0619, 0, 0, 0, 0)
Variances of age-distributions of contemporaries, parents and children	
$\vec{\sigma}_1$	(0.2, 0.34, 0.31, 0.29, 0.35, 0.2, 0.2, 0.2, 0.2, 0.2, 0.24, 0.2, 0.2, 0.39, 0.53)
$\vec{\sigma}_2$	(1, 1, 1, 1, 1, 2.5, 3, 4, 4, 1, 2.6, 2.5, 2.51, 1.56, 2.53)
$\vec{\sigma}_3$	(3.62, 4, 4, 1, 3.95, 1, 2.49, 2.42, 1, 2.5, 1, 1, 1, 1, 1)
Ages at entry and exit from workforce, generation time and longevity	
$W_{\min} = 25$, $W_{\max} = 55$, $G = 30$, $L = 75$	

Appendix 3

The model of Feng et al. [2] is not suitable for age groups, but several of those in Feng et al. [5] are. Suppose, for example, that all newborn individuals are susceptible and enter the first age group, and that people exit age groups at specific rate θ (i.e. age from group $i-1$ to i and die while in group n) and that susceptible ones get vaccinated at rate χ. Then,

$$\begin{aligned}
\tfrac{dS_1}{dt} &= \theta N_n - (\lambda_1(t) + \chi + \theta)\, S_1,\\
\tfrac{dS_i}{dt} &= \theta S_{i-1} - (\lambda_i(t) + \chi + \theta)\, S_i, \quad 1 < i \le n,\\
\tfrac{dI_1}{dt} &= \lambda_1(t) S_1 - (\gamma + \theta)\, I_1,\\
\tfrac{dI_i}{dt} &= \theta I_{i-1} + \lambda_i(t) S_i - (\gamma + \theta)\, I_i, \quad 1 < i \le n,\\
\tfrac{dR_1}{dt} &= \chi S_1 + \gamma I_1 - \theta R_1,\\
\tfrac{dR_i}{dt} &= \chi S_i + \theta R_{i-1} + \gamma I_i - \theta R_i, \quad 1 < i \le n,
\end{aligned}$$

where

$$N_i = S_i + I_i + R_i, \;\; N = \sum\nolimits_i N_i,$$

and

$$\lambda_i(t) = a_i \beta_i \sum\nolimits_j c_{ij} \left(\frac{I_j}{N_j}\right).$$

The θN_n births enter the S_1 class, the $\theta\ (S_n + I_n + R_n) = \theta N_n$ deaths exit the nth age class, and N is constant. At the stable age-distribution, $N_n = N/n$, $1 \le i \le n$. Consider the fractions

$$x_i = \frac{S_i}{N_i}, \quad y_i = \frac{I_i}{N_i}, \quad z_i = \frac{R_i}{N_i}, \quad 1 \le i \le n.$$

The equations for y_i have the same form as those for I_i except that the forces of infection are $\lambda_i(t) = a_i \beta_i \sum_j c_{ij} y_j$. The disease-free equilibrium is $x_i = 1$, $y_i = z_i = 0$, $1 \le i \le n$. Note that $\tau = 1/(\gamma + \theta)$ is the infectious period for all age groups i $(1 \le i \le n)$, and that $p = \theta/(\gamma + \theta)$ denotes the probability that an infectious person in age group i $(1 \le i < n)$ enters infectious age group $i + 1$. Let

$$
\begin{aligned}
&A_{i1} = c_{i1}\tau + c_{i2}p\tau + \cdots + c_{i(n-1)}p^{n-2}\tau + c_{in}p^{n-1}\tau = \sum_{k=1}^{n} c_{ik}p^{k-1}\tau \\
&A_{i2} = c_{i2}\tau + c_{i3}p\tau + \cdots + c_{i(n-1)}p^{n-3}\tau + c_{in}p^{n-2}\tau = \sum_{k=2}^{n} c_{ik}p^{k-2}\tau \\
&\cdots\cdots \\
&A_{i(n-1)} = c_{i(n-1)}\tau + c_{in}p\tau, \quad A_{in} = c_{in}\tau, \quad i = 1, 2, \cdots n.
\end{aligned}
$$

The next-generation matrix is

$$
K = \begin{pmatrix} a_1\beta_1 A_{11} & a_1\beta_1 A_{12} & \cdots & a_1\beta_1 A_{1n} \\ a_2\beta_2 A_{21} & a_2\beta_2 A_{22} & \cdots & a_2\beta_2 A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_n\beta_n A_{n1} & a_n\beta_n A_{n2} & \cdots & a_n\beta_n A_{nn} \end{pmatrix},
$$

and its dominant eigenvalue is $\Re_0$. Define $\Re_{0i} = a_i\beta_i\tau$ for $1 \le i \le n$. When mixing is proportional, $c_{1j} = c_{2j} = c_{nj}$ for all j. Thus, $A_{1j} = A_{2j} = \cdots A_{nj} \doteq A_j$ for all j. Because K has rank 1, the dominant eigenvalue of K is given by its trace:

$$
\Re_0 = \sum_{i=1}^{n} a_i\beta_i A_i = \sum_{i=1}^{n}\sum_{k=i}^{n} c_{ik}p^{k-i}\Re_{0k}.
$$

Acknowledgements We thank the anonymous reviewer for his/her very careful review of this chapter and for all the helpful comments that greatly improved the presentation. We acknowledge helpful discussions with Carlos Castillo-Chavez, Sara Del Valle, Herb Hethcote, Andrew Hill, Mac Hyman, and Andrew Moylan. We are grateful to Elsevier for permission to reuse Figs. 4.1, 4.2, and 4.3 from the Journal of Theoretical Biology, Figs. 4.7, 4.8, and 4.9 from the Lancet Infectious Diseases, and Figs. 4.5 and 4.6 from Mathematical Biosciences, and to the American Institute of Biological Sciences for permission to reuse Fig. 4.4 from Discrete and Continuous Dynamical Systems Series B.

References

1. R. Levins, Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. **15**, 237–240 (1969)
2. Z. Feng, A.N. Hill, P.J. Smith, J.W. Glasser, An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or non-random mixing. J. Theor. Biol. **386**, 177–187 (2015)
3. R.M. Anderson, B.T. Grenfell, Quantitative investigations of different vaccination policies for the control of congenital rubella syndrome (CRS) in the United Kingdom. J. Hyg. **96**, 305–333 (1986)
4. N. Hens, Z. Shkedy, M. Aerts, C. Faes, P. Van Damme, P. Beutels, *Modeling Infectious Disease Parameters Based on Serological and Social Contact Data: A Modern Statistical Perspective* (Springer, New York, 2012)

5. Z. Feng, J. Glasser, M. Andersson, R.-M. Carlsson, P. Tüll, H. Hallander, P. Olin, Modeling risks of infection when immunity wanes: application to Bordetella pertussis in Sweden. J. Theor. Biol. **356**, 123–132 (2014)
6. S. Busenberg, C. Castillo-Chavez, A general solution of the problem of mixing of subpopulations and its application to risk- and age-structured epidemic models for the spread of AIDS. IMA J. Math. Appl. Med. Biol. **8**, 1–29 (1991)
7. J.A. Jacquez, C.P. Simon, J. Koopman, L. Sattenspiel, T. Perry, Modeling and analyzing HIV transmission: the effect of contact patterns. Math. Biosci. **92**, 119–199 (1988)
8. A. Nold, Heterogeneity in disease transmission modeling. Math. Biosci. **52**, 227–240 (1980)
9. K.T.D. Eames, N.L. Tilston, E. Brooks-Pollock, W.J. Edmunds, Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS Comput. Biol. **8**(3), e1002425 (2012)
10. P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
11. K. Dietz, Models for vector-borne parasitic diseases, in *Vito Volterra Symposium on Mathematical Models in Biology*, Lecture Notes in Biomathematics, ed. by C. Barigozzi, vol. 39, (1980), pp. 264–277
12. A.D. Barbour, Macdonald's model and the transmission of bilharzias. Trans. R. Soc. Trop. Med. Hyg. **72**, 6–15 (1978)
13. V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. J. Theor. Biol. **251**, 450–467 (2008)
14. R.M. May, R.M. Anderson, Spatial heterogeneity and the design of immunization programs. Math. Biosci. **72**, 83–111 (1984)
15. J.W. Glasser, Z. Feng, S.B. Omer, P.J. Smith, L.E. Rodewald, The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study. Lancet Infect. Dis. **16**, 599–605 (2016)
16. L. Chow, M. Fan, Z. Feng, Dynamics of a multigroup epidemiological model with group-targeted vaccination strategies. J. Theor. Biol. **291**, 56–64 (2011)
17. S.Y. Del Valle, J.M. Hyman, H.W. Hethcote, S.G. Eubank, Mixing patterns between age groups in social networks. Soc. Networks **29**, 539–554 (2007)
18. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, S.T. Gianpaolo, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W.J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. **5**, 381–391 (2008)
19. J. Wallinga, P. Teunis, M. Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am. J. Epidemiol. **164**, 936–944 (2006)
20. E. Zagheni, F.C. Billari, P. Manfredi, A. Melegaro, J. Mossong, W.J. Edmunds, Using time-use data to parameterize models for the spread of close-contact infectious diseases. Am. J. Epidemiol. **168**, 1082–1090 (2008)
21. J.W. Glasser, Z. Feng, A. Moylan, S. Del Valle, C. Castillo-Chavez, Mixing in age-structured population models of infectious diseases. Math. Biosci. **235**, 1–7 (2012)
22. H.W. Hethcote, Modeling heterogeneous mixing in infectious disease dynamics, in *Models for Infectious Human Diseases: Their Structure and Relation to Data*, ed. by V. Isham, G. Medley, (Cambridge Univ Press, Cambridge, 1996), pp. 215–238
23. Z. Feng, A.N. Hill, A.T. Curns, J.W. Glasser, Evaluating targeted interventions via meta-population models with multi-level mixing. Math. Biosci. **287**, 93–104 (2017)
24. H. Caswell, *Matrix Population Models: Construction, Analysis and Interpretation*, 2nd edn. (Sinauer Associates, Sunderland, 2001)

25. L. Hao, J.W. Glasser, Q. Su, C. Ma, Z.-L. Feng, Z. Yin, J.L. Goodson, N. Wen, C. Fan, H. Yang, L.E. Rodewald, Z.-J. Feng, H. Wang, Evaluating vaccination policies to accelerate measles elimination in China: a meta-population modelling study. Int J Epidemiol (in press) https://doi.org/10.1093/ije/dyz058 (2019)
26. Z. Feng, Q. Han, Z. Qiu, A.N. Hill, J.W. Glasser, Computation of $\mathfrak{R}$ in age-structured epidemiological models with maternal and temporary immunity. Discrete Continuous Dyn. Syst. Ser. B **21**, 399–415 (2016)

Chapter 5
Structured Population Models for Vector-Borne Infection Dynamics

Jianhong Wu

Abstract Dynamical systems provide an appropriate framework to examine whether, where and when a vector species and/or a vector-borne pathogen can establish and spread. Such systems often contain time lags to reflect the transition times from one physiological stage to the next, or from one geographic location to others. We present a brief introduction to dynamical systems generated by delay differential equations with varying delay. We focus on those delay differential equations which are reduced from structured population partial differential equation models, and we discuss the implicit assumption that needs to be made to permit this reduction process. We demonstrate the model formulation from tick population and tick-borne disease infection dynamics, and from bird migration and avian influenza spread dynamics. We show how model parameters, especially time-varying development delays, can be informed from laboratory experiments, field studies and surveillance data, and how these parameters are integrated to a single threshold parameter, the basic reproduction number, to quantify when population establishment and disease persistence are likely.

5.1 Delay Differential Equations for Disease Infection Dynamics

An introduction of mathematical modelling for vector-borne disease infection dynamics often starts with a simplified assumption about the homogeneity in the population in terms of reproduction, transmission contacts and environmental conditions. This assumption yields compartmental systems of ordinary differential equations.

Applications of dynamical systems-based modelling and analysis to informing ecosystem management and disease intervention require however details about

J. Wu (✉)
Laboratory for Industrial and Applied Mathematics, York University, Toronto, ON, Canada
e-mail: wujh@yorku.ca

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_5

the heterogeneities in the physiological status of the vector species (such as ticks in the context of tick-borne diseases such as Lyme disease) and/or geographical location of the vector species (such as migratory birds in the context of avian influenza spread). Associated with this requirement from applications is the gradually improved surveillance and field observation about these physiological status and/or geographical location. Incorporating these heterogeneities into an infection dynamics model gives rise to structured population and epidemic models which, under the assumption of homogeneity within a particular stage or a spatial segment, can be reduced to a system of delay differential equations (DDEs).

Here we start with a short introduction to the basic model framework and some fundamental results about systems of DDEs. We will then focus on the case of delays which are variable due to climate change and environmental condition variations, and focus on reduction from structured to staged models. We will discuss the definition and calculation of the vector population establishment threshold, the basic reproduction number in vector ecology, and show how this combined with environmental and vector behavior data can be used to produce the vector population establishment risk maps. We will then introduce the concept of monotone maps and threshold dynamics and present two illustrative examples: Lyme tick population dynamics with structured life cycles, and bird migration dynamics and spatially structured models. We will finally touch on the persistence theory and illustrate the theory with two examples: avian influenza spread through bird migration, and Lyme disease dynamics through multi-stage systemic transmission.

5.2 Delay Differential Equations: Setting Up the Model

We start with the logistic equation

$$x'(t) = rx(t)[1 - x(t)/K], \quad r, K > 0;$$

or generally,

$$x'(t) = -d(x) + b(x)$$

with $d(x)$ as the death rate and $b(x)$ as the birth rate.

Examples are when $d(x) = -rx^2/K$ and $b(x) = rx$ (logistic equation, when $K > 0$ is the carrying capacity constant and r is the intrinsic growth rate); when $d(x) = dx$ with a constant $d > 0$ and $b(x) = pxe^{-qx}$ (the so-called *Ricker function*, leading to a monostable system); and when $b(x) = px^2e^{-qx}$ (modelling the Allee effect and leading to a bistable system).

In this model formulation, homogeneity is implicitly assumed: every individual can reproduce, and birth into the population is instantaneous. In most biological populations, however, individuals can reproduce only after maturation. A more realistic formulation posits two classes within the population: immature and mature (reproducing) individuals. If there is a uniform maturation time (τ), then the equation becomes

$$x'(t) = -dx(t) + \alpha b(x(t-\tau)) \tag{5.1}$$

with the second term being the maturation rate (birth rate at $t - \tau$ times the survival probability α during the maturation). This gives a delay differential equation.

In what follows, we suggest the readers to keep the following reproduction function in mind:

$$b(x) = pxe^{-qx}.$$

To uniquely define a solution for all future time $t \geq 0$, we need to specify the initial condition $x(s) = \phi(s)$ for $s \in [-\tau, 0]$ with the initial function ϕ given from the phase space $C := C([-\tau, 0])$. The initial value problem of (5.1) subject to the initial condition can then be solved using the method of steps that solves the initial value problem consequently on the intervals $[0, \tau]$, $[\tau, 2\tau]$, $\cdots$, $[n\tau, (n + 1)\tau]$ for any integer $n > 1$.

There are some important properties of DDEs, including the non-existence and possible non-uniqueness of backward extensions from an initial condition, the eventual compactness, non-negativeness and boundedness of the (forward) solutions when the feedback function b is appropriately given. The solutions then give a semiflow in C which has a global attractor. The standard notation x_t is used to denote the segment x on the interval $[t - \tau, t]$ translated into the initial interval $[-\tau, 0]$, i.e.,

$$x_t(s) = x(t + s), s \in [-\tau, 0].$$

Fundamental results can be found in [15]; see also [6, 8, 9, 13, 14, 19, 21, 30, 37] for a collection of textbooks and references.

The local stability of the model system at a given equilibrium x^* is determined by the stability of the zero solution of the linear system describing the perturbation $x(t)$ around x^*:

$$x'(t) = -dx(t) + \alpha b'(x^*)x(t-\tau).$$

The linearization at the zero equilibrium $x^* = 0$ generates a positive semigroup (since $b'(0) > 0$) and hence the stability of this equilibrium is determined by the real eigenvalue of the characteristic equation (Corollary 3.2 of [30])

$$\lambda = -d + \alpha b'(x^*)e^{-\lambda\tau}.$$

So if $\alpha b'(0) < d$ then $x^* = 0$ is locally asymptotically stable. We can also easily check that when $\alpha b'(0) < d$ the system has no positive equilibrium. On the other hand, if $\alpha b'(0) > d$ then there is a positive equilibrium x^* which maybe locally asymptotically stable or unstable, but the zero equilibrium becomes unstable. In some cases, we can also make conclusions regarding the global attractivity of the positive equilibrium using the monotone dynamical systems theory [30]. This is particularly true when the positive equilibrium is within the interval where the function b remains monotonically increasing. In the case where $b'(x^*) < 0$ at the positive equilibrium x^*, we have the situation of a negative feedback around this equilibrium and a Hopf bifurcation of periodic solutions may take place. This is a typical example of delay-induced nonlinear oscillations. Figure 5.1 gives an illustration of possible scenarios of model dynamics, depending on the location of the intersections between the death rate dx and the maturation rate function $\alpha b(x)$.

Note also that the survival probability during the maturation period may depend on the maturation delay τ, the stability analysis of the characteristic equation involving delay-dependent coefficients is very complicated, and the global Hopf bifurcation (the birth, death and global continuation of local Hopf bifurcation) has

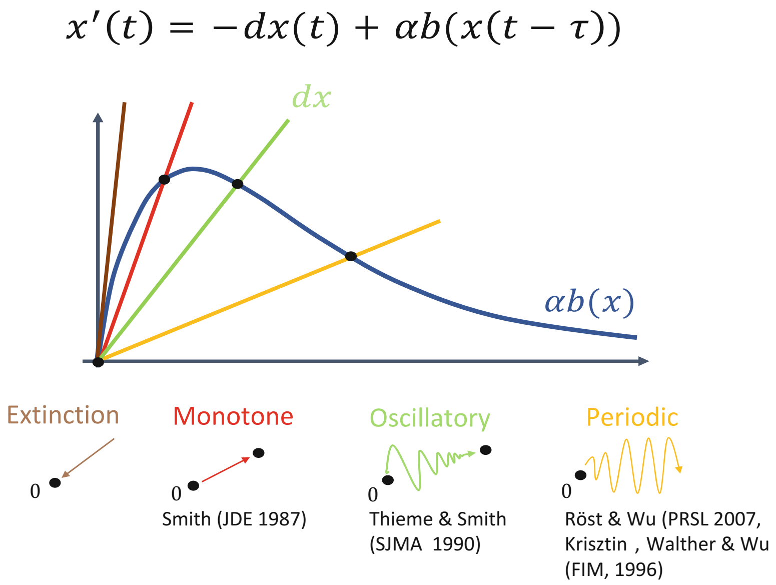


Fig. 5.1 The global dynamics of the delay differential equation $x'(t) = -dx(t) + \alpha b(x(t - \tau))$ with a constant delay $\tau > 0$. Depending on the relative value of $\alpha b'(0)$ and d, the model may have one or two non-negative equilibria. The positive equilibrium, if it exists and is within the interval where the function b is increasing, is the global attractor for all solutions of the equation with non-trivial non-negative initial value (using the monotone dynamical system theory [29]). When this positive equilibrium value is in the interval when the function b is decreasing and when the delay is small, this equilibrium remains stable (using the dynamical system theory in [31] for semiflows which are order-preserving with respect to the so-called exponential ordering). This equilibrium however can lose its stability through the mechanism of Hopf bifurcations. Under certain technical conditions, one can show that the bifurcated periodic solutions are stable [26], and the structure of the global attractor can be described as in [18]

been recently studied in [27, 28]. See also [40] for some further extensions when the scalar equation is replaced by a system of delay differential equations.

A question arises: *What happens if the delay is not a constant, but for example, temporally periodic (maturation regulated by the seasonal variation of the environment)? If this delay is given by* $\tau(t)$, *would the model become*

$$x'(t) = -dx(t) + \alpha b(x(t - \tau(t)))?$$

A positive answer was suggested and used in a number of studies including [11]. However, this answer ignored a key factor as explained below.

To understand why the general answer to the above question should be negative, we call attention to the warning statement in the textbook [8] that one should appropriately start with the description of population dynamics at the individual level or to derive from a probabilistic formulation the system for the matured population dynamics with variable delay.

Let us take the approach using reduction by integration along characteristics of individual-based models. We define $u(t, a)$ as the population density at time t and age a. The dynamics is described by the basic evolutional operator

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) u(t, a) = -\mu(a)u(t, a)$$

subject to boundary condition

$$u(t, 0) = b(M(t)),$$

where the reproductive population is given by

$$M(t) = \int_{\tau(t)}^{\infty} u(t, a)da.$$

Under the assumption that $\mu(a)$ is stage-dependent (not age-dependent) (that is, $\mu(a) = \mu_m$ for a constant independent of the age variable $a \geq \tau(t)$, and $\mu(a) = \mu_i$ for $a < \tau(t)$), we can use integration along characteristics to obtain

$$M'(t) = -\mu_m M(t) + (1 - \tau'(t))e^{-\mu_i \tau(t)} b(M(t - \tau(t))).$$

Note that a factor $(1 - \tau'(t))$ appears, which is one only when the delay is a constant.

An interesting problem for future studies is whether we can transfer the above model with variable delay to a periodic DDE model with a constant delay, with a transformation that is guided by, and can provide with, biological insights into the maturation process with variable maturation time.

5.3 Lyme Tick Population Dynamics

An example to illustrate the importance of considering structured population modelling and variable developmental delays is Lyme disease transmission dynamics.

Lyme disease spread involves complex interaction of a spirochete, multiple vertebrate hosts, and a vector with a two (or three)-year life cycle strongly influenced by the season rhythm. The black-legged tick, *Ixodes scapularis Say*, is the primary vector of *Borrelia burgdorferi*, the bacterial agent of Lyme disease, in eastern and mid-western United States. Northward invasive spread of the tick vectors from United States endemic foci to non-endemic Canadian habitats has been a public health concern. A mathematical model to faithfully describe the development of tick populations and the pathogen spread dynamics is needed to understand the invasion pattern and predict Lyme infection risk under projected environmental condition variations.

In [38], a system of ordinary differential equations with periodic coefficients was proposed for the tick population dynamics. Such a model implicitly makes an assumption of exponentially distributed development delays. However, tick development delay is normally concentrated around a particular value though this value depends on the historical environment conditions up to the time of the completion of the development. A more appropriate model would require the use of time-varying development delay.

An attempt was made in [39] which carefully follows the development of tick populations from one stage to another. In the formulated model, the development delay is not a constant but rather a periodic function of the time due to seasonality in the environmental conditions. The model parameters were estimated from many years of surveillance, lab test and field data, and the theory of Floquet multipliers of periodic systems was used to calculate the threshold condition for the tick population dynamics. In the next subsections, we will introduce the model and some relevant analyses.

5.3.1 Model Formulation and Objective

We now describe key ingredients in the aforementioned model study.

- Model formulation: a general dynamic population model where the development time from one life stage to the next has considerable variation due to temperature change.
- Key assumptions: the transition time between two consecutive stages is constant when the temperature is fixed; the correlation between the fixed temperature and the transition time can be determined from lab data; the temperature in a considered region varies periodically (annually); and therefore the transition time between two consecutive stages (in the considered region) is a temporally varying periodic function (of the time).

- Model variables: The life cycle of a population is divided into n stages, with each stage embodying a specific point of the life of the individual. Let x_j $(1 \leq j \leq n)$ be the size of subpopulations at the jth stage, with stages in order of increasing maturity (e.g., egg, larvae, nymphs, adult…), except x_1 which is the size of the mature subpopulation who are able to produce offspring (egg-laying females).
- Objective: To formulate a closed system for the dynamics of $(x_1(t), \cdots, x_n(t))$ in order to predict the tick establishment risk.

Age-structured model, the starting point: We start with the population's chronological age variable a (time since being produced as an egg), and describe the evolution of $\rho(t, a)$, the density of the female population, by

$$\begin{cases} (\frac{\partial}{\partial t} + \frac{\partial}{\partial a})\rho(t, a) = -\mu(t, a)\rho(t, a), \\ \rho(0, a) = \phi(a), \;\; a \geq 0 \quad \text{(Initial Condition)}, \\ \rho(t, 0) = b(x_1(t)), \;\; t \geq 0 \quad \text{(Boundary Condition)}. \end{cases}$$

Here μ is the death rate. Integrating along characteristics yields

$$\rho(t, a) = \begin{cases} \rho(0, a - t)e^{-\int_0^t \mu(r, a-t+r)\, dr}, & 0 \leq t \leq a, \\ \rho(t - a, 0)e^{-\int_0^a \mu(t-a+r, r)\, dr}, & a < t. \end{cases}$$

A natural question then arises: *What kind of homogeneity needs to be assumed to permit the reduction from a structured population PDE model to a stage-structured DDE model?*

It turns out that the stage-homogeneity assumption about the mortality rate $\mu(t, a)$ given below is (mathematically) sufficient and (practically) justified by how the lab and field observation data is collected. This stage-homogeneity assumption states that each mortality rate in a given stage is a constant independent of the ages within the given stage, but the mortality rates can vary from one stage to another. This is described by

$$\mu(t, a) = \begin{cases} \mu_1(x_1(t)), & a \in [A_n(t), \infty), \\ \mu_i(x_i(t)), & a \in [A_{i-1}(t), A_i(t)], \;\; i = 2, \cdots, n, \end{cases}$$

where $A_{i-1}(t)$ and $A_i(t)$ are the time-dependent minimum and maximum ages of those individuals who are developing within the specific ith stage, and

$$\begin{cases} x_1(t) = \int_{A_n(t)}^{\infty} \rho(t, a)\, da \\ x_i(t) = \int_{A_{i-1}(t)}^{A_i(t)} \rho(t, a)\, da, \;\; i = 2, \cdots, n. \end{cases}$$

Under this stage-homogeneity assumption, we have from the evolution equation the following

$$\begin{aligned} x_1'(t) &= \int_{A_n(t)}^{\infty} \{(\partial_t + \partial_a)\rho(t,a) - \partial_a \rho(t,a)\}\, da - \rho(t, A_n(t))A_n'(t) \\ &= \rho(t,\infty) + \rho(t, A_n(t)) - \int_{A_n(t)}^{\infty} \mu(t,a)\rho(t,a)\, da - \rho(t, A_n(t))A_n'(t) \\ &= \rho(t, A_n(t))(1 - A_n'(t)) - \int_{A_n(t)}^{\infty} \mu(t,a)\rho(t,a)\, da \\ &= \rho(t, A_n(t))(1 - A_n'(t)) - \mu_1(x_1(t))x_1(t). \end{aligned}$$

Similarly, for $i = 2, \cdots, n$, we have

$$x_i'(t) = \rho(t, A_{i-1}(t))(1 - A_{i-1}'(t)) - \rho(t, A_i(t))(1 - A_i'(t)) - \mu_i(x_i(t))x_i(t).$$

Therefore, we obtain the closed system:

$$\begin{cases} x_1'(t) = \rho(t, A_n(t))(1 - A_n'(t)) - \mu_1(x_1(t))x_1(t), \\ x_i'(t) = \rho(t, A_{i-1}(t))(1 - A_{i-1}'(t)) - \rho(t, A_i(t))(1 - A_i'(t)) - \mu_i(x_i(t))x_i(t). \end{cases} \tag{5.2}$$

Note also that x_1 is decoupled from other equations in system (2).

With appropriate assumptions on b and μ_i, we can obtain the non-negativeness, boundedness and the existence of the global compact attractor.

We now address the *practical problem: How to calculate $A_i(t)$ and $\rho(t, A_i(t))$ from the available data?* To answer this question, we let $\tau_i(t)$ represent the length of time that a tick is developed at time t into the $(i+1)$-stage from a tick at the previous i-stage at time $t - \tau_i(t)$. Much of the qualitative analysis requires the condition

$$1 - \tau_i'(t) \geq 0. \tag{5.3}$$

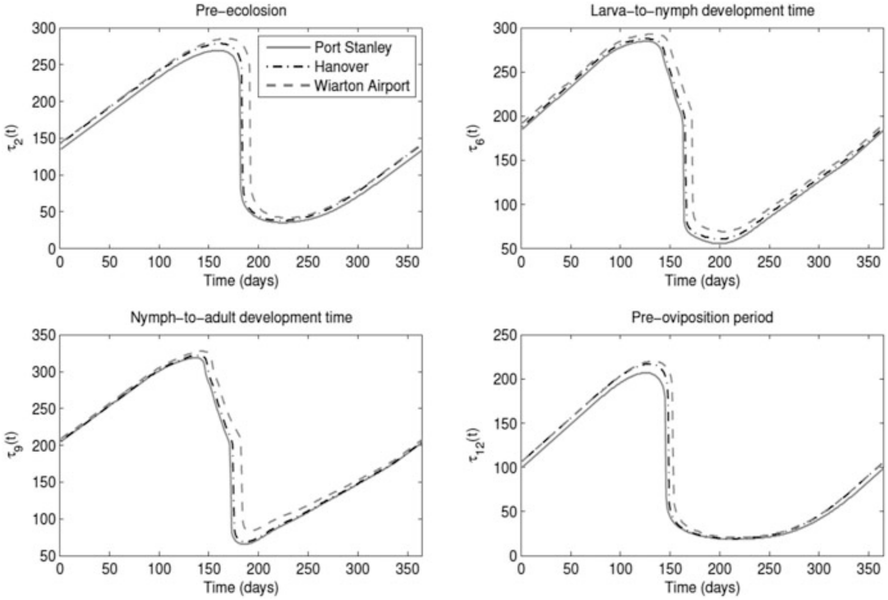
It is important, for resolving the above practical problem, that we note $\tau_i(t)$ can be approximated from lab data. An illustration is given in Fig. 5.2, see also [22–24] for some of the lab data discussions.

Equally importantly, from the biological interpretations between maturation age and chronological ages, we can calculate $A_i(t)$ iteratively from $\tau_i(t)$ using the following formula (formula (9) in [39]):

$$A_i(t) = \sum_{j=2}^{i} \tau_j \left(t - \sum_{k=j+1}^{i} \tau_k \left(t - \sum_{l=k+1}^{i} \tau_l \left(t - \cdots \tau_{i-1}(t - \tau_i(t))\right)\right)\right).$$

With the above discussions, we can then substitute

$$\rho(t, A_i(t)) = \rho(t - A_i(t), 0)\alpha_i(t, t - A_i(t))$$


Pre–ecolosion

Larva–to–nymph development time

Nymph–to–adult development time

Pre–oviposition period

Fig. 5.2 Samples of time-varying development delays, using temperature data during 1971–2000, in Port Stanley, Hanover and Wiarton Airport weather stations (Figure is taken from [39])

to the equation (5.2) coupled with the reproduction condition

$$\rho(t, 0) = b(x_n(t))$$

to get a closed system. Here $\alpha_i(t, t - A_i(t))$ $(i = 2, \cdots, n)$ can now be calculated and represents the density-dependent survival probability of an egg who was born at time $t - A_i(t)$ and is able to live until time t when the egg matures (fully) to the ith stage.

5.3.2 The Ecological Threshold: Calculating Future Generation of Egg-Laying Females

To answer the question whether the population can grow and establish in the environment, we linearize the system at the zero solution to check if the population will undergo exponential growth from a small population. This leads to the introduction of the basic reproduction number R_0.

In particular, the linearized system at the zero solution has a one-dimensional decoupled subsystem

$$x_1'(t) = a(t)x_1(t - A_n(t)) - \mu_1(0)x_1(t),$$

with $a(t) = b'(0)\alpha_n(t, t - A_n(t))(1 - A_n'(t))$ being the change rate of egg-laying females at time t that depends on the number of egg-laying females at time $t - A_n(t)$.

To define the basic reproduction number, we examine the future generation of egg-laying female ticks. We assume that

$$h(t) := t - A_n(t) \text{ is a strictly increasing function of } t.$$

Integration yields

$$x_1(t) = \int_{-\infty}^{t} e^{-\mu_1(0)(t-s)} a(s) x_1(s - A_n(s)) ds.$$

This allows us to look at the number of newly generated egg-laying females per unit time at time t, from an initial introduction of the egg-laying females with an initial distribution of $x(s)$, $s \in R$ (Fig. 5.3).

More specifically, for a fixed time t, the cohort of egg-laying females will produce some newborns who will eventually become egg-laying females at the future time

$$h^{-1}(t) := \tilde{t}, \quad \text{where} \quad h(\tilde{t}) = \tilde{t} - A_n(\tilde{t}).$$

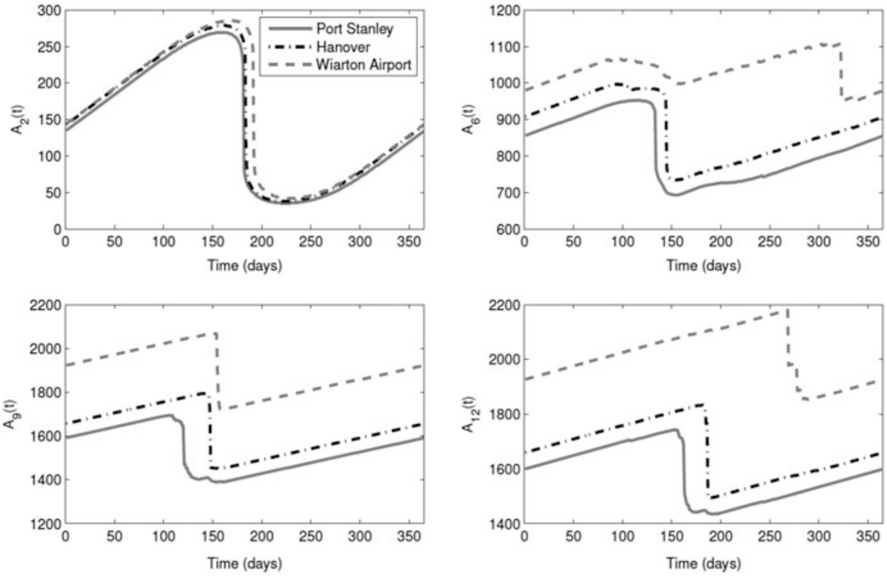


Fig. 5.3 Calculation of $A_{i-1}(t)$ and $A_i(t)$, the time-dependent minimum and maximum ages of those individuals who are developing within the specific ith stage. The calculations are based on the time-varying development delays for the period 1971–2000, in Port Stanley, Hanover and Wiarton Airport weather stations (Figure is taken from [39])

At this future time, we have

$$\frac{d}{dt}x_1(\tilde{t}) = \frac{d}{d\tilde{t}}x_1(\tilde{t})\frac{d\tilde{t}}{dt} = [a(\tilde{t})x_1(h(\tilde{t})) - \mu_1(0)x_1(\tilde{t})]\frac{1}{1 - A_n'(\tilde{t})}.$$

We write

$$\frac{d}{dt}x_1(\tilde{t}) = [a(h^{-1}(t))x_1(t) - \mu_1(0)x_1(h^{-1}(t))]\frac{1}{1 - A_n'(h^{-1}(t))}.$$

That is, the number of newly generated egg-laying females per unit time at time t is given by $y(t) = c(t)x_1(t)$ with

$$c(t) := a(h^{-1}(t))/(1 - A_n'(h^{-1}(t))).$$

Multiplying $x_1(t) = \int_{-\infty}^{t} e^{-\mu_1(0)(t-s)}a(s)x_1(s - A_n(s))ds$ by $c(t)$ gives

$$\begin{aligned} y(t) &= c(t)\int_{-\infty}^{t} e^{-\mu_1(0)(t-s)}\frac{a(s)}{c(s - A_n(s))}y(s - A_n(s))\,ds \\ &= \int_{A_n(t)}^{\infty} c(t)e^{-\mu_1(0)(t-h^{-1}(t-r))}y(t-r)\,dr \\ &= \int_0^{\infty} \mathscr{K}(t,r)y(t-r)\,dr \end{aligned}$$

with

$$\mathscr{K}(t,r) = \begin{cases} b'(0)\hat{\alpha}_n(h^{-1}(t))e^{-\mu_1(0)(t-h^{-1}(t-r))} & , r \geq A_n(t), \\ 0 & , r < A_n(t). \end{cases}$$

Note that $\mathscr{K}(t,r)$ is a periodic function with respect to time t, i.e., $\mathscr{K}(t,r) = \mathscr{K}(t+\omega, r)$. Biologically, this means that at time t, only the cohort of egg-laying females who are still alive before time $t - A_n(t)$ is capable of reproducing eggs which will mature to new generation of egg-laying females.

It is now natural to introduce

$$\mathscr{C}_\omega := \{u : R \to R \text{ is continuous }, u(t+\omega) = u(t)\},$$

equipped with maximum norm $\|\cdot\|$, and let $\mathscr{L} : C_\omega \to C_\omega$ be defined by

$$(\mathscr{L}u)(t) = \int_0^{\infty} \mathscr{K}(t,r)u(t-r)\,dr.$$

One can then show that $\mathscr{L}$ is strongly positive, continuous and compact on $\mathscr{C}_\omega$. This is called the *next generation operator* [7]. The basic reproductive number is defined as the spectral radius of the linear integral operator

$$\mathscr{R}_0 = \rho(\mathscr{L}).$$

In [39], it was proved that when $\mathscr{R}_0 < 1$, the zero solution is locally asymptotically stable; when $\mathscr{R}_0 > 1$, the zero solution is unstable. The proof is based on an application of the Krein–Rutman Theorem.[1] We refer to [17, 33] for some earlier results about the threshold R_0 in our setting. The approach of [39] follows more of [1, 2, 36].

5.3.3 Numerical Calculation of the Threshold: The Mathematics Behind a Lyme Tick Risk Map

Not only the size of R_0 relative to the unity is important to evaluate whether the vector population can establish in the region, but also the value of this R_0 is important to estimate the initial growth rate of population since near the zero equilibrium the solution from a non-trivial initial value grows exponentially with the rate $\ln R_0$ if $R_0 > 1$. Therefore, if we are to apply the above theory in a practical context, it is important to develop algorithms for calculating R_0.

One such algorithm was developed in [39] using the most intuitive discretization and integration. This algorithm links the calculation of $\mathscr{R}_0$ to the calculation of the spectral radius of a Leslie matrix in a periodic environment. In particular, to compute $\mathscr{R}_0$ numerically, we partition the interval $[0, \omega]$ into N (a large integer) subintervals of equal length. Set $t_i = (i-1)\omega/N$ for $i = 1, 2, \cdots, N$ and let $W_i = u(t_i)$. Then the problem of estimating $\mathscr{R}_0$ reduces to the calculation of the spectral radius of a Leslie matrix. Namely, we have the matrix eigenvalue problem of the form $\tilde{R}_0 W = XW$, where $W = (W_1, W_2, \cdots, W_N)^T$, and $\tilde{R}_0$ is the spectral radius of a $N \times N$ positive matrix X. In this matrix, the (i, j) element is given explicitly and with a clear biological interpretation for each metric element.

The calculated R_0 values for different regions and under different observed and predicted environmental conditions can then be used to depict the tick reproduction map for *I. scapularis*, see [38, 39]. This can then be used to estimate the impact of predicted climate change on tick population dynamics [25], as illustrated in Fig. 5.4. We should mention a recent study [5] that shows how remote sensing data can be further used to increase the spatial detail for this Lyme disease risk mapping. This

[1] In functional analysis, the Krein-Rutman theorem is a generalization of the Perron-Frobenius theorem to infinite-dimensional Banach spaces. It was proved by Krein and Rutman in 1948. The Krein-Rutman Theorem states that: Let X be a Banach space, and let $K \subset X$ be a convex cone such that $K - K$ is dense in X. Let $T : X \to X$ be a non-zero compact operator which is positive, meaning that $T(K) \subset K$, and assume that its spectral radius $\rho(T)$ is strictly positive. Then $\rho(T)$ is an eigenvalue of T with positive eigenvector, meaning that there exists $u \in K \setminus \{0\}$ such that $T(u) = \rho(T)u$.

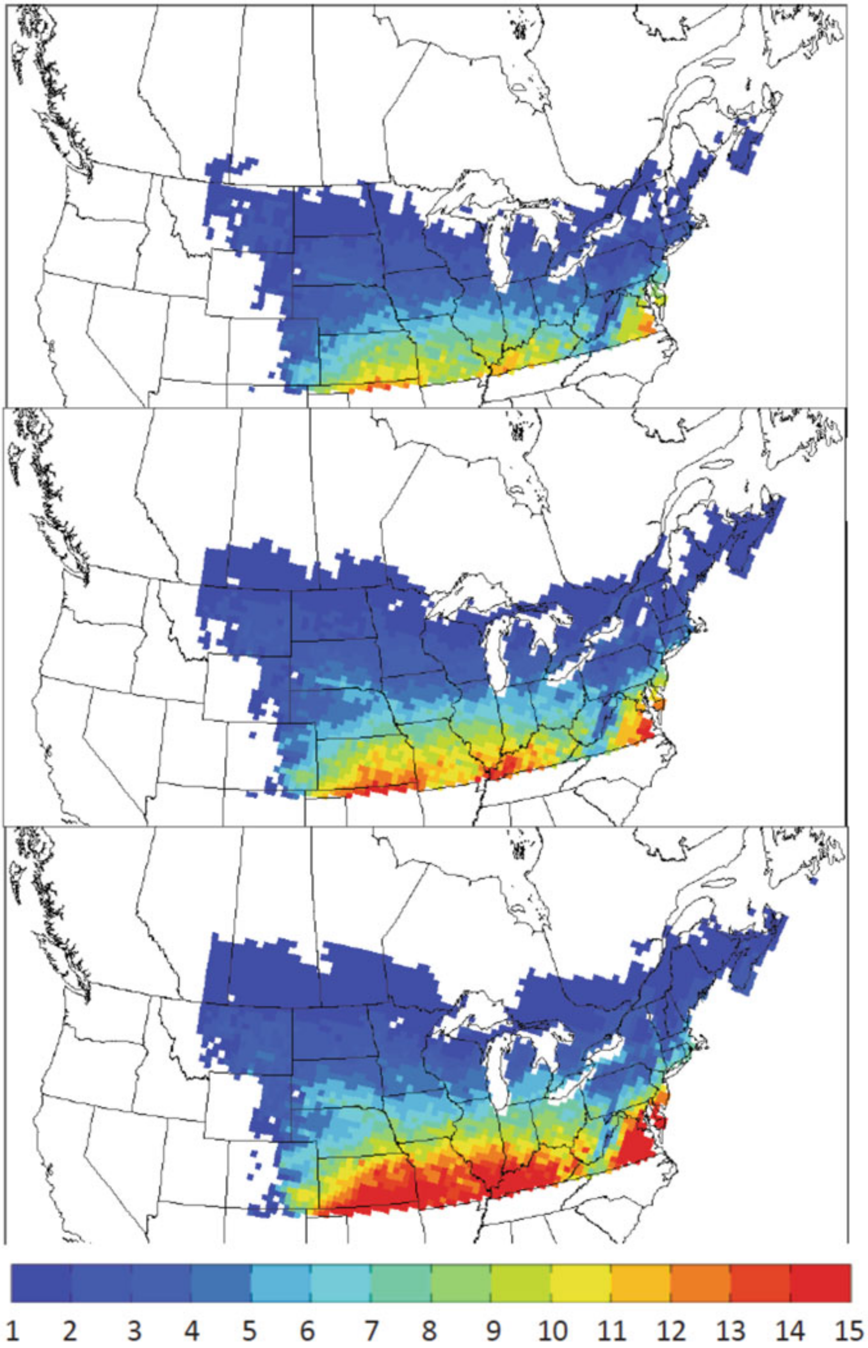


Fig. 5.4 Maps of values of R_0 in North America, estimated from climate observations (1971–2000: upper panel), and projected climate for 2011 to 2040 (middle panel) and for 2041 to 2070 (bottom panel). The color scale indicates R_0 values. Figure is taken from [25], and shows the northward expansion of the tick establishment due to climate warming. See [10] for a recent modelling study about the epidemic propagation speed and patterns in a wave-like environment as illustrated in the above maps

study also shows how well the risk map coincides with the number of ticks submitted to the Public Health Ontario, in the study area.

As another remark, we note that the equation for x_1 is de-coupled from the rest due to the use of the delay. However, for the purpose of Lyme disease risk projection, it is important to describe the density and variation of feeding nymphs and ticks in other stages since ticks in these stages are more involved in sharing the host for the Lyme pathogen transmission, and for human to get infection from the infected ticks. These densities can be described, using the formulation derived from Eq. (5.2), for $i = 2, \cdots, n$, given below:

$$\begin{aligned} x_i' = {} & \alpha_{i-1}(t, t - A_{i-1}(t))b(x_1(t - A_{i-1}(t)))(1 - A_{i-1}'(t)) \\ & -\alpha_i(t, t - A_i(t))b(x_1(t - A_i(t)))(1 - A_i'(t)) - \mu_i(x_i(t))x_i(t). \end{aligned}$$

5.4 Bird Migration Dynamics: Spatial Heterogeneity and Transition Delay

Structured population dynamics arises not only from temporally structured populations, but also spatially segregated populations. We illustrate this here with a model for bird migration. The model described is taken from a series of studies [3, 4, 12, 34] on avian influenza spread modelling. The central issue of this series of studies is seasonal bird migration dynamics and spatial-temporal distribution, and its implications for avian influenza spread patterns.

This series of studies has been guided by some satellite tracking data from the U.S. Geological Survey which recorded the migration path of a dozen bar-headed geese (from Mongolia to India). The data also shows that migration routes are often one-dimensional, as they tend to be funnelled into narrow pathways, often following coastlines or mountain ranges.

It is therefore natural that we start with the spatially explicit bird migration model using advection equations. Let x be the arc length along the continuum. Let $x_1 = 0$ be the summer breeding site, x_n be the winter feeding location and $x_i, i = 2, 3, ..., n - 1$ be the stopover locations where birds stop for short periods to feed. Let l_i be the distance between the locations x_i and x_{i+1} and U_i be the mean flight velocity between these two locations, so that the time taken to fly between x_i and x_{i+1} will be

$$\tau_i = l_i / U_i.$$

The density $s(t, x)$ obeys the advection equation

$$(\frac{\partial}{\partial t} + U_i \frac{\partial}{\partial x}) s(t, x) = -\mu_i s(t, x).$$

Let $S_i(t)$ be the number of birds at location x_i. Then at time t, the rate of birds leaving patch x_i is $d_i(t)S_i(t)$, with $d_i(t)$ being the rate of outward migration from patch i. The rate of birds arriving into patch x_{i+1} is

$$U_i s(t, x_{i+1}) = d_i(t-\tau_i)S_i(t-\tau_i)\alpha_i(t).$$

Using integration along characteristics, one can obtain the following bird migration patchy model:

$$\begin{cases} S_1'(t) = p(t)b(S_1(t)) + \alpha_n d_n(t-\tau_n)S_n(t-\tau_n) - d_1(t)S_1(t) - \mu_1(t)S_1(t), \\ \quad\vdots \\ S_i'(t) = \alpha_{i-1}d_{i-1}(t-\tau_{i-1})S_{i-1}(t-\tau_{i-1}) - d_i(t)S_i(t) - \mu_i(t)S_i(t), 2 \le i \le n. \end{cases}$$

It is natural to choose the phase space $C := \Pi_{i=1}^n C([-\tau_i, 0])$.

To describe the qualitative behaviors of the model equation, we will need the following section on discrete dynamical systems.

5.4.1 Monotone Maps and Threshold Dynamics

We start with introducing a few concepts:

- Let E be an ordered Banach space with positive cone P such that $\text{int}(P) \neq \emptyset$. For $x, y \in E$, we write $x \ge y$ if $x - y \in P$; $x > y$ if $x - y \in P \setminus \{0\}$ and $x >> y$ if $x - y \in \text{int}P$.
- Let $U \subset E$ and $f : U \to U$ be a given continuous map. We say that f is monotone if $x \ge y$ implies $f(x) \ge f(y)$; strongly monotone if $x > y$ implies $f(x) >> f(y)$.
- $f : U \to U$ is said to be strictly subhomogeneous if $f(\lambda x) > \lambda f(x)$ for any $x \in U$ with $x >> 0$ and $\lambda \in (0, 1)$.

In terms of the bird migration model, we define $f : \Pi_{i=1}^n C([-\tau_i, 0]) \to \Pi_{i=1}^n C([-\tau_i, 0])$ by $f(\phi) = (S_i(\phi)_\omega)_{i=1}^n$ for the period (ω)-operator of the model. We also define $P = \Pi_{i=1}^n C([-\tau_i, 0]; R^+)$. Then we have

- E is an ordered Banach space with positive cone P;
- $f : P \to P$ is monotone; S^m is strongly monotone when $m\omega \ge \max \tau_i$;
- Assume all $p(t)$ and $d_i(t)$ are positive and ω-periodic and positive (this assumption can be weaken), and $b(0) = 0$ and $b : [0, \infty) \to [0, \infty)$ being C^1 and strictly monotone. Therefore for any integer m such that $m\omega > \tau$, $S^m : U \to U$ is strongly monotone and precompact (i.e., the image of a bounded set in U under S^m is contained in a compact set).

We then have the following general result on threshold dynamics of monotone maps [41]:

Theorem 5.1 (Threshold Dynamics Theorem) *Let $f : P \to P$ be given such that*

(H1) *f is strongly monotone and strictly subhomogeneous;*
(H2) *f^m is precompact for some positive integer m, and every positive orbit $\{f^n(x); n = 1, 2, \cdots\}$ is bounded;*
(H3) *$f(0) = 0$ and $Df(0)$ is compact and strongly positive.*

Then the following threshold dynamics holds:

(TD1) *if $\rho(Df(0)) \leq 1$, then every positive orbit in P converges to 0;*
(TD2) *if $\rho(Df(0)) > 1$, then there exists a unique fixed point $u^* >> 0$ in P such that every positive orbit in $P \setminus \{0\}$ converges to u^*.*

To apply this for the bird migration model, we define R_0 as $\rho(Df(0))$ with f defined as above, then we conclude that if $R_0 \leq 1$, then $S_i(\phi) \to 0$ as $t \to \infty$ for all $\phi \in P$; if $R_0 > 1$, then the system has a unique positive ω-periodic solution such that every solution of the system starting from $P \setminus \{0\}$ converges to this positive periodic solution. Calculation of R_0 was performed in [35].

Despite this straightforward application of a general threshold dynamics theorem, the established global asymptotical stability of a unique positive solution is significant for the purpose of modelling bird influenza infection dynamics since this global stability result gives us the theoretical foundation to estimate the initial condition for bird influenza epidemic models. Namely, the theoretical result ensures that starting from an arbitrary initial condition, the solution is eventually stabilized at a unique positive periodic solution (assuming the threshold is larger than 1). This unique periodic solution, easily obtained through numerically simulating the bird migration model with an arbitrarily given initial data of birds for a sufficient period of time, gives the initial susceptible birds at the onset of a bird influenza outbreak. The long-term *Limiting* behaviors of an ecological model (bird migration dynamics) give the *Initial Condition* of the epidemic model for a considered bird flu outbreak.

5.5 Global Spread and Disease Epidemiology

The spread of avian flu with a particular strain such as H5N1 combines interactions between local and long-range dynamics. The local dynamics involve interactions/cross-contamination of domesticated birds, local poultry industry and temporary migratory birds. The nonlocal dynamics involve the long-range transportation of industrial material and poultry, and the long-range bird migrations (Fig. 5.5).

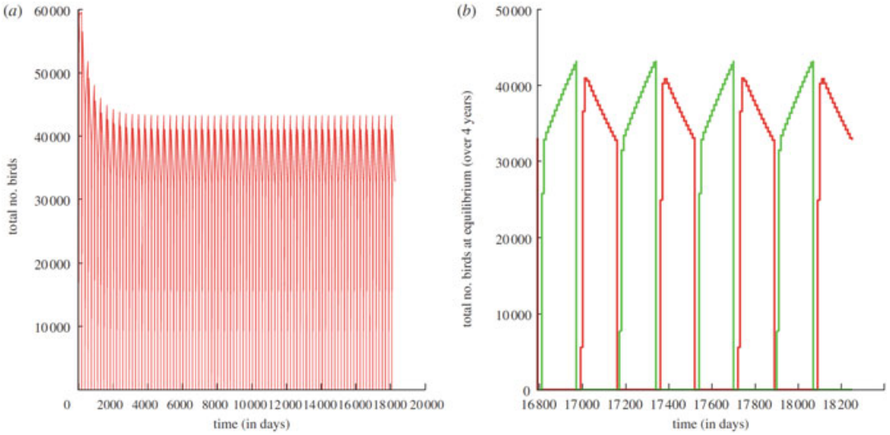


Fig. 5.5 The study [3] chose to focus on bar-headed geese as example species due to their vulnerability to the avian influenza H5N1, as highlighted by the death toll in the 2005 Qinghai Lake outbreak. The study used some satellite tracking data of bar-headed geese to extract the information of arrival, the length of stay and the date and time since deployment, as well as the average distance and time of flight between the current and previous stop sites. This information was then used to parameterize the model and to produce the simulation results showing here. The simulation shows that over a simulation of 50 years, the bird population reaches positive periodic solution. This periodic state is reached for all non-trivial initial conditions, illustrating the theoretically established global asymptotic stability of a unique periodic solution of the model equation

To model the interaction of migratory birds and domestic poultry we must stratify the migratory birds by their disease status and need to add domestic poultry. We use a patch model, where we consider four representative patches: breading ground (b), wintering ground (w), spring onward migration stop over (o) and fall migration, returning to the wintering ground, patch (r). Within each patch, we need to consider the migratory bird (subindex m) and domestic poultry (subindex p) populations, and both are needed to be stratified by their infection status, susceptible (s) and infected (i). Within each patch, we have the standard mass action for the disease transmission, and between patch we assume migration of the migratory birds. This yields the systems of differential equations for *Migratory Bird Dynamics*:

$$\begin{aligned}
\dot{S}_m^b &= B_m(t, S_m^b) + \alpha_{rb}^s d_{rb}^s S_m^r(t - \tau_{rb}^s) - \beta_m^b S_m^b I_m^b - \beta_{pm}^b S_m^b I_p^b - d_{bo}^s S_m^b - \mu_{ms}^b S_m^b,\\
\dot{I}_m^b &= \alpha_{rb}^i d_{rb}^i I_m^r(t - \tau_{rb}^i) + \beta_m^b S_m^b I_m^b + \beta_{pm}^b S_m^b I_p^b - d_{bo}^i I_m^b - \mu_{mi}^b I_m^b,\\
\dot{S}_m^o &= \alpha_{bo}^s d_{bo}^s S_m^b(t - \tau_{bo}^s) - \beta_m^o S_m^o I_m^o - \beta_{pm}^o S_m^o I_p^o - d_{ow}^s S_m^o - \mu_{ms}^o S_m^o,\\
\dot{I}_m^o &= \alpha_{bo}^i d_{bo}^i I_m^b(t - \tau_{bo}^i) + \beta_m^o S_m^o I_m^o + \beta_{pm}^o S_m^o I_p^o - d_{ow}^i I_m^o - \mu_{mi}^o I_m^o,\\
\dot{S}_m^w &= \alpha_{ow}^s d_{ow}^s S_m^o(t - \tau_{ow}^s) - \beta_m^w S_m^w I_m^w - \beta_{pm}^w S_m^w I_p^w - d_{wr}^s S_m^w - \mu_{ms}^w S_m^w,\\
\dot{I}_m^w &= \alpha_{ow}^i d_{ow}^i I_m^o(t - \tau_{ow}^i) + \beta_m^w S_m^w I_m^w + \beta_{pm}^w S_m^w I_p^w - d_{wr}^i I_m^w - \mu_{mi}^w I_m^w,\\
\dot{S}_m^r &= \alpha_{wr}^s d_{wr}^s S_m^w(t - \tau_{wr}^s) - \beta_m^r S_m^r I_m^r - \beta_{pm}^r S_m^r I_p^r - d_{rb}^s S_m^r - \mu_{ms}^r S_m^r,\\
\dot{I}_m^r &= \alpha_{wr}^i d_{wr}^i I_m^w(t - \tau_{wr}^i) + \beta_m^r S_m^r I_m^r + \beta_{pm}^r S_m^r I_p^r - d_{rb}^i I_m^r - \mu_{mi}^r I_m^r
\end{aligned}$$

coupled with the system for *Poultry Population Dynamics*:

$$\begin{aligned}
\dot{I}_p^b &= \beta_p^b(N_p^b - I_p^b)I_p^b + \beta_{mp}^b(N_p^b - I_p^b)I_m^b - \mu_p^b I_p^b,\\
\dot{I}_p^o &= \beta_p^o(N_p^o - I_p^o)I_p^o + \beta_{mp}^o(N_p^o - I_p^o)I_m^o - \mu_p^o I_p^o,\\
\dot{I}_p^w &= \beta_p^w(N_p^w - I_p^w)I_p^w + \beta_{mp}^w(N_p^w - I_p^w)I_m^w - \mu_p^w I_p^w,\\
\dot{I}_p^r &= \beta_p^r(N_p^r - I_p^r)I_p^r + \beta_{mp}^r(N_p^r - I_p^r)I_m^r - \mu_p^r I_p^r.
\end{aligned}$$

In [12], a threshold, given in terms of the spectral radius $r(T_I)$ of the time T-solution operator of the linearized periodic system of delay differential equations at the disease-free equilibrium, was theoretically derived. A closed form in terms of the model parameters is possible in some special cases. It was then shown that this threshold determines whether disease persists or not: the non-trivial disease-free equilibrium is globally asymptotically stable once the threshold is below 1; if the threshold is larger than 1, then the disease is uniformly strongly persistent in the sense that there exists some constant $\eta > 0$, which is independent of the initial conditions, such that, for each $c = b, o, w, r$,

$$\liminf_{t\to\infty} I_m^c(t) \geq \eta, \qquad \liminf_{t\to\infty} I_p^c(t) \geq \eta.$$

This result is based on the persistence theory discussed below: Let X be a complete metric space with the metric d. Let X_0 and ∂X_0 be open and closed subsets of X, respectively, such that $X_0 \cap \partial X_0 = \emptyset$ and $X = X_0 \cup \partial X_0$. Let $S : X \to X$ be a continuous map with $S(X_0) \subset X_0$. We introduce a few concepts here:

- S is uniformly persistent with respect to $(X_0, \partial X_0)$ if there exists $\eta > 0$ such that for any $x \in X_0$, $\liminf_{n\to\infty} d(S^n x, \partial X_0) \geq \eta$;
- A nonempty invariant set $M \subset \partial X_0$ is isolated if it is the maximal invariant set in some neighbourhood of itself;
- An isolated set $A \subset \partial X_0$ is chained to an isolated set $B \subset \partial X_0$, written as $A \to B$, if there exists a full orbit through some $x \notin A \cup B$ such that $\omega(x) \subset B$ and $\alpha(x) \subset A$;
- A finite sequence $\{M_1, \cdots, M_k\}$ of invariant sets is called a chain if $M_1 \to M_2 \to \cdots M_k$. The chain is called a cycle if $M_k = M_1$

We refer to [16, 32, 41] for more systematic treatments of the persistence theory, but the theorem below is what was used in [12]:

Theorem 5.2 *Assume that*

- *$S : X \to X$ has a global attractor;*
- *Let A_δ be the maximal compactor invariant set of S in ∂X_0. $\tilde{A}_\delta = \cup_{x\in A_\delta}\omega(x)$ has an isolated and acyclic covering $\cup_{i=1}^k M_i$ in ∂X_0 (that is, $A_\delta \subset \cup_{i=1}^k M_i$, where $M_1, M_2, \cdots, M_k$ are pairwise disjoint and compact and isolated invariant sets of S in ∂X_0 such that each M_i is also an isolated invariant set in X, and no subset of the M_i's forms a cycle for $S_\delta = S|_{A\delta}$ in A_δ).*

Then S is uniformly persistent if and only if for each M_i, we have $W^s(M_i) \cap X_0 = \emptyset$, where $W^s(M_i) = \{x, x \in X, \omega(x) \neq \emptyset, \omega(x) \subset M_i\}$ is the stable set of M_i.

The persistence theory, when applied to the above avian influenza model, concludes that the avian influenza spread persists in the sense that both infected migratory and domestic poultry birds will remain strictly larger than a unspecified constant. Numerical simulations have indicated that the pattern of disease persistence can be quite complicated, and is not necessarily fluctuating regularly as an annual cycle. This raises an issue about the estimation of inter-pandemic and intra-pandemic intervals. There seems to be no theoretical framework that has been applied to address this important practical issue.

Lyme disease dynamics was also considered in the study [20] with standard stratification of tick populations by the infection status and by tick development stages. The first such model is to assume the development rate is exponentially distributed (and time-independent). This leads to an epidemic system of ordinary differential equations with periodic coefficients. This formulation facilitates refined persistence results about the periodicity of persistent disease spread patterns. It remains to see whether the introduction of periodically varying delay will make the model analysis much more complicated. From the public health prospective, it would be desirable to establish not only the Lyme tick risk map, but Lyme disease risk map—in terms of the threshold values of the epidemic models.

5.6 Summary

In this chapter, we consider modelling environment impact on vector-borne infection dynamics using delay differential equations. This is based on a series of sections which introduce a general framework using delay differential equations, and relevant results on global dynamics and persistence about the implication of environment changes for the interplay of vector species ecology and vector-borne disease epidemiology. General results are illustrated by applications to avian influenza and Lyme disease spread.

We first consider spatiotemporal patterns of bird migration and seasonal stage-activities of tick populations with focus on model formulation and parameterization. Here, we derive, from first-order hyperbolic partial differential equations, prototype delay differential equations describing the spatial dynamics of migratory birds and stage-structured tick population dynamics. The periodicity in model coefficients and delays arises due to seasonality. We illustrate how surveillance, laboratory, field study and satellite/remote sensing data can be integrated to parameterize the models.

We then use the model to describe spatiotemporal patterns of bird migration and seasonal stage-activities of tick populations: global dynamics. We describe the phase space and general framework for the qualitative behaviors of delay differential equations with periodic coefficients/delays and examine the global dynamics of the model systems using the monotone dynamical systems theory. We discuss the impact of climate changes on vector establishment risks.

We finally consider avian influenza spread and Lyme disease epidemics: persistence and irregular infection dynamics. Here we stratify the vector populations in terms of their infection status (susceptible or infectious) and obtain corresponding epidemic models. We introduce the concept and general results of infection persistence and threshold phenomena, and we discuss further challenges depicting the inter-epidemics and intra-epidemic intervals.

There are a number of challenging issues for the modelling, parameterization, dynamic behavior analysis and numerics of structured population models arising from vector-borne disease infection risk assessment consideration. Such a model framework seems to be appropriate given the important role of the physiological or geographical status of the vector species in defining the vector population dynamics and the disease spread. The reduction from the structured population models to delay differential equations is both mandated and facilitated by the fact that surveillance data is normally collected for the vector in a certain physiological stage or a geographic location, and this reduction also renders the well-established dynamical systems theory of delay differential equations applicable to considering some important ecological and epidemiological systems.

References

1. N. Bacaër, Approximation of the basic reproduction number R_0 for vector-borne diseases with a periodic vector population. Bull. Math. Biol. **69**, 1067–1091 (2007)
2. N. Bacaër, R. Ouifki, Growth rate and basic reproduction number for population models with a simple periodic factor. Math. Biosci. **210**, 647–658 (2007)
3. L. Bourouiba, J. Wu, S. Newman, J. Takekawa, T. Natdorj, N. Batbayar, C.M. Bishop, L.A. Hawkes, P.J. Butler, M. Wikelski, Spatial dynamics of bar-headed geese migration in the context of H5N1. J. R. Soc. Interface **7**(52), 1627–1639 (2010)
4. L. Bourouiba, S. Gourley, R. Liu, J. Wu, The interaction of migratory birds and domestic poultry, and its role in sustaining avian influenza. SIAM J. Appl. Math. **71**, 487–516 (2011)
5. A. Cheng, D. Chen, K. Woodstock, O. Ogden, X. Wu, J. Wu, Analyzing the potential risk of climate change on Lyme disease in eastern Ontario, Canada using time series remotely sensed temperature data and tick population modelling. Remote Sens. **9**, 609 (2017)
6. O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Disease: Model Building, Analysis and Interpretation* (Wiley, New York, 2000)
7. O. Diekmann, J.A.P. Heesterbeek, M.G. Roberts, The construction of next-generation matrices for compartmental epidemic models. J. R. Soc. Interface **7**(47), 873–885 (2010)
8. O. Diekmann, S.A. van Gils, S.M. Verduyn Lunel, H.O. Walther, *Delay Equations, Functional-, Complex-, and Nonlinear Analysis* (Springer, New York, 1995)
9. T. Erneux, *Applied Delay Differential Equations* (Springer, Berlin, 2009)
10. J. Fang, Y. Lou, J. Wu, Can pathogen spread keep pace with its host invasion? SIAM J. Appl. Math. **76**(4), 1633–1657 (2016)
11. H. Freedman, J. Wu, Periodic solutions of single-species model with periodic delay. SIAM J. Math. Anal. **23**(3), 689–701 (1992)
12. S. Gourley, R. Liu, J. Wu, Spatiotemporal distributions of migratory birds: patchy models with delay. SIAM J. Appl. Dyn. Syst. **9**(2), 589–610 (2010)

13. S. Guo, J. Wu, *Bifurcation Theory of Functional Differential Equations* (Springer, New York, 2014)
14. J.K. Hale, *Theory of Functional Differential Equations* (Springer, New York, 1977)
15. J.K. Hale, S.M. Verduyn Lunel, *Introduction to Functional Differential Equations* (Springer, New York, 1993)
16. J.K. Hale, P. Waltman, Persistence in infinite-dimensional systems. SIAM J. Math. Anal. **20**, 388–395 (1989)
17. P. Jagers, O. Nerman, Branching processes in periodically varying environment. Ann. Prob. **13**, 254–268 (1985)
18. T. Krisztin, H. Walther, J. Wu, Shape, Smoothness and invariant stratification of an attracting set for delayed positive feedback, in *Fields Institute Monograph Series*, vol. 11 (American Mathematical Society, Providence, 1996)
19. Y. Kuang, *Delay Differential Equations: with Applications in Population Dynamics* (Academic Press, Springer, Berlin, 2013)
20. Y. Lou, J. Wu, X. Wu, Impact of biodiversity and seasonality on Lyme-pathogen transmission. Theor. Biol. Med. Model. **11**(1), 50 (2014)
21. J.A.J. Metz, O. Diekmann, *The Dynamics of Physiologically Structured Population* (Springer, Heidelberg, 1986)
22. N.H. Ogden, L.R. Lindsay, G. Beauchamp, D. Charron, A. Maarouf, C.J. O'Callaghan, D. Waltner-Toews, I.K. Barker, Investigation of relationships between temperature and developmental rates of tick *Ixodes scapularis* (Acari: Ixodidae) in the laboratory and field. J. Med. Entomol. **41**, 622–633 (2004)
23. N.H. Ogden, M. Bigras-Poulin, C.J. O'Callaghan, I.K. Barker, L.R. Lindsay, A. Maarouf, K.E. Smoyer-Tomic, D. Waltner-Toews, D. Charron, A dynamic population model to investigate effects of climate on geographic range and seasonality of the tick *Ixodes scapularis*. Int. J. Parasitol. **35**, 375–389 (2005)
24. N.H. Ogden, A. Maarouf, I.K. Barker, M. Bigras-Poulin, L.R. Lindsay, M.G. Morshed, C.J. O'Callaghan, F. Ramay, D. Waltner-Toews, F.F. Charron, Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. Int. J. Parasitol. **36**, 63–70 (2006)
25. N.H. Ogden, M. Radojevié, X. Wu, V.R. Duvvuri, P. Leighton, J. Wu, Estimated effects of projected climate change on the basic reproductive number of the Lyme disease vector *Ixods scapularis*. Environ. Health. Perspect. **122**(6), 631 (2014)
26. G. Röst, J. Wu, Domain-decomposition method for the global dynamics of delay differential equations with unimodal feedback. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. **463**(2086), 2655–2669 (2007)
27. H. Shu, L. Wang, J. Wu, Global dynamics of Nicholson's blowflies equation revisited: onset and termination of nonlinear oscillations. J. Differ. Equ. **255**(9), 2565–2586 (2013)
28. H. Shu, L. Wang, J. Wu, Bounded global Hopf branches for stage-structured differential equations with unimodal feedback. Nonlinearity **30**, 943–964 (2017)
29. H. Smith, Monotone semiflows generated by functional differential equations. J. Differ. Equ. **66**(3), 420–442 (1987)
30. H.L. Smith, *An Introduction to Delay Differential equations with Applications to the Life Sciences* (Springer, New York, 2010)
31. H. Smith, H. Thieme, Quasi convergence and stability for strongly order-preserving semiflows. SIAM J. Math. Anal. **21**(3), 673–692 (1990)
32. H.L. Smith, H.R. Thieme, *Dynamical Systems and Population Persistence* (American Mathematical Society, Providence, 2011)
33. H.R. Thieme, Renewal theorems for linear periodic Volterra integral equations. J. Integr. Equ. **7**, 253–277 (1984)
34. X.S. Wang, J. Wu, Seasonal migration dynamics: periodicity, transition delay, and finite dimensional reduction. Proc. R. Soc. A. **468**, 634–650 (2012)
35. X.S. Wang, J. Wu, Periodic systems of delay differential equations and avian influenza dynamics. J. Math. Sci. **201**, 693–704 (2014)

36. W. Wang, X.Q. Zhao, Threshold dynamics for compartmental epidemic models in periodic environments. J. Dyn. Diff. Equat. **20**, 699–717 (2008)
37. J. Wu, *Theory of Partial Functional Differential Equations* (Springer-Verlag, New York, 1996)
38. X. Wu, V. Duvvuri, Y. Lou, N. Ogden, Y. Pelcat, J. Wu, Developing a temperature-driven map of the basic reproductive number of the emerging tick vector of Lyme disease *Ixodes scapularis* in Canada. J. Theor. Biol. **319**, 50–61 (2013)
39. X. Wu, F. Magpantay, J. Wu, Z. Zou, Stage-structured population systems with temporally periodic delay. Math. Methods Appl. Sci. **38**, 3464–3481 (2015)
40. X. Zhang, X. Wu, J. Wu, Critical contact rate for vector-host-pathogen oscillation involving co-feeding and diapause. J. Biol. Syst. **25**, 657 (2017)
41. X. Zhao, *Dynamical Systems in Population Biology* (Springer, New York, 2003)

Chapter 6
Stochastic Population Kinetics and Its Underlying Mathematicothermodynamics

Hong Qian

Abstract Based on differential calculus, classical mechanics represents the natural world in terms of featureless point masses and their movements. Chemistry studies molecules each of which has a large number of internal degrees of freedom in terms of atoms, electrons, etc.; the behavior of even a single biomolecule like a protein is often so complex that the foundation of chemical kinetics is essentially based on stochastic mathematics. Stochastic population kinetics is a more powerful and more realistic representation of the biological world. This chapter introduces this new mathematical modeling paradigm and shows the existence of a hidden thermodynamic structure underlying any stochastic nonlinear kinetic description of a multi-population biological system. The mathematicothermodynamics presented here is a generalization of J. W. Gibbs' chemical thermodynamics for equilibrium chemical reaction systems, as heterogeneous matters.

6.1 Introduction

Françis Jacob (1920–2013), one of the leading molecular biologists of the twentieth century, stated in his book "The Possible and the Actual" [13] that Western art had radically changed since the Renaissance from "symbolizing" to "represent" the real world. One can in fact view pure versus applied mathematics as a change from the former to the latter. The ultimate goal of mathematical science is to quantitatively represent the real world in terms of mathematics.

Currently there is a sharp contrast between the mathematical models, or theories, in physics and in biology. While we take Newton's equation of motion as almost the "Truth" under the appropriate conditions, one does not have such a level of confidence for the mathematical models in biology.

H. Qian (✉)
Department of Applied Mathematics, University of Washington, Seattle, WA, USA
e-mail: qian@amath.washington.edu

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_6

In Newtonian mechanics, the natural world is represented by point masses and described by their movements. Each point mass, e.g., a Newtonian particle, has a unique position and velocity. The natural world according to chemistry, however, consists of "identical" molecules made of atoms. While each individual molecule has intrinsic stochasticity, e.g., a molecular individualism [4] due to the atomic motions within, population wise molecules follow statistical rate laws in their syntheses (birth), degradations (death), spatial diffusion (migration), state transitions (character switching), and interactions. Such a formal reaction kinetic system in a small volume V, such as biochemical reaction kinetics in a single cell, can be rigorously treated in terms of an integer-valued, continuous-time Markov process describing its nonlinear behavior, counting the molecules and their reactions, one at a time.

Population dynamics in biology has long been described in terms of nonlinear differential equations [17]. Many of the equations are remarkably similar to the kinetic equation for chemical reactions. In this chapter, we shall introduce in a rigorous fashion the rate law of rare events in term of exponential waiting time and the Poisson process. We shall show that the type of differential equations for population dynamics has a mathematical foundation in the theory of probability and Markov processes.

After introducing the stochastic mathematical representation of population kinetics, in the Sect. 6.9 of the chapter, we present a recently discovered universal mathematical structure that is inherent in any Markov population kinetics. This structure has a remarkable resemblance to the theory of thermodynamics, first developed in the nineteenth century by physicists dealing with heat—the stochastic motions of atoms and molecules. To distinguish the mathematical structure in the stochastic population kinetics from the subject from physics, we coined the term ***mathematicothermodynamics***, within which we axiomatically introduce notions such as closed systems, open-driven systems, entropy production, free energy dissipation, etc. We shall derive two "laws": The first is concerned with the balance of a free energy like function, and the second is concerned with certain monotonicity in the dynamics.

Finally, phase transition in physics, conformational transition in biochemistry, and phenotypic switching in cell biology are all nonlinear phenomena intrinsically related to multi-stability and saddle-node bifurcation, in the limits of time $t \to \infty$ and system's size $V \to \infty$ [12, 26].

6.2 Probability and Stochastic Processes: A New Language for Population Dynamics

There are fundamentally two types of mathematical modeling: (a) representing scientific data in terms of mathematical formula or equations and (b) describing a system's behavior (natural or engineered, physical or biological, electronic, chemical, economical, social, ...) based on existing, established formula and equations. For lack of better terminology, we shall call the former ***data-driven modeling*** and the latter ***mechanistically derived modeling***. Note, according to Karl Popper (1902–1994) and his philosophy of science, the only legitimate scientific activity is falsifying a hypothesis: that requires first to formulate a hypothesis, which sometime is just looking for patterns in the data (e.g., numerical hypothesis) and sometime is proposing a mechanism (e.g., modeling); and (b) to derive rigorous predictions from a hypothesis, which is a form of logical, or mathematical, deduction.

Let us revisit some of the key notions already discussed, or widely used, in many of the other chapters—but let us try to be critical. In Chap. 1 Hillen and Lewis introduced the growth rate through a limiting process: if a population grows two person every 100 days, then it is "equivalent" to one person every 50 days, and half a person every 25 days. In fact, the growth rate is

$$r = \lim_{\Delta t \to 0} \frac{P(t + \Delta t) - P(t)}{\Delta t}.$$

Instantaneous rate (fluxion) is one of the most important concepts of Newton's calculus! But does this make sense to quantify population growth? A half of a person, one tenth of a person? Clearly this theory cannot be true when the Δt is too small: *population change* cannot have non-integer numbers.

Second, has anyone ever seen such a regular population growth with exactly two person in the first 100 days, and another two in the next 100 days? I am sure some of you will say "that is just an average".

Indeed, *discreteness* and *probability* are two fundamental issues in any population dynamics. Both have been ignored in the differential equation-based description of population dynamics. We shall start discussing population kinetics anew below. Most of the materials are taken from [1, 19, 20, 22, 23, 28, 31].

6.2.1 Brief Review of Elementary Probabilities

A ***random variable*** X taking a continuous real value has a probability density function (pdf) $f_X(x)$:

$$\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1, \quad f_X(x) \geq 0. \tag{6.1}$$

The meaning of the $f_X(x)$ is this: for infinitesimal $\mathrm{d}x$, the probability of observing $X \in (x, x + \mathrm{d}x]$ is $f_X(s)\mathrm{d}x$:

$$\Pr\{x < X \leq x + \mathrm{d}x\} = f_X(x)\mathrm{d}x. \tag{6.2}$$

Then, the cumulative probability distribution of X is defined as

$$F_X(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} f_X(z)\mathrm{d}z, \text{ and } f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}. \tag{6.3}$$

The mean (or expected value) and variance of the random variable X then are

$$\langle X \rangle = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x, \tag{6.4}$$

$$\mathrm{Var}[X] = \mathbb{E}\big[(X-\mu)^2\big] = \int_{-\infty}^{\infty} \big(x-\mu\big)^2 f_X(x)\mathrm{d}x, \tag{6.5}$$

in which we have denoted $\mathbb{E}[X]$ by μ. Two most important examples of random variables taking real values are “exponential” and “normal”, also called Gaussian. The former has the standard form

$$f_X(x) = \lambda e^{-\lambda x}, \; x \ge 0, \; \lambda > 0, \tag{6.6}$$

with mean and variance being λ^{-1} and λ^{-2}; the latter has a standard form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \tag{6.7}$$

with mean μ and variance σ^2.

Gaussian normal distribution is widely discussed; including in popular press [11]. It is understood as a consequence of the *central limit theorem*. It is a statistical law emerging from a large collection of identical, independent parts. In the following sections, we shall show that for dynamical processes involving populations, there is a much less known, but equally if not more important statistical law: exponentially distributed time between “rare events”. In stochastic modeling of population dynamics, one’s primary focus is not the random number of individuals at a particular time; rather it is the random time of the next event that changes the number of individuals by one.

The best known discrete, integer-valued random variables are Bernoulli, binomial, Poisson, and geometric [30].

6.2.2 Radioactive Decay and Exponential Time

Let us revisit the simplest differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y, \tag{6.8}$$

where $\lambda > 0$. This equation has been introduced as a mathematical model for the remaining fraction of a radioactive material at time t

$$\frac{y(t)}{y(0)} = e^{-\lambda t}. \tag{6.9}$$

If all the atomic nuclei are *identical and independent*, then

$$\text{Pr}\big\{\text{a nucleus remaining radioactive at time } t\big\} = e^{-\lambda t}. \tag{6.10}$$

However, if T is the random time at which the event of radioactive decay occurs, then

$$\text{Pr}\big\{\text{a nucleus remaining radioactive at time } t\big\} = \text{Pr}\big\{T \geq t\big\}. \tag{6.11}$$

T is a non-negative real-valued random variable with cumulative probability distribution $F_T(t) = \text{Pr}\{T \leq t\} = 1 - e^{-\lambda t}$ and probability density function $f_T(t) = \mathrm{d}F(t)/\mathrm{d}t = \lambda e^{-\lambda t}$.

What types of problems, or more precisely "scenarios" and "mechanisms", will give rise to this exponentially distributed waiting time? Why is it so universal? A good understanding of these questions will provide the reader a deeper understanding of the mathematical foundation of population dynamics, as emergent statistical laws, in terms of seemingly random behavior of a large population of individuals [15].

6.2.2.1 Rare Event

Let T be the random time at which a certain event occurs. If the occurrence of such an event is independent in time intervals $[t_1, t_2]$ and $[t_2, t_3]$, and if its occurrence is uniform in time (e.g., the system and its environment are stationary), then

$$\text{Prob. of no event occurring in } [0, t + \Delta t] = \tag{6.12}$$

$$\text{Prob. of no event occurring in } [0, t] \times \text{Prob. of no event occurring in } [t, t + \Delta t].$$

That is,

$$\text{Pr}\big\{T > t + \Delta t\big\} = \text{Pr}\big\{T > t\big\} \times \text{Prob. of no event occurring in } [t, t + \Delta t].$$

Now if the probability of one such event occurring in the time interval $[t, t + \Delta t]$ is proportional to Δt, and the probability of more than one events is $\propto o(\Delta t)$, then

$$\text{Pr}\big\{T > t + \Delta t\big\} = \text{Pr}\big\{T > t\big\} \times \Big(1 - \lambda \Delta t + o(\Delta t)\Big). \tag{6.13}$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pr}\{T > t\} = -\lambda \mathrm{Pr}\{T > t\}, \implies F_T(t) = e^{-\lambda t}. \tag{6.14}$$

Example: The waiting time for the first shopper coming in a store in the morning on a regular day.

6.2.2.2 Memoryless

One of the most important, in fact defining, properties of exponential distributed waiting time is

$$\frac{\mathrm{Pr}\{T \geq t+\tau\}}{\mathrm{Pr}\{T \geq t\}} = \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda t}} = e^{-\lambda\tau}. \tag{6.15}$$

Example: You and your lazy brother doing experiments to observe the mean time of an exponentially distributed event. Even though your brother starts counting time a whole hour later than you, his resulting statistics will be exactly the same as yours!

More interestingly, the more individuals in a population, the faster the next event to occur. In mathematical terms: if all $T_k \sim \lambda_k e^{-\lambda_k t}$ and they are independently distributed, then $T^* = \min(T_1, T_2, \cdots, T_n)$ also has an exponential distribution

$$\begin{aligned} \mathrm{Pr}\{T^* > t\} &= \mathrm{Pr}\{T_1 > t, \cdots, T_n > t\} \\ &= \mathrm{Pr}\{T_1 > t\} \times \mathrm{Pr}\{T_2 > t\} \times \cdots \times \mathrm{Pr}\{T_n > t\} = e^{-\mu t}, \end{aligned} \tag{6.16}$$

where $\mu = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Thus, $f_{T^*}(t) = \mu e^{-\mu t}$.

6.2.2.3 Minimal Time of a Set of Non-Exponential i.i.d. Random Times

Now consider a set of random times $\{T_k\}$. They are *identical, independently distributed* (i.i.d.) random times with pdf $f_T(t)$ and cumulative probability distribution $F_T(t)$. Then $T^* = \min(T_1, T_2, \cdots, T_n)$ has its distribution

$$\mathrm{Pr}\{T^* > t\} = \Big(1 - F_T(t)\Big)^n. \tag{6.17}$$

Now, introducing scaled $\hat{T}^* = nT^*$ and considering n to be very large, its distribution is

$$\mathrm{Pr}\Big\{\hat{T}^* > t\Big\} = \left(1 - F_T\left(\frac{t}{n}\right)\right)^n \simeq \left(-\frac{F_T'(0)}{n}t + O\big(n^{-2}\big)\right)^n \to e^{-F_T'(0)t}. \tag{6.18}$$

Therefore, if $F'_T(0) = f_T(0)$ is finite, one obtains an exponentially distributed time.

We note the mathematical condition $f_T(0) > 0$: in an application, this implies that the time scale involved in the mechanism for the occurrence of an event is several orders of magnitude faster than the time scale in question.

6.2.3 Known Mechanisms That Yield an Exponential Distribution

In the previous section, we have derived the exponentially distributed waiting time based on some very elementary assumptions concerning (1) time homogeneous and (2) independent. Furthermore, in Sect. 6.2.2.3, we have shown that for non-exponential T, as long as $f_T(0) \neq 0$, the minimum of a large collection of i.i.d. T's will be exponential. This is a strong argument for why one can use, on an appropriate time scale, the equations like (6.8) to model population dynamics.

6.2.3.1 Khinchin's Theorem

Let us consider a house that uses n light bulbs. One bought a large box of new light bulbs, and let us assume all the bulbs having identical, independently distributed life time X with pdf $f_X(x)$. For each light-bulb socket, one puts on a new bulb when the old one is burnt. The time sequence $0, T_1, T_2, \cdots, T_k, \cdots$ is called a *renewal process*, in which $T_k = \sum_{\ell=1}^{k} X^{(\ell)}$, where the $X^{(\ell)}$ with different ℓ are i.i.d. random variables drawn from the distribution $f_X(x)$. Now for the entire house, there are n identical, independent renewal processes. The time sequence of bulb changing form a *superposition* of the n renewal processes [3], as illustrated in Fig. 6.1.

For a single renewal process with renewal time distribution $f_X(x)$, the corresponding counting process, e.g., the number of renewals occurred before time t, N_t, has the distribution

$$\Pr\left\{N_t \geq k\right\} = \Pr\left\{T_k \leq t\right\} = F_{T_k}(t) = \int_0^t f_{T_k}(x)\mathrm{d}x. \tag{6.19}$$

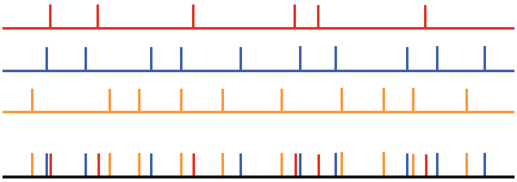


Fig. 6.1 If the red, orange, and blue point processes represent the renewal events of light bulbs for 3 different sockets, then the fourth row is the combined point process for all the bulb changes. It is the superposition of the three individual processes. With more and sockets, a statistical law emerges

Therefore,

$$\Pr\left\{N_t = k\right\} = F_{T_k}(t) - F_{T_{k+1}}(t). \tag{6.20}$$

Now if one randomly picks a time t, and let T_t^* be the waiting time for the next renewal, T_t^* is known as residual time in renewal theory. Its distribution is different from $f_X(x)$. In fact, one has

$$\begin{aligned}\Pr\{T_t^* \le s\} &= \sum_{\ell=0}^{\infty} \Pr\left\{N_t = \ell\right\} \Pr\left\{T_{\ell+1} \le t + s\right\} \\ &= \sum_{\ell=0}^{\infty} \Big(F_{T_\ell}(t) - F_{T_{\ell+1}}(t)\Big) F_{T_{\ell+1}}(t+s). \end{aligned} \tag{6.21}$$

Therefore, the probability density function for the stationary T_t^* is

$$f_{T_t^*}(s) = \frac{\mathrm{d}}{\mathrm{d}s} \Pr\{T_t^* \le s\}. \tag{6.22}$$

$$f_{T_t^*}(0) = \sum_{\ell=1}^{\infty} \Big(F_{T_\ell}(t) - F_{T_{\ell+1}}(t)\Big) f_{T_\ell}(t) \neq 0. \tag{6.23}$$

Applying the result in Sect. 6.2.2.3, we then have the following theorem, which can be found in [3].

Theorem *If $T_k^{(1)}$, $T_k^{(2)} \cdots$, $T_k^{(n)}$ are n i.i.d. renewal processes with waiting time distribution $f_X(x)$, then the superposition of the n renewal processes has an exponential waiting time for the next event in the limit of $n \to \infty$, with rate parameter $n\mathbb{E}^{-1}[X]$.*

6.2.3.2 Kramers' Theory and Saddle-Crossing as a Rare Event

We have discussed the minimal time of a large collection of i.i.d. waiting times, and we have discussed superposition of renewal processes. We now turn to a third mechanisms: the emergence of discrete chemical reactions from a description of atoms continuously moving in a molecule in an aqueous solution.

From a classical mechanics stand point, a molecule is a collection of atoms. For a protein with N number of atoms, a Newtonian mechanical description of its dynamics has $6N$ degrees of freedom, without even considering the atoms in the solvent, which is at least an order of magnitude more. This is what one observes from a molecular dynamics (MD) simulation. It is very complicated.

However, any such mechanical system has a potential energy function (its gradient is called a force field of MD simulations). Treating the solvent as a viscous

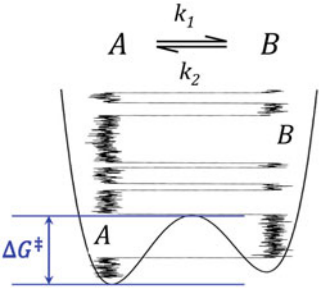


Fig. 6.2 The mathematical description of a chemical reaction of a single molecule. It is an emergent statistical law of a large number of discrete, stochastic reactions. $k_1 \propto e^{-\Delta G^\ddagger/k_BT}$, where the $\Delta G^\ddagger$ is called *activation energy*. Similarly, k_2 has its own activation barrier height. According to this description, the ratio k_1/k_2 becomes independent of the barrier

medium with frictional coefficient η, the dynamics of a protein is over damped and spends most of the time at the bottom of an "energy well", as illustrated in Fig. 6.2. However, since the solvent is not truly continuous, but rather corpuscular, the collisions with the solvent molecules constitute a random force. Therefore, the dynamics can be described by a stochastic differential equation like

$$\mathrm{d}Y(t) = b(Y)\mathrm{d}t + A\mathrm{d}B(t), \tag{6.24}$$

in which $b(y) = -\eta^{-1}\nabla_y U(y)$, and $A = \sqrt{2\eta^{-1}k_BT}$.

With the presence of random forcing term $B(t)$, $Y(t)$ will once a while move against the deterministic force field and even cross the barrier (a saddle point in a high-dimensional space). But this is a rare event. This randomly perturbed nonlinear dynamical systems thus behaves, on a very long time scale, as $A \rightleftharpoons B$, with only two parameters k_1 and k_2. The rate constants are related to the height of the barrier. H. A. Kramers first worked out the mathematical theory for this type of problems in 1940. The idea is not limited to chemical reactions; it is applicable to any nonlinear dynamics with random perturbations [7].

With one line of mathematics from Kramers, $k \propto e^{-\Delta G^\ddagger/k_BT}$ (Fig. 6.2), all the detailed atomic motions are deemed irrelevant—only two parameters, called forward and backward rate constants, are useful to a chemist. Furthermore, the theory shows that the transition from $A \rightarrow B$ spends most of the time in the waiting; the actual transition event is instantaneous! Indeed, one can mathematically prove in the limit of $\Delta G^\ddagger/k_BT \rightarrow \infty$, the waiting time distribution asymptotically approaches to exponential. From a molecular biological function perspective, the notion of discrete conformational states and the events of transitions among them are fundamental.

6.2.4 Population Growth

We have discussed $\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y$ with positive λ: radioactive decay. And it does not seem that a similar discussion can be applied to $\frac{\mathrm{d}x}{\mathrm{d}t} = rx$ with a positive r, the other half of a population dynamics.

The answer turns out to be simple but profound: one should treat the birth as an event! The waiting time for the next birth is expected to be exponential. Furthermore, the rate is expected to be proportional to the number of individuals currently in the population (Exercise 1.2), say $X(t)$. Therefore, *on average the growth is 1 additional person in* $\big(r\mathbb{E}[X]\big)^{-1}$ *time*:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[X(t)\big] = r\mathbb{E}\big[X(t)\big]. \tag{6.25}$$

Death is an event, birth is an event, state transition is an event. Most biological dynamics is about counting the populations, and about biological events that lead to changing populations. Stochasticity is in the timings of the various events. This is why J. D. Murray stated in [17] that continuous growth models for a species at time t have the universal conservation equation:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \text{births} - \text{deaths} + \text{migration}, \tag{6.26}$$

where $Y(t)$ is the population density.

6.2.5 Discrete State Continuous Time Markov (Q) Processes

Discrete state continuous time Markov processes are sometime called quasi Markovian, or Q-processes, a terminology first introduced in Arne Jensen's 1954 book *A Distribution Model, Applicable to Economics* and then by David Freedman in his 1971 book *Markov Chains*. In terms of the probability of state k at time t, $p_k(t)$, one has

$$p_k(t+\mathrm{d}t) - p_k(t) = \left(\sum_{\ell=1}^{N} p_\ell(t)q_{\ell k}\right)\mathrm{d}t, \tag{6.27}$$

where $q_{\ell k}\mathrm{d}t$ is the transition probability from state ℓ to k within the infinitesimal time interval $\mathrm{d}t$. Eq. (6.27) is called a *master equation*. Its fundamental solution is $\mathbf{P}(t) = e^{\mathbf{Q}t}$, where the $\mathbf{Q}$ matrix has off-diagonal elements $q_{ij} \geq 0$ and

$$q_{ii} = -\sum_{j\neq i} q_{ij}. \tag{6.28}$$

Therefore, $\mathbf{Q}$ has each and every row sums to zero. It is often referred to as infinitesimal transition rate matrix. It is easy to show that in this case, the sum

$$\sum_{k=1}^{N} p_k(t)$$

is independent of time t. The total probability is conserved over time. Note several important differences between Eqs. (6.26) and (6.27): The former is an equation for population *density* $Y(t)$ while the latter is an equation for the probability of population *size* $p_k(t) \equiv \Pr\{N(t) = k\}$; the right-hand side of former usually is a nonlinear function of Y while the latter is necessarily linear. The dimension of the latter ODE system, however, is much higher than the former.

6.2.5.1 Kolmogorov Forward and Backward Equations

In matrix form, Eq. (6.27) can be expressed as $\frac{\mathrm{d}}{\mathrm{d}t} p = p\mathbf{Q}$, where $p = (p_1, \cdots, p_N)$ is a row vector. This equation is called *Kolmogorov forward equation*. Note strictly speaking the forward equation is not about the probability distribution (a vector), but about the transition probability matrix (fundamental solution) $\mathbf{P}(t)$ with initial value $\mathbf{P}(0) = \mathbf{I}$. More interestingly,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P} = \mathbf{PQ} = \left(e^{\mathbf{Q}t}\right)\mathbf{Q} = \mathbf{QP}. \tag{6.29}$$

This is a different differential equation:

$$\frac{\mathrm{d}u_k}{\mathrm{d}t} = \sum_{\ell=1}^{N} q_{k\ell} u_\ell, \tag{6.30}$$

which is called *Kolmogorov backward equation*. If $\{\pi_k\}$ is a stationary probability distribution, e.g., the solution to

$$\sum_{\ell=1}^{N} \pi_\ell q_{\ell k} = 0, \quad k = 1, 2, \cdots, N,$$

then the solution to the backward equation, $u_k(t)$ has the important property of

$$\sum_{k=1}^{N} u_k(t)\pi_k$$

being independent of time t, e.g., it is a conserved quantity.

The solutions to the Kolmogorov forward and backward equations also have another important property. Let $p_k(t)$ and $q_k(t)$ be two solutions to a forward equation with different initial distributions $p_k(0)$ and $q_k(0)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{N} p_k(t)\ln\left(\frac{p_k(t)}{q_k(t)}\right) \le 0. \tag{6.31}$$

One special case of this, which is widely known, is the choice of $q_k(t) = \pi_k$, if $\pi_k > 0\ \forall k$.

Similarly, two positive solutions to a Kolmogorov backward equation, $u_k(t)$ and $v_k(t)$ with different initial conditions $u_k(0)$ and $v_k(0)$, respectively, have

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{N}\Big(\pi_k u_k(t)\Big)\ln\left(\frac{u_k(t)}{v_k(t)}\right) \le 0. \tag{6.32}$$

One special case of this is when choosing $v_k(t) \equiv 1$. The quantity in Eq. (6.32) is called an H-function; the quantity in Eq. (6.31) is called relative entropy, or Kullback–Leibler divergence in information theory, or free energy in physical chemistry. These results have a deep implication for the second law of thermodynamics.

6.3 Theory of Chemical and Biochemical Reaction Systems

A general representation for complex chemical reaction systems is

$$\nu_{j1}X_1 + \nu_{j2}X_2 + \cdots \nu_{jn}X_n \xrightarrow{k_j} \kappa_{j1}X_1 + \kappa_{j2}X_2 + \cdots \kappa_{jn}X_n. \tag{6.33}$$

$1 \le j \le m$. There are n species and m reactions. $(\nu_{ji} - \kappa_{ji})$ are called *stoichiometric coefficients*, they relate a species i to the reaction j. In a broader sense, a "reaction" is just a type of "events".

6.3.1 Differential Equation and Nonlinear Dynamics

Because of the conservation of matter,

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j=1}^{m}\big(\kappa_{ji} - \nu_{ji}\big)\hat{\varphi}_j(\mathbf{x}) \tag{6.34}$$

where x_i is the concentration of chemical species X_i, $1 \le i \le n$, and

$$\hat{\varphi}_j(\mathbf{x}) = k_j x_1^{\nu_{j1}} x_2^{\nu_{j2}} \cdots x_n^{\nu_{jn}} \tag{6.35}$$

is called the instantaneous flux of the jth reaction. $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. Eq. (6.34) is called rate equations, and Eq. (6.35) is called *the law of mass action* (LMA).

6.3.2 Delbrück-Gillespie Process (DGP)

Let us now consider probabilistically the discrete, individual events of the m possible reactions in Eq. (6.33), one at a time. The DGP assumes that the jth reaction occurs following an exponentially distributed waiting time, with rate parameter

$$\varphi_j(\mathbf{X}) = k_j V \prod_{\ell=1}^{n} \left(\frac{X_\ell!}{(X_\ell - \nu_{j\ell})! V^{\nu_{j\ell}}} \right), \tag{6.36}$$

when the molecular numbers of ith chemical species being X_i. Note $\varphi_j(\mathbf{X})$ has the dimension of [time]$^{-1}$. Clearly, the first reaction that occurs also follows an exponential time, with the rate being the sum of the rates of the m reactions:

$$\sum_{j=1}^{m} \varphi_j(\mathbf{X}). \tag{6.37}$$

Among the i.i.d. $T_1, T_2, \cdots, T_n$, all exponentially distributed with respective rate parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$, the probability of the smallest one being T_k is

$$\begin{aligned} \Pr\{T^* = T_k\} &= \Pr\Big\{T_k \le \min\big(T_1, \cdots, T_{k-1}, T_{k+1}, \cdots, T_n\big)\Big\} \\ &= \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}. \end{aligned} \tag{6.38}$$

More importantly,

$$\begin{aligned} &\Pr\{T^* = T_k, T^* \ge t\} \\ &= \Pr\Big\{T_1 \ge T_k, \cdots, T_{k-1} \ge T_k, T_k \ge t, T_{k+1} \ge T_k, T_n \ge T_k, \Big\} \\ &= \int_t^\infty \lambda_k e^{-\lambda_k t_k} \prod_{\ell=1, \ell \ne k}^{n} \left(\int_{t_k}^\infty \lambda_\ell e^{-\lambda_\ell t_\ell} \mathrm{d}t_\ell \right) \end{aligned}$$

$$= \int_t^{\infty} \lambda_k e^{-\lambda_k t_k} \prod_{\ell=1,\ell\neq k}^{n} \left(\int_{t_k}^{\infty} \lambda_\ell e^{-\lambda_\ell t_\ell} \mathrm{d}t_\ell \right)$$
$$= \left(\frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n} \right) e^{-(\lambda_1+\cdots+\lambda_n)t}. \tag{6.39}$$

This means the following important fact: the minimal time among $\{T_k\}$ gives two random variables: $T^* \equiv \min_k\{T_k\}$ and $k^* \equiv \arg\min_k\{T_k\}$; the minimal time T^* and the identity k^* are statistically independent.

6.3.3 Integral Representations with Random Time Change

6.3.3.1 Poisson Process

A standard Poisson process $Y(t)$ is an integer-valued, continuous-time Markov process with distribution

$$\mathrm{Pr}\Big\{ Y(t) = k \Big\} = \frac{t^k}{k!} e^{-t}. \tag{6.40}$$

A Poisson process has both a *point process* representation, $T_1, T_2, \cdots, T_n$, and a *counting process* representation $Y(t)$. The former is a positive real-valued, discrete-time Markov process with independent increments, and $T_{i+1} - T_i$ is exponentially distributed with rate 1.

6.3.3.2 Random Time Changed Poisson Representation

In terms of Poisson processes, the stochastic trajectory of a DGP representing the integer number of the molecule X_i at time t,

$$X_i(t) = X_i(0) + \sum_{j=1}^{m} \Big(\kappa_{ji} - \nu_{ji}\Big) Y_j \left(\int_0^t \varphi_j\Big(\mathbf{X}(t)\Big)\mathrm{d}t \right) \tag{6.41}$$

in which $\varphi_j(\mathbf{X})$ is given in (6.36). We have abused the notation X_i as both the symbol of a type of molecule, as in Eq. (6.33), and its number in the reaction system.

We see that in the limit of $\mathbf{X} \to \infty$ and $V \to \infty$,

$$\varphi_j(\mathbf{X}) \to k_j V \prod_{\ell=1}^{n} \left(\frac{X_\ell}{V} \right)^{\nu_{j\ell}} = k_j V \prod_{\ell=1}^{n} x_\ell^{\nu_{j\ell}} = V \hat{\varphi}_j(\mathbf{x}). \tag{6.42}$$

$\varphi_j(\mathbf{X})$ is also called the *propensity* of the jth reaction.

6.3.4 Birth-and-Death Process with State-Dependent Transition Rates

6.3.4.1 One-Dimensional System

Consider the stochastic population kinetics of a single species. Let $p_n(t)$ be the probability of having n individuals in the population at time t. Then $p_n(t)$ satisfies the master equation

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = p_{n-1}u_{n-1} - p_n\big(u_n + w_n\big) + p_{n+1}w_{n+1}, \tag{6.43}$$

in which u_k and w_k are the birth rate and death rate of the population with exactly k individuals. The stationary distribution to Eq. (6.43) can be obtained:

$$\frac{p_n^{ss}}{p_{n-1}^{ss}} = \frac{u_{n-1}}{w_n}. \tag{6.44}$$

Therefore,

$$p_n^{ss} = p_0^{ss}\prod_{k=1}^{n}\left(\frac{u_{k-1}}{w_k}\right), \tag{6.45}$$

in which p_0^{ss} is to be determined by normalization.

Eq. (6.43) is the DGP corresponding to the nonlinear population dynamics of a single species with birth and death rates $\hat{u}(x)$ and $\hat{w}(x)$, with $x(t) \equiv \frac{X(t)}{V}$,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \hat{u}(x) - \hat{w}(x), \tag{6.46}$$

where,

$$\hat{u}(x) = \lim_{V\to\infty}\frac{u_{xV}}{V},\ \ \hat{w}(x) = \lim_{V\to\infty}\frac{w_{xV}}{V}. \tag{6.47}$$

It is easy to verify that the peaks and troughs of stationary probability distribution p_n^{ss} correspond nicely with the stable and unstable fixed points of Eq. (6.47). For the rest of this chapter, this correspondence should be kept in mind.

6.4 Using Mathematics to Articulate a Fundamental Idea in Biology

I want to use the following example to illustrate how to use mathematics, not only as a tool for computation and for modeling, but also for representing fundamental ideas.

Consider a population with many subpopulations $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, all $x_i \geq 0$. In the absence of migration, if we denote per capita growth rate $r_i = b_i - d_i$, then

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i r_i. \tag{6.48}$$

For simplicity, we shall assume that both per capita birth rate b_i and death rate d_i are constants. Then the per capita growth rate for the entire population, which is also the mean per capita growth rate,

$$\bar{r} = \frac{\displaystyle\sum_{i=1}^{n} \frac{\mathrm{d}x_i}{\mathrm{d}t}}{\displaystyle\sum_{i=1}^{n} x_i} = \frac{\displaystyle\sum_{i=1}^{n} x_i r_i}{\displaystyle\sum_{i=1}^{n} x_i}, \quad x_i \geq 0. \tag{6.49}$$

Then,

$$\frac{\mathrm{d}\bar{r}(\mathbf{x})}{\mathrm{d}t} = \left[\frac{\sum_{i=1}^{n} x_i r_i^2}{\sum_{i=1}^{n} x_i} - \left(\frac{\sum_{i=1}^{n} x_i r_i}{\sum_{i=1}^{n} x_i} \right)^2 \right]. \tag{6.50}$$

We note that the term inside $[\cdots]$ on the right-hand side is never negative:

$$\frac{\sum_{i=1}^{n} x_i r_i^2}{\sum_{i=1}^{n} x_i} - \left(\frac{\sum_{i=1}^{n} x_i r_i}{\sum_{i=1}^{n} x_i} \right)^2 = \frac{\sum_{i=1}^{n} x_i \left(r_i - \bar{r} \right)^2}{\sum_{i=1}^{n} x_i} \geq 0. \tag{6.51}$$

In fact, it is exactly the variance of r_i among the different subpopulations. Therefore, it is always positive if there are variations among r_i. This mathematical result is a part of the ideas of both Adam Smith, on economics, and Charles Darwin, on the natural selection. In fact, the term $[\cdots]$ in Eq. (6.50) has been identified by R. A. Fisher, the British statistician and evolutionary biologist, as the "growth of fitness due to natural selection" [6]. Here is a quote from Smith's *magnum opus* "An Inquiry into the Nature and Causes of the Wealth of Nations" (1776):

> As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other eases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good. It is an affectation, indeed, not very common among merchants, and very few words need be employed in dissuading them from it.

6.5 Ecological Dynamics and Nonlinear Chemical Reactions: Two Examples

6.5.1 Predator and Prey System

Let $z(t)$ be the population density of a predator at time t and $x(t)$ be the population density of a prey at the same time. Then the simplest predator-prey dynamics is [17]

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - \beta x z, \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = -\gamma z + \delta x z. \end{cases} \tag{6.52}$$

The detailed analysis of the nonlinear dynamics can be found in many textbooks on mathematical biology or differential equations [17].

Let us now consider the following chemical reaction system:

$$A + X \xrightarrow{k_1} 2X, \quad X + Y \xrightarrow{k_2} 2Y, \quad Y \xrightarrow{k_3} B. \tag{6.53}$$

Then according to the LMA, the concentrations of X and Y, with fixed concentrations of A and B being a and b:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_1 a x - k_2 x y, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -k_3 y + k_2 x y. \tag{6.54}$$

Therefore, we see that dynamics of an ecological predator-prey system is remarkable similar to that of a chemical reaction system with autocatalysis [16]: the first reaction in (6.53) requires an existing X serving as a catalyst for the reaction $A \rightarrow X$. A species that appears on the both sides of a chemical reaction is called a *catalyst*.

6.5.2 A Competition Model

Let us now consider another widely studied ecological dynamics with competition [17]:

$$\begin{cases} \dfrac{\mathrm{d}N_1}{\mathrm{d}t} = r_1 N_1 - a_1 N_1^2 - b_{21} N_1 N_2, \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}t} = r_2 N_2 - a_2 N_2^2 - b_{12} N_2 N_1. \end{cases} \tag{6.55}$$

Can one "design" a system of chemical reactions that yields an identical system of differential equation? Without loss of generality, let us assume that $b_{12} > b_{21}$.

$$\begin{aligned} &A + X \xrightarrow{k_1} 2X, \quad X + X \xrightarrow{k_2} B, \quad A + Y \xrightarrow{k_3} 2Y, \\ &Y + Y \xrightarrow{k_4} B, \quad X + Y \xrightarrow{k_5} B, \quad X + Y \xrightarrow{k_6} X + B, \end{aligned} \tag{6.56}$$

which, according to the LMA,

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = (k_1 a)x - k_2 x^2 - k_5 xy, \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = (k_3 a)y - k_4 y^2 - (k_5 + k_6)xy. \end{cases} \tag{6.57}$$

If we identify x, y with N_1, N_2, and

$$(k_1 a) \leftrightarrow r_1, \ k_2 \leftrightarrow a_1, \ k_5 \leftrightarrow b_{21}, \ (k_3 a) \leftrightarrow r_2, \ k_4 \leftrightarrow a_2, \ (k_5 + k_6) \leftrightarrow b_{12},$$

then (6.57) is the same as (6.55). Note that the last reaction, $X + Y \to X + B$, is introduced to represent $b_{12} > b_{21}$.

A close inspection of the system of chemical reactions in (6.56) indicates that the overall reaction is $2A \to B$. Since each and every reaction is irreversible, there can be no chemical equilibrium. Rather, the system eventually reaches a *nonequilibrium steady state* in which there is a continuous, overall chemical flux converting $2A$ to B.

6.5.3 Logistic Model and Keizer's Paradox

We now turn to studying some issues more in-depth. Let us now consider a much simpler chemical reaction system,

$$A + X \xrightarrow{k_1} 2X, \quad X + X \xrightarrow{k_2} B. \tag{6.58}$$

It is easy to see that the ODE according to the LMA,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = r\left(1 - \frac{x}{K}\right)x, \quad r = k_1 a, \ K = \frac{r}{k_2}, \tag{6.59}$$

is the celebrated *logistic equation* in population dynamics. In the ecological context, r is known as the per capita growth rate in the absence of intra-species competition; and K is known as *carrying capacity*.

The DGP stochastic model has a *chemical master equation* (CME) for the probability of n X molecules in a reaction volume of V:

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = u_{n-1}p_{n-1} - \Big(u_n + w_n\Big)p_n + w_{n+1}p_{n+1}, \tag{6.60a}$$

in which the state-dependent birth and death rates are

$$u_n = rn, \quad w_n = \frac{k_2 n(n-1)}{V}. \tag{6.60b}$$

Then, according to Eq. (6.45),

$$p_0^{eq} = 1 \ \text{ and } \ p_n^{eq} = 0, \ n \geq 1, \tag{6.61}$$

since $u_0 = 0$! In other words, according to this theory, the stationary probability distribution is “population extinction with probability 1”. But the ODE in (6.59) says that the stable steady state is $x = K$, with $x = 0$ being a unstable steady state which is not “relevant”.

This seeming disagreement between the deterministic ODE in (6.59) and stochastic dynamics described by (6.60) is known as *Keizer’s paradox*. We refer the readers to [33] for the resolution of the paradox.

6.6 Chemical Thermodynamics and Entropy Production

6.6.1 Classical Chemical Thermodynamics

A single *reversible* chemical reaction

$$\nu_1 X_1 + \nu_2 X_2 + \cdots \nu_n X_n \underset{k^-}{\overset{k^+}{\rightleftharpoons}} \kappa_1 X_1 + \kappa_2 X_2 + \cdots \kappa_n X_n \tag{6.62}$$

is said to be in a chemical equilibrium when

$$\frac{\hat{\varphi}_k^+(\mathbf{x}^{eq})}{\hat{\varphi}_k^-(\mathbf{x}^{eq})} = 1 \ \Leftrightarrow \ \left(\frac{x_1^{\nu_1}x_2^{\nu_2}\cdots x_n^{\nu_n}}{x_1^{\kappa_1}x_2^{\kappa_2}\cdots x_n^{\kappa_n}}\right)^{eq} = \frac{k^-}{k^+}. \tag{6.63}$$

(k^-/k^+) is known as the *equilibrium constant* of the reaction. The ratio on the lhs is a constant independent of the total amount participating species.

Chemical thermodynamics introduces the notions of chemical energy and chemical potential: for ideal solutions chemical species i has a chemical potential

$$\mu_i = \mu_i^o + k_B T \ln x_i. \tag{6.64}$$

in which μ_i^o is determined by the atomic structure of a molecule, e.g., internal energy. k_B is Boltzmann's constant, and T is temperature in Kelvin. Then the Gibbs free energy of the lhs of (6.62) is the sum of the chemical potential

$$G = \sum_{i=1}^{n} \nu_i \Big(\mu_i^o + k_B T \ln x_i \Big). \tag{6.65}$$

When the reaction reaches its equilibrium, one has the total chemical potentials being equal on both sides:

$$\sum_{i=1}^{n} \big(\nu_i - \kappa_i\big) \Big(\mu_i^o + k_B T \ln x_i^{eq} \Big) = 0. \tag{6.66}$$

This implies

$$\prod_{i=1}^{n} \big(x_i^{eq}\big)^{\nu_i - \kappa_i} = e^{-\frac{(\nu_i - \kappa_i)\mu_i^o}{k_B T}} = \frac{k^-}{k^+}, \tag{6.67}$$

or

$$\Delta G^o = \left(\sum_{i=1}^{n} \kappa_i \mu_i^o \right) - \left(\sum_{i=1}^{n} \nu_i \mu_i^o \right) = k_B T \ln \left(\frac{k^-}{k^+} \right). \tag{6.68}$$

This is a very well-known formula that can be found in every college chemistry textbook.

6.6.2 Mass-Action Kinetics

Following Eqs. (6.34) and (6.35), we have

$$\begin{aligned}
\frac{\mathrm{d}x_i}{\mathrm{d}t} &= \sum_{j=1}^{m} \big(\kappa_{ji} - \nu_{ji}\big) \Big(\hat{\varphi}_j^+ - \hat{\varphi}_j^- \Big) \\
&= \sum_{j=1}^{m} \big(\kappa_{ji} - \nu_{ji}\big) \hat{\varphi}_j^- \left\{ \exp\left[\sum_{\ell=1}^{n} \big(\kappa_{j\ell} - \nu_{j\ell}\big) \ln \left(\frac{x_\ell}{x_\ell^{eq}} \right) \right] - 1 \right\} \\
&= \sum_{j=1}^{m} \big(\kappa_{ji} - \nu_{ji}\big) \hat{\varphi}_j^+ \left\{ 1 - \exp\left[\sum_{\ell=1}^{n} \big(\nu_{ji} - \kappa_{ji}\big) \ln \left(\frac{x_\ell}{x_\ell^{eq}} \right) \right] \right\}.
\end{aligned} \tag{6.69}$$

Equation (6.69) shows that when $x_\ell = x_\ell^{eq}$, the term $[\cdots] = 0$ and the term $\{\cdots\} = 0$ as well, for every j. Therefore, the kinetic equation in (6.69) is consistent with the chemical equilibrium according to thermodynamics, e.g., Eqs. (6.66) and (6.67). Interestingly, recent work has shown that both macroscopic kinetics as in (6.69) and equilibrium thermodynamics in Sect. 6.6.1 are consequences of a stochastic kinetic description of a reaction system [10].

6.6.3 Stochastic Chemical Kinetics

We now apply the above formalism to a nonlinear chemical reaction in a small volume V with small number of molecules, n_A, n_B, and n_C numbers of A, B, and C:

$$A + B \underset{k^-}{\overset{k^+}{\rightleftharpoons}} C. \tag{6.70}$$

We note that the $n_A + n_C$ and $n_B + n_C$ do not change in the reaction. Hence we can denote $n_A + n_C = n_A^o$ and $n_B + n_C = n_B^o$ as the total amount of A and B, including those in C, at the initial time. Now if we use n_C as the non-negative integer-valued random variable to describe the stochastic chemical kinetics, this simple nonlinear chemical reaction, according to DGP, is a one-dimensional birth-and-death process, with state-dependent birth and death rates $u_n = k^+ n_A n_B$ and $w_n = k_- n_C$. Then, according to Eq. (6.45), we have an equilibrium distribution $p^{eq}(m) = \Pr\{n_C^{eq} = m\}$:

$$\frac{p^{eq}(m+1)}{p^{eq}(m)} = \frac{k^+(n_A^o - m)(n_B^o - m)}{k^-(m+1)V}, \tag{6.71}$$

in which $n_A^o = n_A(0) + n_C(0)$ and $n_B^o = n_B(0) + n_C(0)$. Therefore,

$$p^{eq}(m) = \frac{\Xi^{-1}\, n_A^o! n_B^o!}{m!(n_A^o - m)!(n_B^o - m)!} \left(\frac{k^+}{k^- V}\right)^m, \tag{6.72}$$

where Ξ is a normalization factor

$$\Xi(\lambda) = \sum_{m=0}^{\min(n_A^o, n_B^o)} \frac{n_A^o!\, n_B^o!\, \lambda^m}{m!(n_A^o - m)!(n_B^o - m)!}, \quad \lambda = \left(\frac{k^+}{k^- V}\right). \tag{6.73}$$

More importantly, by noting $n_A + n_B + n_C = n_A^0 + n_B^0 - n_C$,

$$\begin{aligned} & -\ln p^{eq}(n_C) \\ & = -\ln\left[\frac{\lambda^{n_C}}{n_C!(n_A^o - n_C)!(n_B^o - n_C)!}\right] + \text{const.} \end{aligned}$$

$$
\begin{aligned}
&= n_A \ln\left(\frac{n_A}{V}\right) - n_A + n_B \ln\left(\frac{n_B}{V}\right) - n_B + n_C \ln\left(\frac{n_C}{V}\right) - n_C - n_C \ln\left(\frac{k^+}{k^-}\right) \\
&= n_A \ln x_A + n_B \ln x_B + n_C \ln x_C + n_C\left(\frac{\mu_C^o - \mu_A^o - \mu_B^0}{k_B T}\right) - (n_A + n_B + n_C) \\
&= \sum_{\sigma=A,B,C} n_\sigma\left(\frac{\mu_\sigma^o}{k_B T} + \ln x_\sigma - 1\right). \qquad (6.74)
\end{aligned}
$$

This agrees with Eq. (6.65).

In classical chemical kinetics, for a given $\mathbf{x}(t)$, the Ideal function of the chemical reaction system is

$$
G^{eq}[\mathbf{x}(t)] = \sum_{\sigma=1}^{n} x_\sigma\Big(\mu_\sigma^o + k_B T \ln x_\sigma - k_B T\Big). \qquad (6.75)
$$

Then, following Eq. (6.34), assuming each and every reaction is reversible with rate constants k_j^+ and k_j^-,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} G^{eq}[\mathbf{x}(t)] &= \sum_{i=1}^{n} \frac{\mathrm{d}x_i}{\mathrm{d}t}\Big(\mu_i^o + k_B T \ln x_i\Big) \\
&= k_B T \sum_{i=1}^{n}\sum_{j=1}^{m} \ln\left(\frac{x_i}{x_i^{eq}}\right)\Big(\kappa_{ji} - \nu_{ji}\Big)\left(k_j^+ \prod_{\ell=1}^{n} x_\ell^{\nu_{j\ell}} - k_j^- \prod_{\ell=1}^{n} x_\ell^{\kappa_{j\ell}}\right) \\
&= -k_B T \sum_{j=1}^{m}\left\{\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_i^{eq}}\right)^{\nu_{ji}-\kappa_{ji}}\right\}\Big(\hat{\varphi}_j^+ - \hat{\varphi}_j^-\Big) \\
&= -k_B T \sum_{j=1}^{m}\Big(\hat{\varphi}_j^+ - \hat{\varphi}_j^-\Big)\ln\left(\frac{\hat{\varphi}_j^+}{\hat{\varphi}_j^-}\right) \qquad (6.76) \\
&\le 0. \qquad (6.77)
\end{aligned}
$$

The minus-log stationary probability distribution is a Lyapunov function for the dynamics. The rhs of Eq. (6.76) is known as *entropy production rate*.

6.6.4 Nonequilibrium Steady-State and Driven Chemical Systems

If a chemical reaction system reaches its chemical equilibrium, then each and every reaction in the system is in *detailed balance* with zero net flux. This puts a very strong condition on the dynamics. When a chemical reaction system has a sustained source and sink with different chemical potentials, it cannot reach a chemical equilibrium. Rather, it reaches a *nonequilibrium steady state* (NESS).

Let us consider the following two examples, the Schlögl model for bistability [34] and Schnakenberg model for nonlinear oscillation [17, 25, 35].

6.6.4.1 Schlögl Model

$$A+2X \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} 3X, \quad X \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} B, \tag{6.78}$$

in which the concentrations (or chemical potentials) of A and B are sustained by an external agent. This reaction is known as *Schlögl model*, whose dynamics can be described by the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_1^+ax^2 - k_1^-x^3 - k_2^+x + k_2^-b = f(x), \tag{6.79}$$

which is a third-order polynomial. It can exhibit bistability and saddle-node bifurcation phenomenon. All of them only occur under driven condition, when $\mu_A \neq \mu_B$. Note in the chemical equilibrium: $\mu_A = \mu_A^o + k_BT\ln a = \mu_B^o + k_BT\ln b$, and

$$\left(\frac{b}{a}\right)^{eq} = \frac{k_1^+k_2^+}{k_1^-k_2^-}. \tag{6.80}$$

Differential equation (6.79), with its parameters $ak_1^+k_2^+ = bk_1^-k_2^-$, has the right-hand-side

$$\begin{aligned} f(x) &= k_1^+ax^2 - k_1^-x^3 - k_2^+x + k_2^-b \\ &= k_1^+ax^2 - k_1^-x^3 - k_2^+x + \frac{ak_1^+k_2^+}{k_1^-} \\ &= \left(x^2 + \frac{k_2^+}{k_1^-}\right)\left(ak_1^+ - k_1^-x\right). \end{aligned} \tag{6.81}$$

Therefore, the $f(x)$ has a unique fixed point at $x = \frac{ak_1^+}{k_1^-}$, the chemical equilibrium. In general, system (6.78) can exhibit chemical bistability; but this is only possible when A and B have a sufficiently large chemical potential difference, e.g., *a chemostat*.

More interestingly, when a and b satisfying (6.80), the DGP of the number of X, $n_X(t)$, is again a one-dimensional birth-and-death process, with

$$u_n = \frac{k_1^+ an(n-1)}{V} + k_2^- bV \;=\; \frac{k_1^+ a}{V}\left(n(n-1) + \frac{k_2^+ V^2}{k_1^-}\right), \tag{6.82}$$

$$\begin{aligned} w_{n+1} &= \frac{k_1^-(n+1)n(n-1)}{V^2} + k_2^+(n+1) \\ &= \frac{k_1^-(n+1)}{V^2}\left(n(n-1) + \frac{k_2^+ V^2}{k_1^-}\right). \end{aligned}$$

Therefore, the stationary distribution, according to Eq. (6.45),

$$p_n^{eq} = C\prod_{\ell=0}^{n-1} \frac{k_1^+ a/V}{k_1^-(\ell+1)/V^2} = \frac{\lambda^n}{n!}e^{-\lambda}, \;\; \lambda = \left(\frac{k_1^+ aV}{k_1^-}\right). \tag{6.83}$$

This is a Poisson distribution, with expected value being $\mathbb{E}\big[n_X^{eq}\big] = \lambda$. Therefore, the expected concentration is $(k_1^+ a/k_1^-)$.

6.6.4.2 Schnakenberg Model

Similarly,

$$A \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} X, \;\; B \xrightarrow{k_2} Y, \;\; 2X + Y \xrightarrow{k_3} 3X, \tag{6.84}$$

is known as *Schnakenberg model*, whose dynamics follow

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = k_1^+ a - k_1^- x - k_3 x^2 y \;=\; f(x,y), \\[2ex] \dfrac{\mathrm{d}z}{\mathrm{d}t} = k_2 b - k_3 x^2 y \;=\; g(x,y). \end{cases} \tag{6.85}$$

This system can exhibit limit cycle oscillation and Hopf bifurcation. In terms of the DGP, it exhibits a rotational diffusion. We refer the readers to [25, 35] for an in-depth analysis of the model.

6.7 The Law of Large Numbers—Kurtz's Theorem

6.7.1 Diffusion Approximation and Kramers–Moyal Expansion

Starting with the master equation in (6.43), let us consider a partial differential equation (PDE) for a continuous density function $f(x,t)\mathrm{d}x = p_{Vx}(t)$ where $x = \frac{n}{V}$, $\mathrm{d}x = \frac{1}{V}$, then

$$
\begin{aligned}
\frac{\partial f(x,t)}{\partial t} &= V\frac{\mathrm{d}p_{Vx}(t)}{\mathrm{d}t}\\
&= \frac{1}{\mathrm{d}x}\Big(f(x-\mathrm{d}x,t)\hat{u}(x-\mathrm{d}x) - f(x,t)\Big(\hat{u}(x)+\hat{w}(x)\Big)\\
&\quad +f(x+\mathrm{d}x,t)\hat{w}(x+\mathrm{d}x)\Big)\\
&= \frac{\partial}{\partial x}\Big(f(x+\mathrm{d}x/2,t)\hat{w}(x+\mathrm{d}x/2) - f(x-\mathrm{d}x/2,t)\hat{u}(x-\mathrm{d}x/2)\Big)\\
&\approx \frac{\partial}{\partial x}\left\{\frac{\partial}{\partial x}\left(\frac{\hat{w}(x)+\hat{u}(x)}{2V}\right)f(x,t) - \Big(\hat{u}(x)-\hat{w}(x)\Big)f(x,t)\right\} + \cdots
\end{aligned}
\tag{6.86}
$$

in which

$$V^{-1}u_{Vx} = \hat{u}(x), \quad V^{-1}w_{Vx} = \hat{w}(x), \tag{6.87}$$

as $V \to \infty$.

6.7.2 Nonlinear Differential Equation, Law of Mass Action

Therefore, in the limit of $V \to \infty$,

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}\Big(\hat{u}(x) - \hat{w}(x)\Big)f(x,t), \tag{6.88}$$

which corresponds to the ordinary differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \hat{u}(x) - \hat{w}(x), \tag{6.89}$$

that defines the characteristic lines of (6.88).

6.7.3 Central Limit Theorem, a Time-Inhomogeneous Gaussian Process

Now consider the process

$$Y(t) = \frac{X(t) - Vx(t)}{\sqrt{V}}, \tag{6.90}$$

which characterizes the "deviation" of $\frac{X(t)}{V}$ from $x(t)$. In the limit of $V \to \infty$, its pdf $f_Y(y,t)$ satisfies a linear PDE with time-varying coefficients

$$\frac{\partial f_Y(y,t)}{\partial t} = \frac{\partial}{\partial y}\left\{\frac{\partial}{\partial y}\left(\frac{\hat{w}(x(t)) + \hat{u}(x(t))}{2}\right) f_Y(y,t) \right.$$
$$\left. -\Big(\hat{u}'(x(t)) - \hat{w}'(x(t))\Big) y f_Y(y,t)\right\}. \tag{6.91}$$

Therefore, $Y(t)$ is a continuous time, real-valued, time-inhomogeneous Markov process. Note the PDE (6.91) is very different from PDE (6.86). They are known in physics literature as the Kramers–Moyal expansion and van Kampen's Ω-expansion, respectively [32]. The former is not related to the central limit theorem.

6.7.4 Diffusion's Dilemma

Truncating the Eq. (6.86) after the second order, it has a stationary distribution

$$-\ln \hat{f}_Y^{st}(y) = 2V \int \left(\frac{\hat{w}(x) - \hat{u}(x)}{\hat{u}(x) + \hat{w}(x)}\right) \mathrm{d}x. \tag{6.92}$$

On the other hand, the stationary solution given in (6.45),

$$p_n^{ss} = p_0^{ss} \prod_{k=1}^{n} \left(\frac{u_{k-1}}{w_k}\right),$$

in the limit of $V \to \infty$ with $V^{-1}u_{Vx} = \hat{u}(x)$, $V^{-1}w_{Vx} = \hat{w}(x)$, and $V^{-1} = \mathrm{d}x$, yields

$$-\ln p_{Vx}^{ss} = -\sum_{k=1}^{n} \ln\left(\frac{u_{k-1}}{w_k}\right) + C \leftrightarrow -\ln f^{ss}(x) = V \int \ln\left(\frac{\hat{w}(x)}{\hat{u}(x)}\right) \mathrm{d}x. \tag{6.93}$$

Is it possible Eqs. (6.92) and (6.93) are actually the same? We notice that both have identical local extrema:

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(-\ln f_Y^{st}(x)\Big) = 2V\left(\frac{\hat{w}(x)-\hat{u}(x)}{\hat{w}(x)+\hat{u}(x)}\right) = 0 \Longrightarrow \hat{w}(x) = \hat{u}(x). \tag{6.94}$$

In fact, the curvature at a local extremum is identical:

$$\begin{aligned}\left[\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(-\ln f_Y^{st}(x)\Big)\right]_{\hat{u}=\hat{w}} &= 2V\left(\frac{\hat{w}'(x)-\hat{u}'(x)}{\hat{w}(x)+\hat{u}(x)}\right) = V\left(\frac{\hat{w}'(x)-\hat{u}'(x)}{\hat{u}(x)}\right)\\ &= \left[\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(-\ln f^{ss}s(x)\Big)\right]_{\hat{u}=\hat{w}}.\end{aligned} \tag{6.95}$$

However, it can be shown, via an example, that the global minimum can be different [20, 37]! This implies that Kramers–Moyal expansion is not a valid approximation for stochastic kinetics with multiple stability. Continuous time, real-valued Markov processes are also called *diffusion processes*. The above result illustrates that there is no globally valid diffusion approximation for stochastic population kinetics in general.

6.8 The Logic of the Mechanical Theory of Heat and Nonequilibrium Thermodynamics

In order to present some rather recent results in Sect. 6.9 and put those results into a proper context, let us first revisit the celebrated work of L. Boltzmann on the *mechanical theory of heat* [8], and the generally accepted macroscopic nonequilibrium thermodynamics presented in the classic treatise of de Groot and Mazur [5]. The readers will recognize the logical threads of both theories in Sect. 6.9, as well as the finding of a missing link between the above two theories.

Boltzmann's theory is based on the general Hamiltonian dynamics and starts with a definition of an entropy function $S = -k_B \ln \Omega(E)$. Section 6.9 will be based on the general Markov dynamics and starts with a definition of an entropy function according to Shannon [29]. Note that Boltzmann's entropy is a static quantity, the entropy in Sect. 6.9, Eq. (6.110) below, is a function of time.

De Groot and Mazur's theory is based on continuity equations for mass and energy, relating time changes of the density of these quantities to transport processes in three-dimensional space, and identifies entropy productions as "transport flux × driving force", *à la* Onsager [18]. Section 6.9 is based on a continuity equation for the probability in the state space, relating time change of probability to its transport, and also identifies the entropy production as "probability flux × chemical potential difference".

There is a missing link between Boltzmann's theory and the nonequilibrium thermodynamics. In addition to the continuity equations, the de Groot-Mazur approach also requires the *entropy balance equation* [5],

$$\frac{\mathrm{d}S}{\mathrm{d}t} = e_p + J_S, \tag{6.96}$$

as one of its fundamental premises, where e_p is the entropy production rate and J_S is the rate of entropy supplied to a system by its surroundings. The second law of thermodynamics, e.g., Clausius inequality, dictates that $e_p \geq 0$. Unfortunately, Boltzmann's mechanical theory of heat is not able to derive an equation like (6.96) from a Hamiltonian dynamics without resorting to additional assumptions based on a *stosszahlansatz*.[1] As one will see from Sect. 6.9, however, Markov dynamics is able to provide nicely an equation like (6.96). A stochastic dynamic approach to nonequilibrium thermodynamics is able to fill this logic gap, as was first demonstrated by Bergmann and Lebowitz in 1955 [2].

6.8.1 Boltzmann's Mechanical Theory of Heat

The entire world, as long as one is interested in phenomena that are at not too small a scale (e.g., quantum) and not too close to the speed of light (e.g., relativity), follows the Newtonian mechanics which can be represented mathematically in terms of a Hamiltonian system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\partial H(x,y)}{\partial y}, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{\partial H(x,y)}{\partial x}. \tag{6.97}$$

One of the most important result concerning the Eq. (6.97) is the dynamics invariance of $H(x(t), y(t))$:

$$\frac{\mathrm{d}}{\mathrm{d}t} H\big(x(t), y(t)\big) = \frac{\partial H}{\partial x}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right) + \frac{\partial H}{\partial y}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right) = 0. \tag{6.98}$$

[1] In the phase space, the Hamiltonian system has a Liouville equation

$$\frac{\partial u(x,y,t)}{\partial t} = -\left(\frac{\partial H}{\partial y}\right)\frac{\partial u}{\partial x} + \left(\frac{\partial H}{\partial x}\right)\frac{\partial u}{\partial y}.$$

It is easy to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\iint u(x,y,t)\ln u(x,y,t)\mathrm{d}x\mathrm{d}y = 0.$$

Therefore, the information-entropy like quantity is time invariant under a deterministic diffeomorphism [36].

Now, let us assume that the Hamiltonian function contains also several parameters, $H(x, y, V, N)$ where V is the box size of a mechanical system and N is the number of particles in the box, then the next question which an applied mathematician might ask, but interestingly which has not been extensively studied, is this: "What is the long-time behavior of the system as a function of V, N, and other parameters?"

A Hamiltonian system, however, is fundamentally different from the earlier systems we have studied, which have attractive fixed point(s). In fact, it is clear that the long-time behavior is a function of the initial condition $H\big(x(0), y(0)\big) = E$. Helmholtz and Boltzmann (1884) realized that a "thermodynamic equilibrium state" of a mechanical system is *not a single point in the phase space, but rather, it is an entire invariant manifold* defined by the level set $H(x, y, V, N) = E$. It was Boltzmann's ingenuity to realize that one can define

$$\begin{aligned} S\big(E, V, N\big) &= k_B \ln \big\{\text{phase volume contained by the surface } H(x, y) = E\big\} \\ &= k_B \ln \int_{H(x,y)\le E} \mathrm{d}x\mathrm{d}y. \end{aligned} \tag{6.99}$$

Since $S(E)$ is monotonic, one has an implicit function $E = E(S, V)$. Then

$$\begin{aligned} \mathrm{d}E &= \left(\frac{\partial E}{\partial S}\right)_{V,N} \mathrm{d}S + \left(\frac{\partial E}{\partial V}\right)_{S,N} \mathrm{d}V + \left(\frac{\partial E}{\partial N}\right)_{S,V} \mathrm{d}N \\ &= T\mathrm{d}S - p\mathrm{d}V + \mu\mathrm{d}N. \end{aligned} \tag{6.100}$$

What is the significance of Eq. (6.100)? First, it is completely based on the fact that a Hamiltonian system has a *conservation of mechanical energy H*. Furthermore, however, this conservation of energy is valid not only for a single Hamiltonian system on a single invariant torus, but also the Hamiltonian system with multiple level sets, and even among an entire class of Hamiltonian systems with varying V and N, and other parameters. It becomes a universally valid equation, known as *the First Law of Thermodynamics*. Note, according to this theory, thermodynamic quantities like T, p, μ are mathematically defined via Eq. (6.100). They are emergent phenomena.

T and p have mechanical interpretations, though not perfect, as mean kinetic energy and mean momentum transfer to a wall. μ, however, has no interpretation in terms of classical motion; rather, it has an interpretation in terms of Brownian motion:

$$\frac{\partial \rho(x, t)}{\partial t} = D\frac{\partial^2 \rho(x, t)}{\partial x^2} = -\frac{1}{\eta}\frac{\partial(\hat{F}\rho)}{\partial x}, \tag{6.101}$$

where

$$\hat{F} = -\frac{\partial \mu}{\partial x}, \;\; \text{and} \;\; \mu = D\eta \ln \rho(x, t) = k_B T \ln \rho(x, t). \tag{6.102}$$

$\hat{F}$ is known as *entropic force* in chemistry, and μ is known as chemical potential.

6.8.2 *Classical Macroscopic Nonequilibrium Thermodynamics*

Equation (6.100) is valid only when the entire torus $H(x, y) = E$ is visited in the long time limit; this is known as *ergodicity*. In other words, with time t in mind, the equation is valid only when the $\mathrm{d}S$ and $\mathrm{d}V$ are very slowly changing. What happens if the changes are not slow? Then, the *Second Law of Thermodynamics* states that

$$T\mathrm{d}S \geq \text{đ}Q = \mathrm{d}E - \text{đ}W, \tag{6.103}$$

in which $\text{đ}Q$ is the amount of heat that flows into the system, and $\text{đ}W$ is the amount of work done to the system. Both are path dependent, as indicated by the đ. Eq. (6.103) is known as the Clausius inequality. The notion of *entropy production* is introduced to account for the inequality:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = e_p - \frac{h_d}{T}, \quad e_p \geq 0, \tag{6.104}$$

in which e_p is called entropy production, which is never negative. $h_d = -\text{đ}Q/\mathrm{d}t$ is called heat dissipation. In general, neither e_p nor h_d is a time derivative. Eq. (6.104) is known as an *entropy balance equation.*

6.8.2.1 Local Equilibrium Assumption and Classical Derivation of Entropy Production

If one assumes that Eq. (6.100) is valid locally in space and time, then one has

$$\frac{\partial s(x,t)}{\partial t} = \frac{1}{T}\frac{\partial u(x,t)}{\partial t} - \sum_{i=1}^{n} \mu_i \frac{\partial c_i(x,t)}{\partial t}, \tag{6.105}$$

in which we have assumed imcompressibility $\mathrm{d}V = 0$. $s(x, t)$, $u(x, t)$, and $c_i(x, t)$ are entropy density, energy density, and concentration of the ith species.

Realizing that both energy and particles follow continuity equation in space-time, one has

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\partial J_u(x,t)}{\partial x}, \quad \frac{\partial c_i(x,t)}{\partial t} = -\frac{\partial J_i(x,t)}{\partial x}. \tag{6.106}$$

Then, substituting these into Eq. (6.105), and use a certain amount of physical intuition, one arrives at

$$\frac{\partial s(x,t)}{\partial t} = e_p(x,t) + J_S(x,t) \tag{6.107a}$$

where entropy production rate per unit volume

$$e_p(x,t) = J_u \frac{\partial}{\partial x}\left(\frac{1}{T}\right) - \sum_{i=1}^{n} J_i \frac{\partial}{\partial x}\left(\frac{\mu_i}{T}\right) - \sum_{j=1}^{m} \frac{\Delta\mu_j \hat{\varphi}_j}{T}, \tag{6.107b}$$

and entropy flux

$$J_S(x,t) = \frac{\partial}{\partial x}\left(\frac{J_u}{T} - \sum_{i=1}^{n} \frac{\mu_j J_j}{T}\right). \tag{6.107c}$$

According to Onsager's theory [18], each term in the entropy production e_p is a

$$\text{transport flux} \times \text{driving force} \tag{6.108}$$

which should be non-negative. The theory of nonequilibrium thermodynamics concerns with transport processes of various kinds: diffusion, heat, charge, chemical, etc. More information on the various transport fluxes can only be obtained, phenomenologically, from engineering.

6.9 Mathematicothermodynamics of Markov Dynamics

We now consider discrete-state Markov system with stochastic dynamics in terms of "continuity equation for probability in state space", e.g., Chapman–Kolmogorov equation, or master equation

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N} \Big(p_j q_{ji} - p_i q_{ij}\Big), \tag{6.109}$$

in which q_{ij} are the infinitesimal transition probability rate given in (6.27).

We shall now follow the same logic steps of Boltzmann, illustrated in Sect. 6.8.1, to develop a "thermodynamic theory" based on the general dynamics by introducing the notion of entropy. Eq. (6.109) replaces the Hamiltonian system (6.97), and in the place of Boltzmann's celebrated $S = k_B \ln \Omega(E)$ will be the Gibbs-Shannon entropy:

$$S(t) = -\sum_{i=1}^{N} p_i(t) \ln p_i(t). \tag{6.110}$$

Then, one has

$$\frac{\mathrm{d}S}{\mathrm{d}t} = e_p + J_S, \tag{6.111a}$$

where

$$e_p(t) = \frac{1}{2}\sum_{i,j=1}^{N}\Big(p_i(t)q_{ij} - p_j(t)q_{ji}\Big)\ln\left(\frac{p_i(t)q_{ij}}{p_j(t)q_{ji}}\right), \tag{6.111b}$$

$$J_S(t) = \frac{1}{2}\sum_{i,j=1}^{N}\Big(p_i(t)q_{ij} - p_j(t)q_{ji}\Big)\ln\left(\frac{q_{ji}}{q_{ij}}\right). \tag{6.111c}$$

It is immediately obvious that $e_p \geq 0$ since for every pair of i, j in Eq. (6.111b), the term has the form of $(a-b)\ln(a/b) \geq 0$. We also note the resemblance of (6.111b) to Eq. (6.76).

Therefore, we have derived an entropy balance equation based on Markov dynamics, without the assumption of local equilibrium. Equations (6.111b) and (6.111c) further give explicit expressions, in terms of the $\{p_i(t)\}$, for the entropy flux J_S the non-negative entropy production e_p. As we shall show below, there is a complete nonequilibrium thermodynamics on the mesoscopic scale, in state space. This theory is effectively an isothermal theory with the "temperature" being 1.

6.9.1 Non-Decreasing Entropy in Systems with Uniform Stationary Distribution

If the master Eq. (6.109) has a stationary distribution $p_n^{ss} = 1\ \forall n$, then

$$\sum_{j=1}^{N}\Big(q_{ji} - q_{ij}\Big) = \sum_{j=1}^{N} q_{ji} = 0, \ \ \forall i.$$

In this case,

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= -\sum_{i=1}^{N}\left(\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}\right)\ln p_i \ = \ -\sum_{i,j=1}^{N}\Big(p_jq_{ji} - p_iq_{ij}\Big)\ln p_i \\ &= \sum_{i,j=1}^{N} p_iq_{ij}\ln\left(\frac{p_i}{p_j}\right) \ \geq \ \sum_{i,j=1}^{N} p_iq_{ij}\left(\frac{p_j}{p_i} - 1\right) \\ &= \sum_{j=1}^{N} p_j\left(\sum_{i=1}^{N} q_{ij}\right) \ = \ 0.\end{aligned} \tag{6.112}$$

We therefore have a "theorem" stating that if the stationary probability distribution is uniform, then the entropy S is non-decreasing function of time.

6.9.2 Q-Processes with Detailed Balance

If a Q process has a stationary distribution $p_i^{ss} q_{ij} = p_j^{ss} q_{ji}$, known as *detailed balance*, then

$$
\begin{aligned}
J_S(t) &= \frac{1}{2} \sum_{i,j=1}^{N} \Big(p_i(t) q_{ij} - p_j(t) q_{ji} \Big) \ln \left(\frac{q_{ji}}{q_{ij}} \right) \\
&= \frac{1}{2} \sum_{i,j=1}^{N} \Big(p_i(t) q_{ij} - p_j(t) q_{ji} \Big) \ln \left(\frac{p_i^{ss}}{p_j^{ss}} \right) \\
&= \sum_{i,j=1}^{N} \Big(p_j(t) q_{ji} - p_i(t) q_{ij} \Big) \ln p_j^{ss} \;=\; - \sum_{j=1}^{N} \frac{\mathrm{d}p_j(t)}{\mathrm{d}t} \ln p_j^{ss} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left(\sum_{j=1}^{N} p_j(t) \Big(-\ln p_j^{ss} \Big) \right) \;=\; \frac{1}{T} \frac{\mathrm{d}\overline{E}}{\mathrm{d}t},
\end{aligned}
\tag{6.113}
$$

in which

$$
\overline{E} = \sum_{j=1}^{N} p_j(t) E_j,
\tag{6.114}
$$

should be identified as the mean energy, with $E_j = -T \ln p_j^{ss}$ as the "energy" of the state i according to Boltzmann's law. Then, Eq. (6.111a) becomes

$$
\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\overline{E}}{T} - S \right) = -e_p \leq 0.
\tag{6.115}
$$

$F = \overline{E} - TS$ is known as the "free energy" of a thermodynamic system. It is expected to monotonically decreases with time in an isothermal system approaching to equilibrium. In an equilibrium steady state, the free energy reaches its minimum.

6.9.3 Monotonicity of F Change in General Q-Processes

Encouraged by the above results, let us consider the Kullback–Leibler divergence, also known as relative entropy:

$$F(t) = \sum_{i=1}^{N} p_i(t)\Big(- \ln p_i^{ss} + \ln p_i(t)\Big) = \sum_{i=1}^{N} p_i(t) \ln\left(\frac{p_i(t)}{p_i^{ss}}\right) \geq 0. \tag{6.116}$$

One can actually show that $\mathrm{d}F/\mathrm{d}t \leq 0$ for general Q-process without the detailed balance:

$$\begin{aligned}\frac{\mathrm{d}F(t)}{\mathrm{d}t} &= \sum_{i=1}^{N}\left(\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}\right)\ln\left(\frac{p_i(t)}{p_i^{ss}}\right) \;=\; \sum_{i,j=1}^{N}\Big(p_j q_{ji} - p_i q_{ij}\Big)\ln\left(\frac{p_i(t)}{p_i^{ss}}\right)\\ &= \sum_{i,j=1}^{N} p_j q_{ji} \ln\left(\frac{p_i(t)p_j^{ss}}{p_i^{ss}p_j(t)}\right) \;\leq\; \sum_{i,j=1}^{N} p_j q_{ji}\left(\frac{p_i(t)p_j^{ss}}{p_i^{ss}p_j(t)} - 1\right)\\ &= \sum_{i=1}^{N}\frac{p_i}{p_i^{ss}}\left(\sum_{j=1}^{N}\Big(p_j^{ss}q_{ji} - p_i^{ss}q_{ij}\Big)\right) \;=\; 0.\end{aligned} \tag{6.117}$$

6.9.4 F Balance Equation of Markov Dynamics

More interestingly, we have a new, balance equation for the $F(t)$:

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = E_{in}(t) - e_p(t), \tag{6.118a}$$

where $e_p(t) \geq 0$ is given in (6.111b), and

$$E_{in}(t) = \frac{1}{2}\sum_{i,j=1}^{N}\Big(p_i(t)q_{ij} - p_j(t)q_{ji}\Big)\ln\left(\frac{p_i^{ss}q_{ij}}{p_j^{ss}q_{ji}}\right) \geq 0. \tag{6.118b}$$

See [9] for the proof of this inequality. Both $E_{in}(t)$ and $e_p(t)$ are non-negative which means that Eq. (6.118a) can be interpreted as "the $F(t)$ has a source and a sink", its change equals to an input $E_{in}(t)$, a source term, and dissipation $e_p(t)$, a sink term. There is a mesoscopic conservation of the quantity F. Equation (6.118a) is more meaningful than the Eq. (6.111a), in which J_S does not have a definitive sign.

The balance Eq. (6.118a) and the monotonicity of $\mathrm{d}F/\mathrm{d}t \leq 0$ have remarkable resemblance to the first and the second laws of thermodynamics. But they are really a part of a mathematical structure of any stochastic Markov dynamics.

To emphasize this mathematical nature, we call all the results in this section, collectively, ***mathematicothermodynamics*** [9, 10, 21, 24].

6.9.5 Driven System and Cycle Decomposition

The entropy production given in (6.111b) can be written as

$$e_p = \sum_{\text{all edge } ij}^{N} \left(\varphi_{ij} - \varphi_{ji}\right) \ln\left(\frac{\varphi_{ij}}{\varphi_{ji}}\right), \tag{6.119}$$

where $\varphi_{ij} = p_i(t)q_{ij}$ is the one-way probability flux from state i to j. It can be proven that, in a stationary Q-process, the above expression can be expressed also as [14]

$$e_p = \sum_{\text{all cycles } \Gamma}^{N} \left(\varphi_\Gamma^+ - \varphi_\Gamma^-\right) \ln\left(\frac{\varphi_\Gamma^+}{\varphi_\Gamma^-}\right), \tag{6.120}$$

in which $\varphi_\Gamma^\pm$ is the number of Γ cycle completed in a unit time, in the forward and backward direction. Most importantly, for cycle $\Gamma = (i_0, i_1, \cdots, i_n, i_0)$

$$\frac{\varphi_\Gamma^+}{\varphi_\Gamma^-} = \frac{q_{i_0i_1}q_{i_1i_2}\cdots q_{i_{n-1}i_n}q_{i_ni_0}}{q_{i_1i_0}q_{i_2i_1}\cdots q_{i_ni_{n-1}}q_{i_0i_n}}, \tag{6.121}$$

which is independent of the probabilities! Therefore, $\ln\left(\varphi_\Gamma^+/\varphi_\Gamma^-\right)$ can and should be understood as the entropy production per cycle, and the term $\left(\varphi_\Gamma^+ - \varphi_\Gamma^-\right)$ is simply a kinematic term that counts the number of cycle completed along a trajectory. All the nonequilibrium thermodynamics is contained in the (6.121); it is about kinetic cycles [27]. If a Markov process is detail balanced, then its entropy production is zero on each and every kinetic cycle.

It is well known since the work of A. N. Kolmogorov that the quantity in (6.121) equals unity for each and every cycle if and only if the Markov process is detailed balanced. Therefore, the mathematical notion of *detailed balance* provides a fitting description of a non-driven kinetic system whose steady state is an equilibrium. For a driven kinetic system, at least one of the cycles in the state space Γ has unbalanced circulation: $\varphi_\Gamma^+ \neq \varphi_\Gamma^-$.

6.9.6 Macroscopic Thermodynamics in the Kurtz Limit

For a DGP with N species and M reactions, the F function introduced in Sect. 6.9.4 is a functional of the probability distribution $p_V(\mathbf{n}, t)$ which is itself a function of the reaction system's volume V. Then one naturally asks what its macroscopic limit is as $V \to \infty$ as in the Kurtz limit? It can be shown that [10]

$$\begin{aligned}\lim_{V\to\infty} \frac{F\big[p_V(\mathbf{n},t)\big]}{V} &= \lim_{V\to\infty} \frac{1}{V} \sum_{\mathbf{n}} p_V(\mathbf{n},t) \ln\left[\frac{p_V(\mathbf{n},t)}{p_V^{ss}(\mathbf{n})}\right] \\ &= -\lim_{V\to\infty} \frac{1}{V} \sum_{\mathbf{n}} p_V(\mathbf{n},t) \ln p_V^{ss}(\mathbf{n}) \\ &= G^{ss}\big[\mathbf{x}(t)\big], \end{aligned} \tag{6.122}$$

in which $\mathbf{n} = (n_1, n_2, \cdots, n_N)$, n_k is the number of molecules of the kth species, $\mathbf{x} = (x_1, \cdots, x_N)$ is the corresponding number density $\mathbf{x} = \frac{\mathbf{n}}{V}$. The Kurtz theorem in Sect. 6.7 states that the stochastic trajectory of a DGP, $\mathbf{n}_V(t)$,

$$\lim_{V\to\infty} \frac{\mathbf{n}_V(t)}{V} = \mathbf{x}(t), \tag{6.123}$$

where $\mathbf{x}(t)$ is the solution to the deterministic, nonlinear rate equation (e.g., Eq. (6.89)). Most interestingly, according to the large deviation principle from the theory of probability, when the steady state probability $p_V^{ss}(\mathbf{n})$ converges to a Dirac-δ function, its tail probability has an asymptotic expression

$$-\lim_{V\to\infty} \frac{\ln p_V^{ss}(\mathbf{n})}{V} = -\lim_{V\to\infty} \frac{\ln p_V^{ss}(V\mathbf{x})}{V} = G^{ss}(\mathbf{x}). \tag{6.124}$$

This steady state large deviation rate function $G^{ss}(\mathbf{x})$ can be identified as a generalized Gibbs function for nonequilibrium chemical reaction systems. It can be shown that

$$\frac{\mathrm{d}}{\mathrm{d}t} G^{ss}\big[\mathbf{x}(t)\big] = \left(\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t}\right) \cdot \nabla_{\mathbf{x}} G^{ss}(\mathbf{x}) \le 0. \tag{6.125}$$

This is a generalization of the inequality in Eq. (6.77). See [10] for the proof.

6.10 Summary and Conclusion

This chapter presents a new modeling paradigm for biological systems and processes that consist of multiple populations of individuals, each with an infinite many internal degrees of freedom. The individuals are grouped into subpopulations and mathematically represented by their statistical behaviors in terms of birth, death, migration, and state switching. We show that the population kinetics in terms of nonlinear ordinary differential equations (ODEs) widely employed in mathematical biology is fundamentally a stochastic kinetic theory. This stochastic population kinetic representation of biological reality can be introduced quite rigorously, thus it provides one with confidence in the conclusions drawn from mathematical analysis. We called this formalism *Delbrück-Gillespie process*. In the large population limit, T. G. Kurtz's theorem, a law of large numbers, yields a system of nonlinear rate equations that is consistent with the traditional ODEs. In Sect. 6.9, very recent results on mesoscopic nonequilibrium thermodynamics and its corresponding macroscopic nonequilibrium thermodynamics are presented. Together the three parts, (1) stochastic kinetics in terms of DGP, (2) deterministic nonlinear dynamics in terms of ODEs, and (3) the mathematicothermodynamics, provide a comprehensive mathematical theory for a wide range of biological systems and processes from biochemistry to ecology.

6.11 Exercises: Simple and Challenging

6.11.1 Simple Exercises

1. Compute the expected value and the variance of an exponentially distributed random variable X with rate λ.

2. Let $X_1, \cdots, X_n$ be n i.i.d. exponential random variables with rate λ. Let $X^* = \min\{X_1, X_2, \cdots, X_n\}$. Show that $f_{T^*}(t) = n\lambda e^{-n\lambda t}$.

3. If a set of n i.i.d. random times all with distribution $f_T(t)$, $f_T(0) = 0$ but $f'_T(0) \neq 0$, what is the distribution for $T^* = \min\{T_1, T_2, \cdots, T_n\}$ in the limit of $n \to \infty$?

6.11.2 More Challenging Exercises

4. Consider a population consisting of identical and independent individual organisms, each with an exponentially distributed time for giving "birth", with rate λ, and going "death", with rate μ.

(i) Now when the population has exactly n individuals, what is the probability distribution for the waiting time to the next birth? What is the probability distribution for the waiting time to the next death? What is the probability distribution for the waiting time to the next birth or death event?

(ii) Let $p_n(t)$ be the probability of having exactly n individuals in the population at time t:

$$\sum_{n=0}^{\infty} p_n(t) = 1.$$

What system of differential equations should $p_n(t)$ satisfy?

(iii) The mean population at time t is defined as

$$\langle n \rangle(t) = \sum_{n=0}^{\infty} n p_n(t).$$

Based on the system of differential equations you obtained in (ii), show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle n \rangle = \Big(\lambda - \mu\Big)\langle n \rangle.$$

5. The 3-state Markov system,

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} A, \tag{6.126}$$

has been widely used in biochemistry to model the conformational changes of a single protein molecule undergoing through its three different states A, B, and C. For example, A is non-active, B is partially active, and C is fully active.

(a) The probabilities for the states, $\mathbf{p} = (p_A, p_B, p_C)$, satisfies a differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}(t) = \mathbf{p}(t)\mathbf{Q},$$

where $\mathbf{Q}$ is a 3×3 matrix. Write the $\mathbf{Q}$ out in terms of the k's. Show that the sum of each and every row is zero. Discuss in probabilistic terms, what is the meaning of this result?

(b) Compute the steady state probabilities p_A^{ss}, p_B^{ss}, and p_C^{ss}, and show that, in the steady state, the net (probabilistic) flux from state A to B,

$$J_{A\to B}^{ss} = k_1 p_A^{ss} - k_{-1} p_B^{ss},$$

is the same as the net flux from state $B \to$ state C, and also the net flux from $C \to A$. Since they are all the same, it is called the steady state flux J^{ss} of the biochemical reaction cycle in (6.126).

(c) What is the condition, in terms of all the k's, for $J^{ss} = 0$?

6. Consider a single enzyme E in the sea of substrate molecule S. The Michaelis–Menten kinetics is

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E^* + P. \tag{6.127}$$

Because there is only a single enzyme molecule working, the concentration of S can be assumed as always constant, at the value c_S.

Write the differential equations for the probability of the enzyme being in state E, ES, and E^*: $p_E(t)$, $p_{ES}(t)$, and $p_{E^*}(t)$.

Given initial condition $p_E(0) = 1$, $p_{ES}(0) = 0$, and $p_{E^*}(0) = 0$, try to solve $p_{E^*}(t)$.

It is clear that the time for the enzyme to move from state E to E^* is stochastic. Let T be the random time. What is the probability distribution for T, $f_T(t)$? How is it related to $p_{E^*}(t)$?

Compute expected value $\mathbb{E}[T]$. Compare your result with the Michaelis–Menten formula.

References

1. D.A. Beard, H. Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems*, in Texts in Biomedical Engineering (Cambridge University Press, London, 2008)
2. P.G. Bergmann, J.L. Lebowitz, New approach to nonequilibrium processes. Phys. Rev. **99**, 578–587 (1955)
3. R.D. Cox, *Renewal Theory* (Methuen & Company, New York, 1970)
4. P.G. de Gennes, Molecular individualism. Science **276**, 1999–2000 (1997)
5. S.R. de Groot, P. Mazur, *Nonequilibrium Thermodynamics* (Dover, New York, 1984)
6. A.W.F. Edwards, The fundamental theorem of natural selection. Biol. Rev. **69**, 443–474 (1994)
7. M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, 3rd ed. (Springer, New York, 2012)
8. G. Gallavotti, *Statistical Mechanics: A Short Treatise* (Springer, Berlin, 1999)
9. H. Ge, H. Qian, The physical origins of entropy production, free energy dissipation and their mathematical representations. Phys. Rev. E **81**, 051133 (2010)
10. H. Ge, H. Qian, Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory. Phys. Rev. E **94**, 052150 (2016)
11. R.J. Herrnstein, C. Murray, *Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York, 1996)
12. S. Huang, F. Li, J.X. Zhou, H. Qian, Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics: differentiations in living matters (review). J. R. Soc. Interface **14**, 20170097 (2017)
13. F. Jacob, *Possible and Actual—Jessie and John Danz Lectures* (University Washington Press, Seattle, 1994)
14. D.-Q. Jiang, M. Qian, M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States*. LNM, vol. 1833 (Springer, New York, 2004)
15. S.D. Levitt, S.J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (William Morrow, New York, 2005)

16. A.J. Lotka, Analytical note on certain rhythmic relations in organic systems. Proc. Natl. Acad. Sci. USA **6**, 410–415 (1920)
17. J.D. Murray, *Mathematical Biology I: An Introduction*, 3rd ed. (Springer, New York, 2002)
18. L. Onsager, Reciprocal relations in irreversible processes, I. Phys. Rev. **37**, 405–426 (1931)
19. H. Qian, Cellular biology in terms of stochastic nonlinear biochemical dynamics: emergent properties, isogenetic variations and chemical system inheritability. J. Stat. Phys. **141**, 990–1013 (2010)
20. H. Qian, Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems–an analytical theory (invited article). Nonlinearity **24**, R19–R49 (2011)
21. H. Qian, A decomposition of irreversible diffusion processes without detailed balance. J. Math. Phys. **54**, 053302 (2013)
22. H. Qian, *Nonlinear Stochastic Dynamics of Complex Systems, I: A Chemical Reaction Kinetic Perspective with Mesoscopic Nonequilibrium Thermodynamics* (2016). ArXiv:1605.08070
23. H. Qian, L.M. Bishop, The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks (Review). Int. J. Mol. Sci. **11**, 3472–3500 (2010)
24. H. Qian, M. Qian, X. Tang, Thermodynamics of the general diffusion process: time-reversibility and entropy production. J. Stat. Phys. **107**, 1129–1141 (2002)
25. H. Qian, S. Saffarian, E.L. Elson, Concentration fluctuations in a mesoscopic oscillating chemical reaction system. Proc. Natl. Acad. Sci. USA **99**, 10376–10381 (2002)
26. H. Qian, P. Ao, Y. Tu, J. Wang, A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales. Chem. Phys. Lett. **665**, 153–161 (2016)
27. H. Qian, S. Kjelstrup, A.B. Kolomeisky, D. Bedeaux, Entropy production in mesoscopic stochastic thermodynamics—Nonequilibrium kinetic cycles driven by chemical potentials, temperatures, and mechanical forces (Topical review). J. Phys. Condens. Matter **28**, 153004 (2016)
28. D.B. Saakian, H. Qian, *Nonlinear Stochastic Dynamics of Complex Systems, III: Nonequilibrium Thermodynamics of Self-Replication Kinetics* (2016). ArXiv:1606.02391
29. C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (University Illinois Press, Chicago, 1963)
30. H.M. Taylor, S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed. (Academic Press, New York, 1998)
31. L.F. Thompson, H. Qian, Nonlinear Stochastic Dynamics of Complex Systems, II: Potential of Entropic Force in Markov Systems with Nonequilibrium Steady State, Generalized Gibbs Function and Criticality (2016). ArXiv:1605.08071
32. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, revised and enlarged ed. (Elsevier, Amsterdam, 1992)
33. M. Vellela, H. Qian, A quasi-stationary analysis of a stochastic chemical reaction: Keizer's paradox. Bull. Math. Biol. **69**, 1727–1746 (2007)
34. M. Vellela, H. Qian, Stochastic dynamics and nonequilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. J. R. Soc. Interface **6**, 925–940 (2009)
35. M. Vellela, H. Qian, On Poincaré-Hill cycle map of rotational random walk: locating stochastic limit cycle in reversible Schnakenberg model. Proc. R. Soc. A Math. Phys. Eng. **466**, 771–788 (2010)
36. J. Voigt, Stochastic operators, information, and entropy. Commun. Math. Phys. **81**, 31–38 (1981)
37. D. Zhou, H. Qian, Fixation, transient landscape and diffusion's dilemma in stochastic evolutionary game dynamics. Phys. Rev. E **84**, 031907 (2011)

Chapter 7
The Turing Model for Biological Pattern Formation

Philip K. Maini and Thomas E. Woolley

Abstract How spatial patterning arises in biological systems is still an unresolved mystery. Here, we consider the first model for spatial pattern formation, proposed by Alan Turing, which showed that structure could emerge from processes that, in themselves, are non-patterning. He therefore went against the reductionist approach, arguing that biological function arises from the *integration* of processes, rather than being attributed to a single, unique, process. While still controversial, some 65 years on, his model still inspires mathematical and experimental advances.

7.1 Biological Pattern Formation

Biological systems exhibit a diverse range of patterns, such as animal pigmentation patterns, limb skeletal structures, etc. (Fig. 7.1). Despite decades of research, a detailed understanding of how these patterns arise still eludes us. We know many of the genes involved and can map out the spatiotemporal dynamics of some of them, but how these dynamics arise is still largely a mystery. In 1952, the logician, computer scientist, code breaker and mathematician Alan Turing proposed a novel mathematical model for pattern formation [1]. He hypothesised that the patterns we see arise due to cells responding to underlying *pre-patterns* of chemical concentrations. He termed these chemicals *morphogens* and showed that spatially heterogeneous patterns could arise in systems in which these chemicals reacted with each other and also underwent diffusion—a phenomenon termed *diffusion-driven instability*. Making the further assumption that cell fate was determined in a morphogen concentration-dependent manner, the chemical pre-pattern would

P. K. Maini (✉)
Wolfson Centre for Mathematical Biology, Mathematical Institute,
University of Oxford, Oxford, UK
e-mail: maini@maths.ox.ac.uk

T. E. Woolley
School of Mathematics, Cardiff University, Cardiff, UK
e-mail: WoolleyT1@cardiff.ac.uk

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics
of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_7

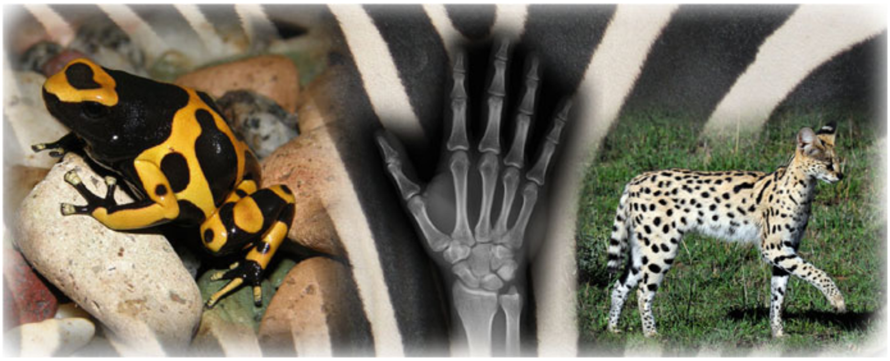


Fig. 7.1 Examples of biological pattern formation. Zebra stripes are shown in the background and going from left to right: poison arrow frog labyrinthine pigmentation pattern; digit pattern of a human; serval spots transitioning to stripes on the tail

manifest itself in a pattern composed of spatially heterogeneous cell fates. In Sect. 7.2 we describe the phenomenon of diffusion-driven instability and deduce the properties exhibited by the resultant patterns. We will also give some examples of reaction-diffusion systems. In Sect. 7.3 we present some applications and in Sect. 7.4 we present conclusions and discussion.

7.2 Mathematical Model

7.2.1 Diffusion is Stabilising

Let us consider the case of a chemical, concentration $u(x, t)$, diffusing in space x (assumed to be in one dimension for simplicity), where t is time. Let us also assume that the chemical is being produced at a rate $f(u)$ where f is typically either a polynomial, or rational, function of u. Then

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + f(u), \tag{7.1}$$

where $D > 0$ is the diffusion coefficient (assumed constant), is the *reaction-diffusion* equation satisfied by $u(x, t)$.

We will assume further that the spatial domain is $[0, L]$ for some $L > 0$ and that the chemical concentration at the edge of the domain is fixed at some value u_0, that is,

$$u(x, t) = u_0 \text{ at } x = 0, L \text{ and } \forall\, t. \tag{7.2}$$

This is often called a Dirichlet boundary condition. Furthermore, suppose that $f(u_0) = 0$. Then, $u(x, t) = u_0$ satisfies Eq. (7.1) and the boundary conditions (7.2) and is termed a *spatially uniform steady state* for u. To find the linear stability of this state we wish to determine if a small perturbation $\hat{u}(x, t)$ from the steady state will grow or decay in time. Substituting into Eq. (7.1), expanding f in a Taylor series and keeping only the linear terms, we have:

$$\frac{\partial \hat{u}}{\partial t} = D\frac{\partial^2 \hat{u}}{\partial x^2} + f'(u_0)\hat{u} \tag{7.3}$$

where $f' = df/du$ and we have used the fact that $f(u_0) = 0$. Furthermore, $\hat{u}$ satisfies $\hat{u}(x, t) = 0$ at $x = 0, L$.

In the case where $D = 0$, Eq. (7.3) has the solution

$$\hat{u}(x, t) = \hat{u}_0 \exp(f'(x_0)t) \tag{7.4}$$

where $\hat{u}_0$ is the initial perturbation. Clearly, if $f'(u_0) < 0$ then the steady state is linearly stable (as t tends to infinity, $\hat{u}(t)$ tends to zero) while if $f'(u_0) > 0$, then the perturbation grows and the steady state is linearly unstable.

Now suppose that $D > 0$ in (7.3). Then, using the method of separation of variables, and taking into account the boundary conditions (7.2), the solution for $\hat{u}(t)$ is the Fourier sine series

$$\hat{u}(x, t) = \sum_{n=1}^{n=\infty} a_n \sin\left(\frac{n\pi x}{L}\right) \exp(\lambda_n t) \tag{7.5}$$

where $\lambda_n = f'(u_0) - D(n\pi/L)^2$, for $n = 1, 2, \ldots$, and a_n are determined by equating the solution to the Fourier sine series of the initial condition for the perturbation. Now we see that even when $f'(u_0) > 0$, if $D > f'(u_0)(L/\pi)^2$, it follows that $\lambda_n < 0 \; \forall \; n$, that is, each term $\sin(n\pi x/L)$ in the Fourier expansion, termed *an admissible mode*, will have either an exponentially decaying amplitude (if $a_n \neq 0$) or zero amplitude (if $a_n = 0$) and so $\hat{u}(x, t)$ tends to zero as t tends to infinity. Therefore, the steady state $u = u_0$, although unstable in the absence of diffusion, is stabilised by the presence of diffusion. Hence, diffusion is stabilising.

Note that if the boundary conditions were instead zero flux (so-called homogeneous Neumann) boundary conditions ($\partial u/\partial x = 0$ at $x = 0, L \; \forall \; t$) then the solution to the linearized system would be a Fourier cosine series and so, in the presence of diffusion, the zeroth mode (constant) term in the Fourier expansion could still grow but every spatially heterogeneous (patterned) term, $\cos(n\pi x/L)$, would have negative growth rate for sufficiently large D. For periodic boundary conditions ($u(0, t) = u(L, t) \; \forall \; t$), the Fourier series solution would now be a combination of sines and cosines but the above arguments still hold.

7.2.2 Diffusion is De-stabilising

The previous mathematical result, namely that diffusion is a stabilising process, also agrees with our intuition, for example if we think of heat. The genius of Turing was to show that this was not necessarily the case if there was more than one chemical, that is, if we had a reaction-diffusion *system*. Let $u(x,t)$ and $v(x,t)$ be two chemicals satisfying the equations:

$$\frac{\partial u}{\partial t} = D_1 \frac{\partial^2 u}{\partial x^2} + f(u,v), \quad \frac{\partial v}{\partial t} = D_2 \frac{\partial^2 v}{\partial x^2} + g(u,v), \tag{7.6}$$

where $f(u,v)$ and $g(u,v)$ are functions describing the reaction kinetics of the morphogens represented by u and v, and D_1 and D_2 are constant (positive) diffusion coefficients. For simplicity, let us assume that once again x is the finite domain $[0, L]$ and that u and v satisfy homogeneous Neumann boundary conditions.

Now suppose there are positive values (u_0, v_0) such that $f(u_0, v_0) = g(u_0, v_0) = 0$. Then (u_0, v_0) is spatially uniform steady state of the system (7.6). To examine the linear stability of this steady state we extend the analysis in Sect. 7.2.1 by deriving equations for small perturbations $(\hat{u}(x,t), \hat{v}(x,t))$ to the steady state. Substituting into Eq. (7.6), expanding f and g in Taylor series and recalling that $f(u_0, v_0) = g(u_0, v_0) = 0$, we arrive (ignoring higher order terms) at the linearized system:

$$\frac{\partial \hat{u}}{\partial t} = D_1 \frac{\partial^2 \hat{u}}{\partial x^2} + f_u \hat{u} + f_v \hat{v}, \quad \frac{\partial \hat{v}}{\partial t} = D_2 \frac{\partial^2 \hat{v}}{\partial x^2} + g_u \hat{u} + g_v \hat{v}, \tag{7.7}$$

where f_u, f_v, g_u, g_v denote the partial derivatives of f and g evaluated at the steady state (u_0, v_0). We may re-write this in the more concise form:

$$\frac{\partial \hat{\boldsymbol{u}}}{\partial t} = \boldsymbol{D} \frac{\partial^2 \hat{\boldsymbol{u}}}{\partial x^2} + \boldsymbol{J}\left(\hat{\boldsymbol{u}}\right), \tag{7.8}$$

where

$$\hat{\boldsymbol{u}} = \begin{pmatrix} \hat{u}(x,t) \\ \hat{v}(x,t) \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \text{ and } \boldsymbol{J} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}. \tag{7.9}$$

We generalise the analysis in Sect. 7.2.1 by looking for a solution of the form $\hat{\boldsymbol{u}}(x,t) = \boldsymbol{a} \exp(ikx + \lambda(k^2)t)$ where, again, we are looking for a separable solution, in this case with $\boldsymbol{a}$ a constant vector and the x component of the solution is written as $\exp(ikx)$—a convenient way to encompass the Fourier components. Substituting this into Eq. (7.8), we arrive at the equation

$$\left(\boldsymbol{J} - \boldsymbol{D}k^2 - \lambda \boldsymbol{I}\right)\boldsymbol{a} = \boldsymbol{0}, \tag{7.10}$$

where $\boldsymbol{I}$ is the 2×2 unit matrix. For non-trivial solutions, we thus require that the matrix multiplying the vector $\boldsymbol{a}$ is singular, that is,

$$Det\left(\boldsymbol{J} - \boldsymbol{D}k^2 - \lambda \boldsymbol{I}\right) = 0, \tag{7.11}$$

where Det denotes the determinant. This is an eigenvalue problem, that is, the temporal growth rate, λ, is the eigenvalue of the matrix $\boldsymbol{J} - \boldsymbol{D}k^2$ and is, in fact, a function of the wave number, k.

Now, in the previous subsection, we showed that in the case of a single reaction-diffusion equation, a spatially uniform steady state, linearly stable in the absence of diffusion, could be stabilised in the present of diffusion. Here, Turing showed the opposite. Let us consider the case when $D_1 = D_2 = 0$. Then λ is simply the eigenvalue of the matrix $\boldsymbol{J}$ and satisfies the *eigenvalue problem*

$$\lambda^2 - (f_u + g_v)\lambda + (f_u g_v - f_v g_u) = 0. \tag{7.12}$$

For the spatially uniform steady state to be stable, we require both solutions to the eigenvalue Eq. (7.12) to have negative real part, and this will be true if the following two conditions hold:

$$f_u + g_v < 0, \text{ and } f_u g_v - f_v g_u > 0. \tag{7.13}$$

Now, in the presence of diffusion (D_1 and D_2 both non-zero), the eigenvalue problem, from Eq. (7.11), relating the growth rate, λ, to the wave number, k, is

$$\lambda^2 - b(k^2)\lambda + c(k^2) = 0, \tag{7.14}$$

where

$$b(k^2) = f_u + g_v - (D_1 + D_2)k^2 \text{ and } c(k^2) = D_1 D_2 k^4 - (D_2 f_u + D_1 g_v)k^2 + f_u g_v - f_v g_u. \tag{7.15}$$

In this case, we wish diffusion to be de-stabilising and a necessary condition for this to hold true is that at least one of the roots, $\lambda(k^2)$, of Eq. (7.14) must have a positive real part for some non-zero (positive) k^2. This can happen if either $b(k^2) > 0$ or $c(k^2) < 0$. However, the first condition in (7.13), and the fact that the diffusion coefficients are non-negative, ensures that $b(k^2) < 0$, so we require $c(k^2) < 0$. For this to occur, the second condition in (7.13) forces $D_2 f_u + D_1 g_v$ to be positive as a necessary condition. More precisely, we require

$$D_2 f_u + D_1 g_v > 2\sqrt{D_1 D_2 (f_u g_v - f_v g_u)} > 0. \tag{7.16}$$

Conditions (7.13) and (7.16) ensure that the uniform steady state is linearly stable in the absence of diffusion but has at least one k for which $\lambda(k^2)$ has positive real part. However, to satisfy the zero flux boundary conditions, admissible modes are restricted to $k = n\pi/L$ for at least one integer value n. This leads to the 4th condition:

$$k_-^2 < \left(\frac{n\pi}{L}\right)^2 < k_+^2 \text{ where } k_\pm^2 = \frac{f_u + g_v \pm \sqrt{(f_u + g_v)^2 - 4D_1D_2(f_ug_v - f_vg_u)}}{2D_1D_2}. \tag{7.17}$$

Hence, if these conditions are satisfied for at least one integer value, $n > 0$, we see that a spatially uniform steady state, stable in the absence of diffusion, becomes unstable in the presence of diffusion. This is termed *diffusion-driven instability* and is an example of *self-organisation* (or sometimes termed an *emergent* phenomenon).

A number of key properties are immediately apparent from these conditions [2]:

1. The diffusion coefficients must be unequal. This follows from the first inequality in (7.13) and (7.16), because if $D_1 = D_2 = D > 0$ then we can divide inequality (7.16) by D to obtain $f_u + g_v > 0$, which contradicts the first inequality of (7.13).
2. The matrix of partial derivatives $\boldsymbol{J}$ must take one of the following two forms:

$$\boldsymbol{J_p} = \begin{pmatrix} + & - \\ + & - \end{pmatrix} \text{ or } \boldsymbol{J_c} = \begin{pmatrix} - & - \\ + & + \end{pmatrix}. \tag{7.18}$$

   This follows from the first inequality in (7.13) and (7.16) and observing that the matrix forms

$$\boldsymbol{J}_{\boldsymbol{p}}' = \begin{pmatrix} - & + \\ - & + \end{pmatrix} \text{ and } \boldsymbol{J}_{\boldsymbol{c}}' = \begin{pmatrix} + & + \\ - & - \end{pmatrix} \tag{7.19}$$

   are in fact captured, respectively, by $\boldsymbol{J_p}$ and $\boldsymbol{J_c}$, by appropriately re-defining u and v or f and g. In detail, from the first inequality in (7.13) and (7.16) it follows that f_u and g_v must have opposite signs. Hence $f_ug_v < 0$, and the second inequality of (7.13) forces f_vg_u to be less than zero, implying that f_v and g_u have opposite signs.
3. There is a minimum domain size for pattern formation. This follows from inequality (7.17). For fixed parameter values in the reaction-diffusion model, for L sufficiently small, this inequality cannot be satisfied for non-zero n.
4. As the domain size increases, the pattern becomes more complicated for two reasons: (1) the lower inequality of Eq. (7.17) means that the minimum allowable wave mode increases. Explicitly, the pattern appearing on a larger domain will have a larger number of peaks and troughs (i.e. larger n), when compared to the pattern on a smaller domain; (2) the range of allowable modes increases. This again follows from inequality (7.17) by observing that as L increases the number of viable integers can also increase.

5. The idea from point 4 can be extended to higher dimensions. For example, if the spatial domain is the 2-dimensional rectangle $[0, L_x] \times [0, L_y]$ then the admissible modes take the form (for zero flux boundary conditions)

$$\cos(n\pi x/L_x)\cos\left(m\pi y/L_y\right),$$

where $k^2 = (n\pi/L_x)^2 + (m\pi/L_y)^2$ and $n = 0, 1, 2, \ldots, m = 0, 1, 2, \ldots$. Clearly, if L_y is very small, while L_x is large (that is, the domain is long and thin), then from the obvious extension of condition (7.17) to this case, it follows that $m = 0$ and so any spatially patterned structure will vary only in the x-direction, that is, the system will exhibit stripes. However, if L_x and L_y are both large, then (7.17) can hold for n and m both non-zero. In this case, we have spots. Note that in 2 dimensions we have the issue of degeneracy. For example, suppose $L_x = L_y = 1$ and (u_0, v_0) was unstable to a mode with $k^2 = 25\pi^2$, then the admissible modes would have wave number pairs $(5, 0)$, $(0, 5)$ (both corresponding to stripes) and $(3, 4)$, $(4, 3)$ (both corresponding to spots). In this case, initial conditions and the form of the non-linear terms determine which mode is selected (or indeed the solution could be a combination of all possibilities). More recently, these results have been generalised using an energy function to show that pattern selection can be determined by investigating the stationary solutions of an associated Fokker–Planck equation [3].

It is important to point out that the above analysis, and properties derived, hold for linear theory, while the original system is non-linear. The obvious question to ask is, do the linear results hold for the non-linear system? While this can be answered to some extent by carrying out a weakly non-linear analysis in the vicinity of a primary bifurcation point [4, 5], we need to resort to numerical solution of the non-linear system for a fuller answer. In Fig. 7.2 we show some results of numerical simulations of the full non-linear system (see Sect. 7.2.3) to illustrate the properties 3–5.

In the above, if the linearized kinetics are represented at (u_0, v_0) by $\boldsymbol{J_p}$, the system is termed a pure *activator-inhibitor* system while, if they are represented by $\boldsymbol{J_c}$, the system is called a cross activator-inhibitor system or a substrate-depletion system. We now explain this terminology. For the case $\boldsymbol{J_p}$ we see that at the spatially uniform steady state, $f_u > 0$ and $f_v < 0$. Hence, at steady state, u is activating its own production, but v is inhibiting the production of u. Moreover, $g_u > 0$ and $g_v < 0$ which means that u activates the production of v. Hence, u is termed an *activator* and v is termed an *inhibitor*. Note further that from the first inequality in (7.13), and (7.16) it follows that $D_2 > D_1$. That is, the inhibitor diffuses more rapidly than the activator. This leads to the self-organising patterning principle of *short-range activation long-range inhibition* [6]. For the case $\boldsymbol{J_c}$, u is a substrate that produces v but is itself depleted. Note that if we calculate the eigenvector $\boldsymbol{a}$ in the case of $\boldsymbol{J_p}$ then equating the second component to zero in the vector Eq. (7.8) forces the components of u and v to have the same sign, that is, the solutions are *in phase*. Conversely, for $\boldsymbol{J_c}$, equating the first component to zero in the vector Eq. (7.8)

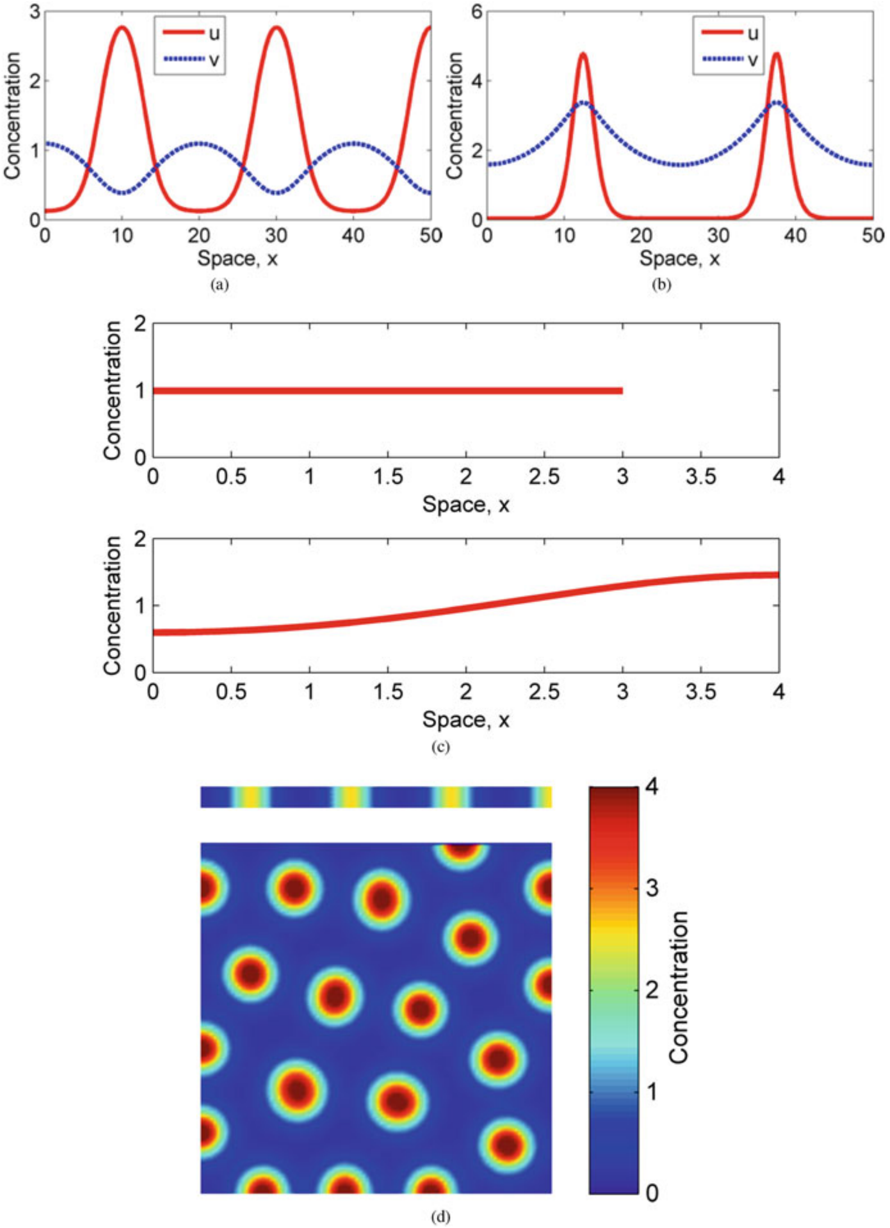


Fig. 7.2 Turing pattern properties. (**a**) Schnakenberg kinetics $D_1 = 1$, $D_2 = 40$, $a = 0.1$ and $b = 0.9$. (**b**) Gierer-Meinhardt kinetics $D_1 = 0.7$, $D_2 = 70$, $a = 0.03$ and $b = 1$. (**c**) If the Schnakenberg kinetics are on a domain of length 3 no pattern emerges. However, a domain length of 4 allows heterogeneity to appear. (**d**) Two-dimensional simulation of the Schnakenberg kinetics. The top simulation shows a thin rectangle that is only able to support stripes across the domain. However, when we increase the vertical height we see that the pattern can produce spots. Note that only one chemical concentration is shown in (**c**) and (**d**), the other one will be 180° out of phase

implies that u and v must have opposite signs and therefore the solutions are *180° out of phase* (see Fig. 7.2a, b).

7.2.3 Defining the Reaction Kinetics

The functions $f(u, v)$ and $g(u, v)$ can take many forms, too numerous for us to list them all, so we simply give a small sample here. Perhaps the best known is the Gierer-Meinhardt model variant [6] that, when non-dimensionalised, takes the form

$$f(u, v) = a - bu + \frac{u^2}{v}; \;\; g(u, v) = u^2 - v, \tag{7.20}$$

where a and b are positive constants. Here, the model has been constructed such that v inhibits u but u activates v. Another model of this class presented by Gierer and Meinhardt, but later derived from a hypothetical chemical reaction using the Law of Mass Action, is the Schnakenberg model [7], which takes the form

$$f(u, v) = a - u + u^2 v; \;\; g(u, v) = b - u^2 v, \tag{7.21}$$

where a and b are constants. The Thomas model [8], on the other hand, describes the interaction of uric acid, u, with oxygen, v, where both reactants diffuse from a reservoir maintained at fixed concentrations, and interact via kinetics empirically determined by data fitting:

$$f(u, v) = \alpha(a - u) - \frac{uv}{c + u + du^2}; \;\; g(u, v) = \beta(b - v) - \frac{uv}{c + u + du^2}, \tag{7.22}$$

where a, b, c, d, α and β are positive constants. More recently, Barrio et al. [9] proposed a caricature model (which we will denote as BVAM) for ease of analysis. In this model, they simply postulated a system in which the linear, quadratic and cubic terms are explicit:

$$f(u, v) = au + v - r_1 uv^2 - r_2 uv; \;\; g(u, v) = bu + cv + r_1 uv^2 + r_2 uv, \tag{7.23}$$

where a, b, c, r_1, and r_2 are non-negative constants. The inclusion of non-linear terms of this form allowed them to easily consider the case where the linear system is degenerate and the non-linear terms then specify the pattern, with quadratic terms favouring spots and cubic terms favouring stripes [10]. Of course, this model can exhibit negative values of u and v which at first, appear unphysical, but u and v should not be interpreted as concentrations, but rather as *deviations* from some positive spatially uniform steady state. This seemingly simple model actually gives rise to an incredibly large range of different types of patterns [11]. In Fig. 7.3, we present a selection of stationary patterns while in Fig. 7.4 we illustrate some

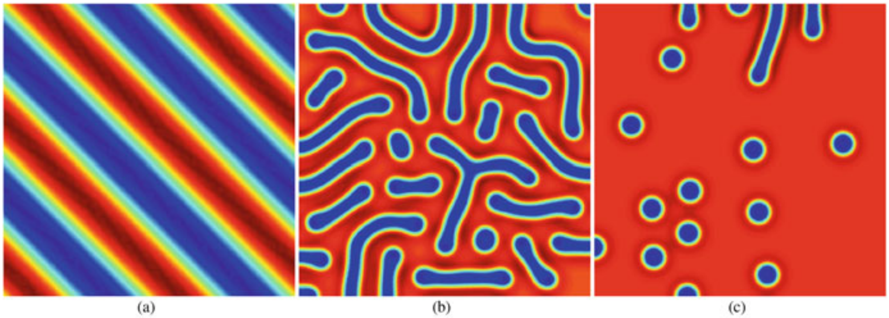


Fig. 7.3 Stationary patterns appearing in the BVAM model. We see that we can generate stripes, labyrinthine patterns and spots, respectively

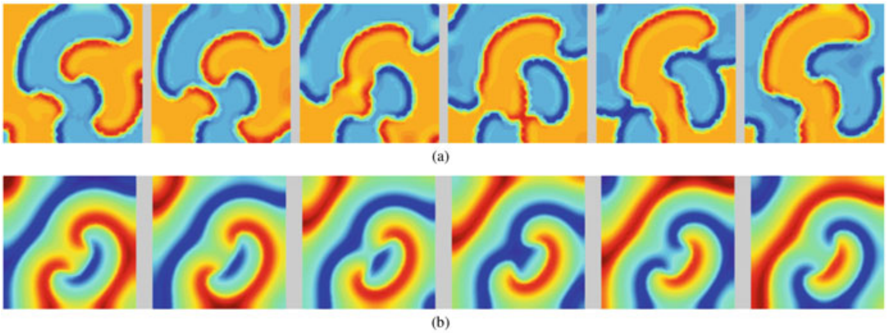


Fig. 7.4 Non-stationary patterns appearing in the BVAM model. The images illustrate snapshots of non-stationary patterns, with time increasing to the right along each row. In (**a**) we see travelling waves, whilst (**b**) shows scroll waves

temporally evolving patterns. The latter are non-stationary, and their analysis is beyond the scope of this chapter.

7.3 Applications

The ability of the Turing model to produce patterning has meant that it has been used in a bewildering array of applications, ranging from regeneration in Hydra, to digit patterning, to animal pigmentation, shells, hair, teeth and feather patterns, etc. (see [12–14] and Fig. 7.5). While the model can produce an astonishing array of patterns, properties 3 and 4 suggest that the patterns that are formed are constricted by domain size. Specifically, property 3 states that if the patterning region is too small, then no pattern will form even if the reaction parameters are chosen to produce a Turing pattern. Further, property 4 suggests that as the domain increases in size, the derived linear theory predicts that patterns increase in complexity and,

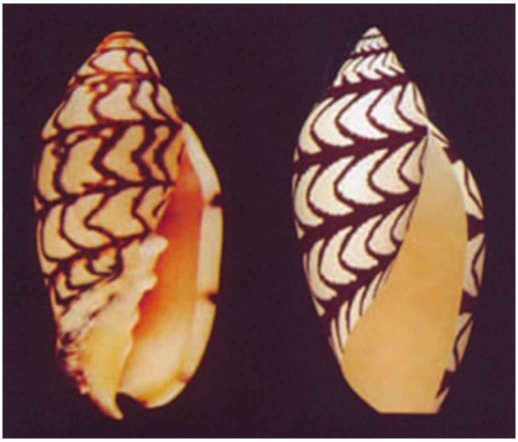


Fig. 7.5 Illustration of more complex Turing dynamics that can reproduce the patterns seen on seashells. Taken from [13]

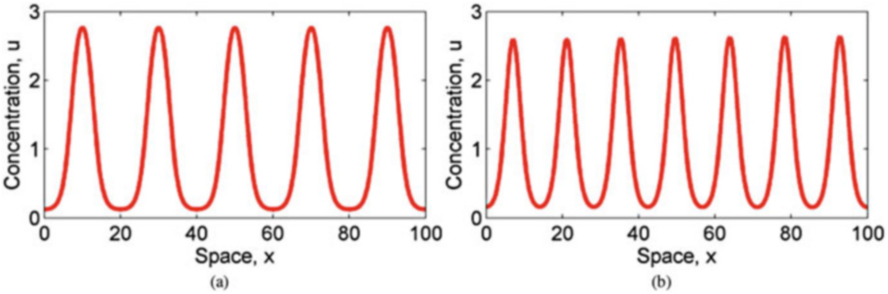
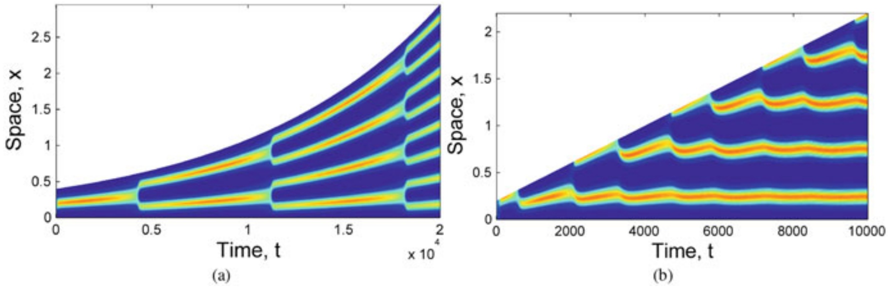


Fig. 7.6 Illustrating the robustness issue. All solutions are stationary. The simulations are of the Schnakenberg model [7] and the parameters are the same for both simulations, the only difference being that the initial conditions were two different randomly chosen states which were uniformly distributed around their unique spatially uniform steady state

conversely, a decrease in domain dimensions would reduce patterning complexity. This is an example of a *developmental constraint* [15].

One of the problems with Turing reaction-diffusion models is that the patterns they produce can be very sensitive to small variations in parameter values and to variations in initial conditions, questioning their applicability to situations where robustness is essential—for example, we only want one head!—and this was first pointed out by Bard and Lauder [16]. This sensitivity can either arise due to the fact that the parameter space in which Turing patterns can form can be very small [17] or because the system can exhibit multiple stable spatially heterogeneous solutions (see, for example, Fig. 7.6).

This issue is still not fully resolved. Dillon et al. [18] showed that choosing different types of boundary conditions could enhance the robustness of some modes



(a)

(b)

Fig. 7.7 Deterministic simulations of the Schnakenberg model [7] on a domain growing (**a**) exponentially uniformly everywhere, and (**b**) linearly, but only at the tip (apical growth)

while eliminating the admissibility of other modes. Crampin et al. [19] reformulated the model on a growing domain, noting equations of the form (7.1) will transform to

$$\frac{\partial u}{\partial t} + \frac{\partial s u}{\partial x} = D\frac{\partial^2 u}{\partial x^2} + f(u), \tag{7.24}$$

where s is the velocity of flow induced by growth (in higher dimensions, this term would take the form $\nabla.(\boldsymbol{s}u)$, see the original paper for the derivation of this form). They showed that this system could robustly generate mode doubling for the case of uniform domain growth as well as generating, in a robust fashion, a sequence of consecutive modes for the case of apical growth (Fig. 7.7).

An obvious question to ask now is, if pattern complexity increases with domain size under this theory, then shouldn't we have a large number of heads? However, cells can only respond to signals for a certain time window before they differentiate and therefore can no longer respond to changing signals. One dramatic situation in which this is not the case is pigmentation patterning in certain fishes where, as the domain grows, the pattern continually changes to preserve wavelength, again consistent with the Turing model [20].

Another key issue is the identification of the morphogens involved. While this remains a controversial issue, there have been many potential activator-inhibitor pairs identified (see, for example [21–23]). Moreover, it has been posited that the activator-inhibitor system might actually be composed of cells—a particular example being the interaction of melanophores and xanthophores in zebrafish pigmentation patterning [24–27].

While in the above we have considered robustness in response to different initial conditions or parameter values, none of these studies investigated the effect of noise, which we would expect to be present in a biological system throughout the patterning process. Woolley et al. [28] showed the presence of noise could disrupt the robust period-doubling patterning sequence seen by Crampin et al. [19] (Fig. 7.7) but robustness was preserved in the case of apical growth (Fig. 7.8).

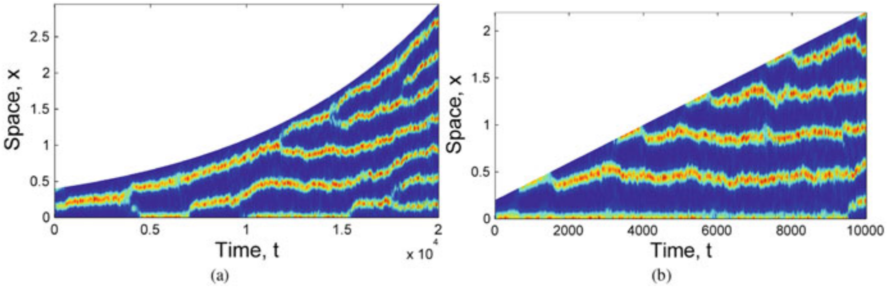


Fig. 7.8 Stochastic simulations of the Schnakenberg model [7] on a domain growing (**a**) exponentially uniformly everywhere, and (**b**) linearly, but only at the tip (apical growth)

This offers a possible reason why, in biology, we usually see patterns forming behind a propagating front, rather than simultaneously across the full domain. A propagating front allows patterns to form in a sequentially controlled and robust fashion.

The robustness issue has recently been tackled in a different way by Kurics et al. [29] who showed that extending the Turing model to be more biologically realistic by including receptor and feedback dynamics actually could greatly enhance the parameter space in which patterns are predicted.

While still controversial biologically, Turing structures have been found in chemistry—the first example being the chloride-iodide-malonic-acid (or CIMA) reaction [30]. One of the reasons why it had been difficult to find Turing structures in chemistry was due to property 1 in Sect. 7.2.2, namely that the diffusion coefficients D_1 and D_2 must be different. While it is possible, theoretically, to obtain Turing structures for D_1 and D_2 arbitrarily close to one another [31], for robust patterning D_1 and D_2 have to be quite different from each other and, typically, when chemicals react with each other the chemical molecules have similar sizes and therefore quite similar diffusion coefficients. In the CIMA reaction, however, one of the reactants was bound to starch (added as an indicator) and this changed its diffusion coefficient significantly to move the system into the Turing patterning regime. This was modelled by Lengyel and Epstein [32] in the following way:

$$\frac{\partial u}{\partial t} = D_1 \frac{\partial^2 u}{\partial x^2} + f(u, v) - c_0 u p_+ + c p_-, \tag{7.25}$$

$$\frac{\partial v}{\partial t} = D_2 \frac{\partial^2 v}{\partial x^2} + g(u, v), \tag{7.26}$$

$$\frac{\partial c}{\partial t} = c_0 u p_+ - c p_-. \tag{7.27}$$

Here, u and v are the concentrations of the chemical species and u is assumed to be interacting with the indicator (starch). Assuming that the starch is in excess we can take its concentration to be fixed at c_0. Then, by the Law of Mass Action, the rate at which u binds with starch to create the complex c is $c_0 u p_+$, where p_+ is a rate constant, while the rate at which u is recovered from the complex is cp_-. Assuming the indicator, and therefore the complex, is immobile, we obtain the equation for the complex c.

Adding Eqs. (7.25) and (7.27) we obtain

$$\frac{\partial(u+c)}{\partial t} = D_1 \frac{\partial^2 u}{\partial x^2} + f(u, v). \tag{7.28}$$

The further assumption that the binding between u and the indicator is fast, allows us to replace c in the Eq. (7.27) by Pu, where $P = c_0 p_+ / p_-$, which essentially rescales the diffusion coefficient D_1 by a factor $1/(1 + P)$ (recall that for diffusion-driven instability $D_1 < D_2$ assuming u is the activator and v the inhibitor. Hence, D_1 is effectively lowered).

Turing's original derivation of his reaction-diffusion model was on a discrete array of "cells" or compartments in which reactions took place while the chemicals were transported down chemical gradients to neighbouring compartments. He essentially arrived at a spatially discretised version of the system described in Sect. 7.2.2. Recently, Tompkins et al. [33] actually made a physical model of this set up with compartments in which chemicals reacted and diffused to neighbouring compartments. They showed that this system could produce patterns.

7.4 Conclusions and Discussion

We have shown how spatial patterning can arise from a coupled system of two reaction-diffusion equations and given some examples of applications of the theory of diffusion-driven instability in biology. We have only looked at the basic Turing model but, since Turing's original paper, there have been many extensions made of the model. For example, in a series of papers Nagorcka and colleagues [34, 35] proposed that the initial structures formed by a Turing model could serve as sources or sinks of further Turing models, leading to very complex patterned structures similar to those observed in hair follicles and feather primordia. Kondo and colleagues have carried out extensive experimental studies on fish pattern regeneration, patterns on fish mutants and addressed the issue of how one could link the parameters in a Turing model with more refined genetic information (see the review: [36]).

It is important to point out that there are many other self-organisation models that can produce patterns. For example, in 1983, Oster, Murray and Harris [37, 38] proposed that patterns arose due to cells *mechanically* interacting with each other, leading to spatially heterogeneous patterns of cells themselves, which they then assumed differentiated into structures. It is also known that cells can move in response to gradients in chemicals (*chemotaxis*), and it has been shown that such chemotaxis models can also lead to spatial pattern formation Keller and Segel [39].

Painter et al. [40, 41] showed how a Turing system combined with chemotaxis could lead to patterns of varying wavelengths, consistent with those formed in *Pomacanthus* and generalising the concept of *positional information* [42].

In summary the Turing model has generated a great deal of experimental and mathematical interest, which continues to this day.

Acknowledgements PKM and TEW would like to thank the Mathematical Biosciences Institute (MBI) at Ohio State University, for financially supporting this research through the National Science Foundation grant DMS 1440386. TEW would also like to acknowledge financial support from St John's College, Oxford and BBSRC grant BKNXBKOO BK00.16.

The zebra skin, serval, arrow frog and x-ray pictures from Fig. 7.1 are used under creative commons license 2.0. They were downloaded from flickr.com and owned by David Richards, Vince Smith, The Reptilarium and Rahim Packir Saibo, respectively.

References

1. A.M. Turing, The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B **237**, 37–72 (1952)
2. J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications* (Springer-Verlag, 2003)
3. T.T. Marquez-Lago, P. Padilla, A selection criterion for patterns in reaction–diffusion systems. Theor. Biol. Med. Modell. **11**(1), 7 (2014)
4. P. Grindrod, *The Theory and Applications of Reaction-Diffusion Equations: Patterns and Waves* (Clarendon Press, Oxford, 1996)
5. N.F. Britton, *Reaction-diffusion Equations and Their Applications to Biology* (Academic Press, London, 1986)
6. A. Gierer, H. Meinhardt, A theory of biological pattern formation. Biol. Cybern. **12**(1), 30–39 (1972)
7. J. Schnakenberg, Simple chemical reaction systems with limit cycle behaviour. J. Theor. Biol. **81**(3), 389–400 (1979)
8. D. Thomas, *Analysis and Control of Immobilised Enzyme Systems*. Chapter Artificial Enzyme Membranes, Transport, Memory, and Oscillatory Phenomena (Springer, Berlin, 1975), pp. 115–150
9. R.A. Barrio, C. Varea, J.L. Aragón, P.K. Maini, A two-dimensional numerical study of spatial pattern formation in interacting Turing systems. Bull. Math. Biol. **61**(3), 483–505 (1999)
10. B. Ermentrout, Stripes or spots? Nonlinear effects in bifurcation of reaction-diffusion equations on the square. Proc. Math. Phys. Sci. **434**(1891), 413–417 (1991)
11. T.E. Woolley, R.E. Baker, P.K. Maini, J.L. Aragón, R.A. Barrio, Analysis of stationary droplets in a generic Turing reaction-diffusion system. Phys. Rev. E **82**(5), 051929 (2010)
12. H. Meinhardt, *Models of Biological Pattern Formation* (Academic Press, London, 1982)
13. H. Meinhardt, *The Algorithmic Beauty of Sea Shells* (Springer, Berlin, 2009)
14. J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, vol. 2, 3rd edn. (Springer, Cham, 2003)
15. G.F. Oster, N. Shubin, J.D. Murray, P. Alberch, Evolution and morphogenetic rules: the shape of the vertebrate limb in ontogeny and phylogeny. Evolution **42**(5), pp. 862–884 (1988)
16. J. Bard, I. Lauder, How well does Turing's theory of morphogenesis work? J. Theor. Biol. **45**(2), 501–31 (1974)
17. J.D. Murray, Parameter space for Turing instability in reaction diffusion mechanisms: a comparison of models. J. Theor. Biol. **98** (1), 143 (1982)

18. R. Dillon, P.K. Maini, H.G. Othmer, Pattern formation in generalized Turing systems. J. Math. Biol. **32**(4), 345–393 (1994)
19. E.J. Crampin, E.A. Gaffney, P.K. Maini, Reaction and diffusion on growing domains: scenarios for robust pattern formation. Bull. Math. Biol. **61**(6), 1093–1120 (1999)
20. S. Kondo, R. Asai, A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus. Nature **376**, 765–768 (1995)
21. S.A. Newman, R. Bhat, Activator-inhibitor dynamics of vertebrate limb pattern formation. Birth Defects Res. C Embryo Today **81**(4), 305–319 (2007)
22. S. Sick, S. Reinker, J. Timmer, T. Schlake, WNT and DKK determine hair follicle spacing through a reaction-diffusion mechanism. Science **314**(5804), 1447–1450 (2006)
23. A. Garfinkel, Y. Tintut, D. Petrasek, K. Boström, L.L. Demer, Pattern formation by vascular mesenchymal cells. Proc. Nat. Acad. Sci. **101**(25), 9247 (2004)
24. A. Nakamasu, G. Takahashi, A. Kanbe, S. Kondo, Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. Proc. Nat. Acad. Sci. **106**(21), 8429–8434 (2009)
25. T.E. Woolley, Pattern production through a chiral chasing mechanism. Phys. Rev. E **96**(3), 032401 (2017)
26. T.E. Woolley, P.K. Maini, E.A. Gaffney, Is pigment cell pattern formation in zebrafish a game of cops and robbers? Pigment Cell Melanoma Res. **27**(5), 686–687 (2014)
27. T. Woolley, Pattern production through a chiral chasing mechanism. Phys. Rev. E **96**, 32401 (2017)
28. T.E. Woolley, R.E. Baker, E.A. Gaffney, P.K. Maini, Stochastic reaction and diffusion on growing domains: understanding the breakdown of robust pattern formation. Phys. Rev. E **84**(4), 046216 (2011)
29. T. Kurics, D. Menshykau, D. Iber, Feedback, receptor clustering, and receptor restriction to single cells yield large Turing spaces for ligand-receptor-based Turing models. Phys. Rev. E **90**(2), 022716 (2014)
30. V. Castets, E. Dulos, J. Boissonade, P. De Kepper, Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. Phys. Rev. Lett. **64**(24), 2953–2956 (1990)
31. J.E. Pearson, W. Horsthemke, Turing instabilities with nearly equal diffusion coefficients. J. Chem. Phys. **90**, 1588 (1989)
32. I. Lengyel, I.R. Epstein, A chemical approach to designing Turing patterns in reaction-diffusion systems. Proc. Nat. Acad. Sci. **89**(9), 3977–3979 (1992)
33. N. Tompkins, N. Li, C. Girabawe, M. Heymann, G.B. Ermentrout, I.R. Epstein, S. Fraden, Testing Turing's theory of morphogenesis in chemical cells. Proc. Natl. Acad. Sci. **111**(12), 4397–4402 (2014)
34. B.N. Nagorcka, Wavelike isomorphic prepatterns in development. J. Theor. Biol. **137**(2), 127–162 (1989)
35. J.R. Mooney, B.N. Nagorcka, Spatial patterns produced by a reaction-diffusion system in primary hair follicles. J. Theor. Biol. **115**(2), 299–317 (1985)
36. S. Kondo, M. Iwashita, M. Yamaguchi, How animals get their skin patterns: fish pigment pattern as a live Turing wave. Int. J. Dev. Biol. **53**(5–6), 851 (2009)
37. G.F. Oster, J.D. Murray, A.K. Harris, Mechanical aspects of mesenchymal morphogenesis. Development **78**(1), 83–125 (1983)
38. J.D. Murray, G.F. Oster, A.K. Harris, A mechanical model for mesenchymal morphogenesis. J. Math. Biol. **17**(1), 125–129 (1983)
39. E.F. Keller, L.A. Segel, Initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. **26**(3), 399–415 (1970)
40. K.J. Painter, P.K. Maini, H.G. Othmer, Stripe formation in juvenile Pomacanthus explained by a generalized Turing mechanism with chemotaxis. Proc. Nat. Acad. Sci. **96**(10), 5549 (1999)
41. K.J. Painter, P.K. Maini, H.G. Othmer, Development and applications of a model for cellular response to multiple chemotactic cues. J. Math. Biol. **41**(4), 285–314 (2000)
42. L. Wolpert, Positional information and pattern formation. Curr. Top. Dev. Biol. **6**, 183–224 (1971)

Chapter 8
Persistence, Competition, and Evolution

King-Yeung Lam and Yuan Lou

Abstract In this chapter we discuss some reaction–diffusion models for single and multiple populations in spatially heterogeneous environments and advective environments. Our goal is to illustrate some interesting, and perhaps surprising, effects of spatial heterogeneity and diffusion on the population dynamics. Specific topics include the logistic model, linear eigenvalue problem with indefinite weight, Lotka–Volterra competition models, reaction–diffusion models in advective environments, and the evolution of dispersal. We will introduce some basic tools for reaction–diffusion equations such as the super-sub solution method, the variational principle for principal eigenvalues, Lyapunov functionals, comparison principles for parabolic equations and systems, etc. Some recent developments will be discussed. In addition, problems with various difficulties ranging from elementary exercises to open research questions will be presented.

8.1 Introduction

Understanding the population dynamics of a single and multiple interacting species, which disperse in spatially heterogeneous environments, is an important topic in spatial ecology. Reaction–diffusion models have played a major role in the modeling

K.-Y. Lam
Department of Mathematics, The Ohio State University, Columbus, OH, USA
e-mail: lam.184@math.ohio-state.edu

Y. Lou (✉)
Institute for Mathematical Sciences, Renmin University of China, Beijing, People's Republic of China

Department of Mathematics, The Ohio State University, Columbus, OH, USA
e-mail: lou@math.ohio-state.edu

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_8

and understanding of population dynamics in heterogeneous environments [8]. In this chapter we will restrict ourselves to a few selected topics in spatial ecology and discuss some reaction–diffusion models for the persistence of a single species, the competition of two populations, and the evolution of dispersal in spatially heterogeneous environments and advective environments.

Our first goal is to illustrate some interesting, and perhaps surprising, effects of spatial heterogeneity on the population dynamics. In Sect. 8.2 we will discuss the logistic model, including the derivation of the continuous-time logistic model from the discrete-time counterpart, a linear eigenvalue problem with indefinite weight, and the maximization of the biomass of a single species at equilibrium. In Sect. 8.3 we will discuss the classical two-species Lotka–Volterra competition models, in both homogeneous and heterogeneous environments. Section 8.4 is devoted to the evolution of random dispersal in heterogeneous environments, where it is shown that the slower dispersal rate will be selected. In Sect. 8.5 we will study the persistence of a single species and the competition of two populations in advective environments. We show that the faster dispersal rate could be selected in advective environments.

Another goal of this chapter is to introduce some basic mathematical tools for reaction–diffusion equations and systems. They include the super-solution and sub-solution method, the variational principle for principal eigenvalues, Lyapunov functionals, linear stability analysis, and the comparison principles for parabolic equations and systems. These materials will be covered in Sects. 8.2–8.5.

Beyond our two main goals, in Sect. 8.6 we will discuss some recent works and point interested readers to the related literature. In addition, some mathematical problems with various degrees of difficulties, ranging from elementary exercises to open research questions, will also be presented.

8.2 Diffusion Models for a Single Species

The dynamics of reaction–diffusion models for a single species are not only of independent interest, they are also building blocks in studying the dynamics of multiple interacting species, especially issues concerning the invasions of exotic species. In this section we focus on logistic type population models with diffusion. Many reaction–diffusion models for a single population are of the form

$$\begin{cases} u_t = d\Delta u + uf(x,u) & \text{in } \Omega \times (0,\infty), \\ \nabla u \cdot n = 0 & \text{on } \partial\Omega \times (0,\infty), \\ u(x,0) = u_0(x) & \text{in } \Omega. \end{cases} \tag{8.1}$$

Here $u(x,t)$ is the population density, $d > 0$ is the diffusion coefficient, $f(x,u)$ represents the growth rate of the population and is differentiable in both x and u. The habitat Ω is a bounded domain in Euclidean space $\mathbb{R}^N$ with smooth boundary $\partial\Omega$, and n is the outward unit normal vector on $\partial\Omega$. The zero Neumann boundary condition means that there is no net movement across the boundary. The initial condition u_0 is assumed to be non-negative and not identically zero.

We present two preliminary results on the dynamics of (8.1), adapted from [8].

Proposition 8.1 *Suppose that $f(x,u) \leq g_0(x)$ for some function g which is Hölder continuous in $\bar{\Omega}$. If the principal eigenvalue, denoted as σ_1, of*

$$\begin{cases} d\Delta\psi + g_0\psi + \sigma\psi = 0 \ \mathit{in}\ \Omega, \\ \nabla\psi \cdot n = 0 \qquad\qquad\qquad \mathit{on}\ \partial\Omega \end{cases} \tag{8.2}$$

is positive, then (8.1) has no positive steady state and all non-negative solutions of (8.1) decay exponentially to zero as $t \to \infty$.

Proof Let $\psi(x) > 0$ be an eigenfunction to σ_1. Set $\overline{u}(x,t) = Ce^{-\sigma_1 t}\psi(x)$. Then $\overline{u}$ satisfies

$$\overline{u}_t - d\Delta\overline{u} - \overline{u}f(x,\overline{u}) = g_0(x)\overline{u} - \overline{u}f(x,\overline{u}) \geq 0.$$

Choose $C > 0$ large such that $u(x,0) \leq \overline{u}(x,0)$. By the comparison principle for parabolic equations [48], $u(x,t) \leq \overline{u}(x,t) \leq C\|\psi\|_\infty e^{-\sigma_1 t}$. ■

Proposition 8.2 *Suppose that there exists some $C > 0$ such that $f(x,u) < 0$ for $u \geq C$. If the principal eigenvalue, denoted as σ_1, of*

$$\begin{cases} d\Delta\psi + f(x,0)\psi + \sigma\psi = 0 \ \mathit{in}\ \Omega, \\ \nabla\psi \cdot n = 0 \qquad\qquad\qquad\qquad \mathit{on}\ \partial\Omega \end{cases} \tag{8.3}$$

is negative, then (8.1) has at least one positive steady state.

Proof Consider the steady state problem of (8.1), i.e.,

$$\begin{cases} d\Delta u + uf(x,u) = 0 \text{ in } \Omega, \\ \nabla u \cdot n = 0 \qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{8.4}$$

Write $f(x,u) = f(x,0) + f_1(x,u)$ so that $f_1(x,u) = O(u)$ for small u. Let $\psi > 0$ be an eigenfunction to σ_1. For sufficiently small $\epsilon > 0$,

$$d\Delta(\epsilon\psi) + (\epsilon\psi)f(x,\epsilon\psi) = \epsilon\psi[-\sigma_1 + f_1(x,\epsilon\psi)] > 0,$$

i.e., $\underline{u} = \epsilon\psi$ is a sub-solution of (8.4). Since $\overline{u} = C$ is a super-solution of (8.4) and for small ϵ we have $\overline{u} \geq \underline{u}$, by the super-solution and sub-solution method [48] we see that (8.1) has a positive steady state $u(x)$ such that $\underline{u}(x) \leq u(x) \leq \overline{u}(x)$ for $x \in \Omega$. ■

A classic example of $f(x,u)$ in (8.1) is the logistic growth model:

$$f(x,u) = r(x)\left(1 - \frac{u}{K(x)}\right), \tag{8.5}$$

where $r(x)$, $K(x) \in C(\overline{\Omega})$ are positive functions.

Exercise Consider problem (8.1). If $f(x,u) < g_0(x)$ for $x \in \overline{\Omega}$ and $u > 0$, and that the principal eigenvalue σ_1 of (8.2) is non-negative, then all non-negative solutions of (8.1) decay to zero as $t \to \infty$ (albeit not necessarily exponentially).

8.2.1 Logistic Model: From Discrete to Continuous

In this subsection we present a derivation of the logistic model from discrete-time models. The continuous-time logistic ODE model (Verhulst, 1838) is given by

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}), \quad t > 0. \tag{8.6}$$

Here r and K are two positive constants: r is the intrinsic growth rate (1/time), and K is the carrying capacity (same unit as population size).

One way to derive (8.6) is to start with discrete-time models. Let N_t denote the population size at time $t = 0, 1, 2, \cdots$. The general model is usually of the form $N_{t+1} = f(N_t)$, where f is the growth function.

The geometric growth model is given by $N_{t+1} = RN_t$, where the biological meaning of parameter R can be seen from

$$R = \frac{N_{t+1}}{N_t} = \frac{\text{numbers of offsprings}}{\text{numbers of parents}}. \tag{8.7}$$

For the geometric model, $N_t/N_{t+1} = 1/R$, i.e., the parent vs offspring ratio is constant. The next level of models in terms of modeling complexity is

$$\frac{N_t}{N_{t+1}} = \text{a linear function of } N_t. \tag{8.8}$$

When $N_t \approx 0$ (N_t is rare), we expect the geometric model to be a good approximation, so that $\frac{N_t}{N_{t+1}} = \frac{1}{R}$. When $N_t \approx K$ (N_t is near the carrying capacity), we expect the population to level off, so that $\frac{N_t}{N_{t+1}} = 1$. Hence,

$$\frac{N_t}{N_{t+1}} = \text{the line passing through } (0, \frac{1}{R}) \text{ and } (K, 1) \tag{8.9}$$

$$= \frac{1}{R} + \frac{N_t}{K}\left(1 - \frac{1}{R}\right). \tag{8.10}$$

After simplification, one obtains the Beverton–Holt model

$$N_{t+1} = \frac{RN_t}{1 + \frac{R-1}{K} N_t}, \quad t = 0, 1, 2, \cdots, \tag{8.11}$$

where the constant K is the population size at which the parent vs offspring ratio is equal to one.

In the derivation above, the duration of each generation is 1. Now, if we let the duration of each generation to be some small constant $h > 0$, then over the (short) time interval h, the population multiplies by a factor of R^h, and thus we may modify (8.11) as

$$N_{t+h} = \frac{R^h N_t}{1 + \frac{R^h-1}{K} N_t}, \quad h > 0.$$

We can rewrite the above as

$$\frac{N_{t+h} - N_t}{h} = \frac{1}{h} \left(\frac{R^h N_t}{1+\frac{R^h-1}{K} N_t} - N_t \right) \tag{8.12}$$

$$= \frac{1}{h} \cdot \frac{(R^h-1)N_t - \frac{R^h-1}{K} N_t^2}{1+\frac{R^h-1}{K} N_t} \tag{8.13}$$

$$= \frac{R^h-1}{h} \cdot \frac{(1-\frac{N_t}{K})N_t}{1+\frac{R^h-1}{K} N_t}. \tag{8.14}$$

Letting $h \to 0$, we obtain the continuous-time logistic model, which relates the instantaneous rate of change of population at time t to the population at time t:

$$\frac{d}{dt} N_t = rN_t(1 - \frac{N_t}{K}), \quad t > 0, \tag{8.15}$$

where $r = \log R$.

Exercise Use the discrete-time two-species model

$$\begin{cases} N_1(t+1) = \frac{R_1 N_1(t)}{1+\alpha_1 N_1(t)+\beta_1 N_2(t)} \\ N_2(t+1) = \frac{R_2 N_2(t)}{1+\alpha_2 N_1(t)+\beta_2 N_2(t)} \end{cases} \tag{8.16}$$

to derive the corresponding continuous-time model

$$\begin{cases} \frac{dN_1}{dt} = r_1 N_1(1 - C_1 N_1 - D_1 N_2) \\ \frac{dN_2}{dt} = r_2 N_2(1 - C_2 N_1 - D_2 N_2) \end{cases}. \tag{8.17}$$

8.2.2 Logistic PDE Model

We consider a special case of the reaction–diffusion model (8.1) with logistic nonlinearity (8.5) and heterogeneous coefficients $r(x) = m(x)$ and $K(x) = m(x)$:

$$\begin{cases} u_t = d\Delta u + u(m(x) - u) & \text{in } \Omega \times (0, \infty), \\ \nabla u \cdot n = 0 & \text{on } \partial\Omega \times (0, \infty), \\ u(x, 0) = u_0(x) & \text{in } \Omega. \end{cases} \tag{8.18}$$

Another example is to add spatially heterogeneous harvesting effect to the logistic nonlinearity $ru(1 - \frac{u}{K})$ with constant coefficients r, K, so that the growth function of the population is given by

$$r(1 - \frac{u}{K}) - h(x) = r - h(x) - \frac{r}{K}u;$$

i.e., $m(x) = r - h(x)$ in this example, where $h(x)$ is the harvesting rate.

By Propositions 8.1 and 8.2, the problem of determining the existence and non-existence of positive steady state is connected with the sign of the principal eigenvalue σ_1 (i.e., the unique eigenvalue possessing a positive eigenfunction) of

$$\begin{cases} d\Delta\psi + m(x)\psi + \sigma\psi = 0 \text{ in } \Omega, \\ \nabla\psi \cdot n = 0 \qquad\qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{8.19}$$

By the variational characterization of elliptic eigenvalues, we have

$$\sigma_1 = \inf_{\psi \in H^1(\Omega), \psi \neq 0} \frac{\int_\Omega [d|\nabla\psi|^2 - m(x)\psi^2]\, dx}{\int_\Omega \psi^2\, dx},$$

one can deduce the following result. (See, e.g., Proposition 4.4 of [47].)

Lemma 8.1 *Suppose that m is non-constant. Then σ_1 is a strictly monotone increasing function of d. Moreover,*

$$\lim_{d \to 0} \sigma_1 = -\max_{\bar{\Omega}} m, \tag{8.20}$$

$$\lim_{d \to \infty} \sigma_1 = -\tfrac{1}{|\Omega|} \int_\Omega m. \tag{8.21}$$

Furthermore, the mapping $d \mapsto \sigma_1$ is concave.

Exercise Prove Lemma 8.1.

By Lemma 8.1, we see that if $\int_\Omega m \geq 0$, then $\sigma_1 < 0$ for any $d > 0$. Hence, by Proposition 8.2, (8.18) has at least one positive steady state for any $d > 0$. If $\int_\Omega m < 0$ and $\max_{\bar{\Omega}} m > 0$, by Lemma 8.1 there exists a unique $d^* \in (0, +\infty)$ such that $\sigma_1 < 0$ for $d < d^*$; and $\sigma_1 > 0$ when $d > d^*$. Again by Propositions 8.1 and 8.2

we see that if $d > d^*$, then every non-negative solution of (8.18) converges to zero; if $d < d^*$, then (8.18) has at least one positive steady state. These discussions lead to the following result. (See, e.g., Proposition 3.3 of [8] or Theorem 4.1 of [47].)

Theorem 8.1 *Suppose that m is non-constant, positive somewhere in Ω and Hölder continuous in $\bar{\Omega}$.*

(i) *If $\int_\Omega m \geq 0$, then for $d > 0$, (8.18) has a unique positive steady state and it is globally asymptotically stable among non-negative, not identically zero, continuous initial data.*
(ii) *If $\int_\Omega m < 0$ and $\max_{\bar{\Omega}} m > 0$, then there exists some $d^* > 0$ such that if $d < d^*$, (8.18) has a unique positive steady state which is globally asymptotically stable; if $d > d^*$, all non-negative solutions of (8.18) converge to zero as $t \to \infty$.*

The existence and non-existence results are addressed by the discussion preceding Theorem 8.1. We now sketch the proof of the uniqueness of the positive steady state, whenever it exists. The proof of the uniqueness of positive steady state is based upon the super-solution and sub-solution method. For this purpose, we consider the steady state problem of (8.18):

$$\begin{cases} d\Delta u + u(m(x) - u) = 0 & \text{in } \Omega, \\ \nabla u \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.22}$$

We say that $\overline{u} \in C^2(\bar{\Omega})$ is a super-solution of (8.22) if it satisfies

$$\begin{cases} d\Delta \overline{u} + \overline{u}(m - \overline{u}) \leq 0 \text{ in } \Omega, \\ \nabla \overline{u} \cdot n \geq 0 \qquad\qquad\quad \text{on } \partial\Omega. \end{cases} \tag{8.23}$$

We can similarly define the sub-solution of (8.22) by reversing the inequalities. The following comparison principle is well known. (See, e.g., [50, 51].)

Theorem 8.2 *Suppose that (8.22) has a pair of super-solution and sub-solution such that $\underline{u} \leq \overline{u}$ in Ω. Then (8.18) has a minimal steady state u_m and a maximal steady state u_M, such that (i) $\underline{u} \leq u_m \leq u_M \leq \overline{u}$ in Ω and (ii) for each solution v of (8.22) satisfying $\underline{u} \leq v \leq \overline{u}$ in Ω, then it must hold that $u_m \leq v \leq u_M$ in Ω.*

We proceed to sketch the proof of the uniqueness result, as claimed in Theorem 8.1. Suppose, for contradiction, that there are two distinct positive steady states when $\sigma_1 < 0$, denoted by u_i, $i = 1, 2$. Since Eq. (8.22) has arbitrary large super-solutions (e.g., any constant C larger than the maximum of function m) and arbitrary small positive sub-solutions, e.g., $\epsilon\psi$, where $\epsilon > 0$ is small and $\psi > 0$ is an eigenfunction of (8.19), we can choose ϵ and C such that $\epsilon\psi \leq u_1, u_2 \leq C$. Hence, by Theorem 8.2, there exists a minimal solution and a maximal solution, denoted by u_m and u_M, respectively, satisfying $u_m \leq u_1, u_2 \leq u_M$ in Ω. Since $u_1 \not\equiv u_2$, we have $u_M \geq, \not\equiv u_m$. Multiplying the equation of u_m by u_M, the equation of u_M by u_m, subtracting and integrating the result in Ω, we see that

$$\int_\Omega u_m u_M (u_M - u_m) = 0,$$

which is a contradiction to the fact that $u_M \geq u_m$ and $u_M \not\equiv u_m$. Hence it is impossible for Eq. (8.22) to have two distinct positive solutions u_1, u_2. This proves the uniqueness result.

8.2.3 An Eigenvalue Problem with Indefinite Weight

Recall that by Theorem 8.1, there exists a critical diffusion rate $d^* > 0$ so that the population as modeled by (8.18) persists if and only if $d \in (0, d^*)$. In this subsection, we will give a characterization of $1/d^*$ via the following linear eigenvalue problem with indefinite weight:

$$\begin{cases} \Delta\varphi + \lambda m(x)\varphi = 0 & \text{in } \Omega, \\ \nabla\varphi \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.24}$$

Problem (8.24) and its variants have been extensively investigated for the last two decades, since they play crucial roles in studying nonlinear models from population biology.

We call λ a *principal eigenvalue* of (8.24) if λ has a positive eigenfunction $\varphi \in H^1(\Omega)$. Clearly, $\lambda = 0$ is a principal eigenvalue of (8.24) with positive constants as its eigenfunctions. Of particular importance is the existence of *positive* principal eigenvalues.

If (8.24) has a positive eigenvalue, denoted by $\lambda_1(m)$, with corresponding positive eigenfunction φ_1, integrating the equation of φ_1 we have

$$\int_\Omega m\varphi_1 = 0,$$

which implies that $m(x)$ changes sign in Ω, i.e., that both Ω_+ and Ω_- have positive Lebesgue measure, where

$$\Omega_+ = \{x \in \Omega : m(x) > 0\}, \qquad \Omega_- = \{x \in \Omega : m(x) < 0\}.$$

Dividing the equation of φ_1 by φ_1 and then integrating in Ω, we find

$$\lambda_1(m) \int_\Omega m = \int_\Omega \frac{\Delta\varphi}{\varphi} = -\int_\Omega \frac{|\nabla\varphi_1|^2}{\varphi_1^2} < 0$$

since φ_1 is not equal to any positive constant (as m is not identically equal to any constant). In summary, the condition

(A1) The set Ω_+ has positive Lebesgue measure, and $\int_\Omega m < 0$

is necessary for the existence of a positive principal eigenvalue. This condition turns out to be also sufficient as shown by the following result [3]:

Theorem 8.3 *The eigenvalue problem (8.24) has a positive principal eigenvalue (denoted by $\lambda_1(m)$) if and only if (A1) holds. Moreover, $\lambda_1(m)$ is the only positive principal eigenvalue and it is simple; it is also the smallest positive eigenvalue of (8.24), and is given by*

$$\lambda_1(m) = \inf_{\varphi \in \mathscr{S}(m)} \frac{\int_\Omega |\nabla \varphi|^2}{\int_\Omega m(x)\varphi^2}, \tag{8.25}$$

where

$$\mathscr{S}(m) := \left\{ \varphi \in H^1(\Omega) : \int_\Omega m(x)\varphi^2 > 0 \right\}.$$

By Theorem 8.3, one may observe that d^* in Theorem 8.1 is characterized by $d^* = 1/\lambda_1(m)$. In fact, the following well-known and useful result holds.

Proposition 8.3 *Suppose that* **(A1)** *holds. Let $d^* := 1/\lambda_1(m)$, where $\lambda_1(m)$ is the principal eigenvalue of (8.24). Then $d^* > 0$. Furthermore, let $\sigma_1(d, m)$ be the principal eigenvalue of (8.19). Then*

(i) $\sigma_1(d, m) < 0$ *when* $0 < d < d^*$*;*
(ii) $\sigma_1(d, m) = 0$ *when* $d = d^*$*;*
(iii) $\sigma_1(d, m) > 0$ *when* $d > d^*$*.*

Exercise Prove Proposition 8.3 using the facts that (i) $\sigma_1(d, m)$ is concave in d; and that (ii) $\sigma_1(d, m) = 0$ if and only if $d = d^*$.

Consider the scenario where there is limited total resource in a bounded domain Ω. What is the optimal way to distribute the resource, so as to maximize the survivorship of the population?

Given $\mu \in (0, 1)$ and $\kappa > 0$, we define

$$\mathscr{M} = \left\{ m \in L^\infty(\Omega) : m(x) \text{ satisfies (A1) and (A2)} \right\}, \tag{8.26}$$

where **(A2)** is the constraint on the resource distribution:

(A2) $-1 \le m(x) \le \kappa$ a.e. in Ω, and $\int_\Omega m \le -\mu|\Omega|$.

Roughly speaking, **(A2)** says that the habitat is unfavorable on average. Also, the resource distribution is bounded from above by κ, and below by -1. We aim to determine the optimal arrangement of the resource so as to maximize d^*, which is equivalent to minimizing $\lambda_1(m)$. Therefore, we set

$$\lambda_{inf} := \inf_{m \in \mathscr{M}} \lambda_1(m).$$

The existence and the profile of global minimizers of $\lambda_1(m)$ in $\mathscr{M}$ with Dirichlet boundary condition was first addressed in [4]. For Neumann boundary conditions, we have the following result [40]:

Theorem 8.4 *The infimum λ_{inf} is attained by some $m \in \mathscr{M}$. Moreover, if $\lambda_1(m) = \lambda_{inf}$, then m can be represented as $m(x) = \kappa\chi_E - \chi_{\Omega\setminus E}$ a.e. in Ω for some measurable set $E \subset \Omega$.*

Proof We only prove the second part of Theorem 8.4. Suppose that $m \in \mathscr{M}$ and $\lambda_1(m) = \lambda_{inf}$. Let φ be the eigenfunction of $\lambda_1(m)$ with the normalization $\sup_\Omega \varphi = 1$. For every $\eta \geq 0$, set $E_\eta := \{x \in \Omega : \varphi(x) > \eta\}$. Note that $|E_\eta|$, the Lebesgue measure of E_η, is a monotone decreasing function of η, $|E_0| = |\Omega|$ and $|E_\eta| = 0$ for $\eta > 1$.

Case 1. There exists some $\eta^* > 0$ such that

$$-\mu|\Omega| = \kappa|E_{\eta^*}| - |\Omega \setminus E_{\eta^*}|, \tag{8.27}$$

i.e., $|E_{\eta^*}| = (1-\mu)|\Omega|/(1+\kappa) > 0$. For this case, define $E^* := E_{\eta^*}$.

Case 2. There is no $\eta > 0$ such that $|E_\eta| = (1-\mu)|\Omega|/(1+\kappa)$. For this case, there exists some $\eta^* > 0$ such that $\lim_{\eta\to\eta^*+}|E_\eta| < (1-\mu)|\Omega|/(1+\kappa) \leq \lim_{\eta\to\eta^*-}|E_\eta|$. Therefore, there exists some measurable set E^* such that $E_{\eta^*} \subset E^* \subset \{x \in \Omega : \varphi(x) \geq \eta^*\}$ and $|E^*| = (1-\mu)|\Omega|/(1+\kappa)$.

Define $m^*(x) = \kappa\chi_{E^*} - \chi_{\Omega/E^*}$. Equation (8.27) ensures that $\int_\Omega m^* = -\mu|\Omega|$. Hence, we have $m^* \in \mathscr{M}$, which implies that $\lambda_{inf} \leq \lambda_1(m^*)$.

We claim that $m(x) = m^*(x)$ a.e. in Ω. To establish our assertion, we first have

$$\begin{aligned}\int_\Omega (m^* - m)\varphi^2 &= \int_{E^*}(\kappa - m)\varphi^2 - \int_{\Omega\setminus E^*}(1+m)\varphi^2 \\ &\geq (\eta^*)^2\int_{E^*}(\kappa - m) - (\eta^*)^2\int_{\Omega\setminus E^*}(1+m) \\ &\geq 0,\end{aligned} \tag{8.28}$$

where the last inequality follows from (8.27) and $\int_\Omega m \leq -\mu|\Omega|$. Since $\int_\Omega m\varphi^2 > 0$, we have $\int_\Omega m^*\varphi^2 > 0$. Hence, $\varphi \in \mathscr{S}(m^*)$. Therefore, applying (8.25) we have

$$\lambda_1(m^*) \leq \frac{\int_\Omega |\nabla\varphi|^2}{\int_\Omega m^*\varphi^2} \leq \frac{\int_\Omega |\nabla\varphi|^2}{\int_\Omega m\varphi^2} = \lambda_1(m). \tag{8.29}$$

Since $\lambda_1(m) = \lambda_{inf} \leq \lambda_1(m^*)$, equalities must hold in (8.29). In particular, $\lambda_1(m^*) = \lambda_1(m)$ and

$$\lambda_1(m^*) = \frac{\int_\Omega |\nabla\varphi|^2}{\int_\Omega m^*\varphi^2}.$$

Therefore, from Theorem 8.3 we see that φ is also an eigenfunction of $\lambda_1(m^*)$, and it satisfies

$$\Delta\varphi + \lambda_1(m^*)m^*\varphi = 0 \quad \text{in } \Omega, \qquad \nabla\varphi \cdot n = 0 \quad \text{on } \partial\Omega$$

in $W^{2,q}(\Omega)$ for every $q > 1$. Since $\varphi > 0$ in $\overline{\Omega}$, we have

$$m^* = -\frac{\Delta\varphi}{\lambda_1(m^*)\varphi} = -\frac{\Delta\varphi}{\lambda_1(m)\varphi} = m$$

a.e. in Ω. This completes the proof of Theorem 8.4. ■

Remark 8.1 Theorem 8.4 implies that the global minimizers of $\lambda_1(m)$ in $\mathscr{M}$ are of "bang-bang" type, i.e., when the habitat is unfavorable on average, the survivorship of the population is maximized when conservational effort and resources are concentrated within a protection zone, even when the rest of the habitat is in poor condition. The original proof of Theorem 8.4 in [40] requires the additional assumption that E_η is continuous in η. The modified proof presented here does not make use of this assumption; see also [46]. We refer to [30] for the case when Ω is a rectangular domain in $\mathbb{R}^2$.

8.2.4 Population Size

In this subsection we study the effects of dispersal and spatial heterogeneity of the environment on the total population size of a single species. Such a consideration is not only out of curiosity, but also useful in studying the invasion of species.

Consider the steady state problem of the diffusive logistic model:

$$\begin{cases} d\Delta\theta + \theta\big[m(x) - \theta\big] = 0 & \text{in } \Omega, \\ \theta > 0 & \text{in } \Omega, \\ \nabla\theta \cdot n = 0 & \text{on } \partial\Omega, \end{cases} \tag{8.30}$$

where the diffusion rate d is assumed to be a positive constant, $m(x)$ is the habitat quality at location x, and the function $\theta = \theta(x, d)$ represents the density of the species at location x. For the sake of clarity we posit

(A3) $m(x)$ is non-constant, bounded, and measurable, and $\int_\Omega m(x)\,dx > 0$.

For solutions of (8.30), the following results are well known.

Theorem 8.5 *Suppose that assumption (A3) holds.*

(i) *For $d > 0$, (8.30) has a unique positive solution $\theta(x, d)$ such that $\theta \in W^{2,p}(\Omega)$ for every $p \geq 1$.*
(ii) *As $d \to 0+$, $\theta(x, d) \to m_+(x)$ in $L^p(\Omega)$ for every $p \geq 1$, where $m_+(x) = \sup\{m(x), 0\}$; as $d \to \infty$, $\theta(x, d) \to \frac{1}{|\Omega|}\int_\Omega m(x)\,dx$ in $W^{2,p}(\Omega)$ for every $p \geq 1$.*
(iii) *If $m(x)$ is Hölder continuous in $\overline{\Omega}$, then $\theta \in C^2(\bar{\Omega})$. Moreover, $\theta(x, d) \to m_+(x)$ in $L^\infty(\Omega)$ as $d \to 0$, and $\theta(x, d) \to \frac{1}{|\Omega|}\int_\Omega m(x)\,dx$ in $C^2(\overline{\Omega})$ as $d \to \infty$.*

Proof We only illustrate that if $m \in C(\bar{\Omega})$ and $m > 0$, then $\theta(x, d) \to m(x)$ in $L^\infty(\Omega)$ as $d \to 0$. Given any $\epsilon > 0$, choose $\overline{u} \in C^2(\bar{\Omega})$ such that $\nabla\overline{u} \cdot n = 0$ on $\partial\Omega$, and

$$m + \frac{\epsilon}{2} \leq \overline{u}(x) \leq m + \epsilon \quad \text{for } x \in \bar{\Omega}.$$

Then,

$$d\Delta\overline{u} + \overline{u}(m - \overline{u}) \leq d\Delta\overline{u} + \overline{u}(m - m - \frac{\epsilon}{2}) = d\Delta\overline{u} - \frac{\epsilon}{2}\overline{u} \leq d\Delta\overline{u} - \frac{\epsilon^2}{4} \leq 0$$

in Ω, where the last inequality holds if d is chosen sufficiently small. Hence, $\overline{u}$ is a super-solution of (8.30). Similarly, choose $\underline{u} \in C^2(\bar{\Omega})$ such that $m - \epsilon \leq \underline{u}(x) \leq m - \frac{\epsilon}{2}$ for any $x \in \bar{\Omega}$, and $\nabla\underline{u} \cdot n = 0$ on $\partial\Omega$. One can proceed similarly to show that $\underline{u}$ is a sub-solution for small d. Hence by the super-solution and sub-solution method [48],

$$m - \epsilon \leq \underline{u}(x) \leq \liminf_{d\to 0+} \theta(x, d) \leq \limsup_{d\to 0+} \theta(x, d) \leq \overline{u}(x) \leq m + \epsilon$$

holds in Ω. Finally, the conclusion follows from letting $\epsilon \to 0+$. ■

Exercise If $\Omega = (0, 1)$ and $m_x > 0$ in $[0, 1]$, show that $\theta_x > 0$ in $(0, 1)$.

Exercise Suppose that m is non-constant and $m \in C^1(\bar{\Omega})$. Show that $\int_\Omega |\nabla\theta|^2\,dx$ is a strictly decreasing function of d and for any $d > 0$,

$$\int_\Omega |\nabla\theta|^2\,dx < \int_\Omega |\nabla m|^2\,dx. \tag{8.31}$$

Since $|\nabla\theta|$ measures the steepness of the population density distribution, we may envision $\int_\Omega |\nabla\theta|^2$ as the average steepness of the population distribution. Similarly, $\int_\Omega |\nabla m|^2$ measures the average steepness of the environmental gradient. This result suggests that the population distribution becomes flatter in average if we increase the dispersal rate. In particular, (8.31) shows that the population distribution is always less steep than the environmental gradient, at least in some average sense.

Open Problem Is $\int_\Omega (\theta - \overline{\theta})^2\, dx$ monotone decreasing in d, where $\overline{\theta} = |\Omega|^{-1} \int_\Omega \theta$? Are $\max_{\bar{\Omega}} \theta$ and $\min_{\bar{\Omega}} \theta$ also monotone in d? Is $\|\theta\|_{L^p}$ monotone decreasing in d for large p?

Remark 8.2 For a fixed $m \in C^\alpha(\bar{\Omega})$ where $\alpha \in (0, 1)$, is it true that there exists some positive constant C, which is independent of d, such that $\|\theta(\cdot, d)\|_{C^\alpha(\bar{\Omega})} \le C$? Averill et al. [1] showed that if $m \in C^2(\bar{\Omega})$ and $m \ge 0$ in $\bar{\Omega}$, then $\theta_d \to m$ in $W^{1,2}(\Omega)$.

In view of part (ii) of Theorem 8.5, it is natural to introduce the function

$$F(d) := \begin{cases} \int_\Omega m_+(x)\, dx, & d = 0, \\ \int_\Omega \theta(x, d) dx, & d > 0, \\ \int_\Omega m(x)\, dx, & d = \infty, \end{cases} \tag{8.32}$$

which can be interpreted as the total population size of the species. By assumption (A2) and part (ii) of Theorem 8.5, F is a continuous, positive function in $[0, \infty]$.

If the spatial environment is homogeneous, i.e., $m(x)$ is equal to some positive constant $\overline{m}$, then $\theta(x, d) \equiv \overline{m}$ is the unique positive solution of (8.30) for every $d > 0$. In this case, the total population size of the species is given by $F(d) = |\Omega|\overline{m}$, which is independent of d. However, if the spatial environment is heterogeneous, i.e., $m(x)$ is a non-constant function, the story changes dramatically:

Theorem 8.6 ([37]) *Suppose that assumption (A1) holds.*

(i) $F(d) > F(\infty)$ *for every* $d \in (0, \infty)$*;*
(ii) *If* $m(x) \ge 0$ *in* Ω*, then for* $d \in (0, \infty)$*,* $F(d)$ *satisfies*

$$F(0) = F(\infty) < F(d).$$

Proof Divide the equation of θ_x by θ,

$$d\frac{\Delta\theta}{\theta} + m - \theta = 0.$$

Integrating the above in Ω, we have

$$\int_\Omega \theta - \int_\Omega m = d\int_\Omega \frac{|\nabla\theta|^2}{\theta^2} > 0;$$

i.e., $\int_\Omega \theta > \int_\Omega m$. The rest of the proof follows from the limiting behaviors of θ as $d \to 0$ and $d \to \infty$, as stated in Theorem 8.5. ■

Remark 8.3 Part (i) of Theorem 8.6 implies that spatial heterogeneity increases the population size of species. To make this assertion precise, set $\overline{m} = \int_\Omega m(x)\,dx/|\Omega|$, and write $F = F(d, m)$ instead of $F(d)$ to indicate the dependence of F on the function m. Part (a) implies that $F(d, m) > F(d, \overline{m})$ for every $d > 0$. In other words, given any $d > 0$ and any function g with $\int_\Omega g(x)\,dx = 0$ and $g \not\equiv 0$, we have $F(d, \overline{m} + \lambda g) > F(d, \overline{m})$ for every $\lambda \neq 0$. Hence, with the dispersal rate being fixed, the population size $F(d, \overline{m} + \lambda g)$, as a function of $\lambda \in \mathbb{R}$, attains a strict global minimum at $\lambda = 0$. We refer to [14] and the references therein for more recent developments.

Exercise Compute $F(d, \overline{m} + \tau g)$ for small τ. What conclusion can you draw from it?

Part (ii) of Theorem 8.6 implies that when $m(x)$ is non-negative, the total population size is minimized at $d = 0$ and $d = \infty$, and maximized at some intermediate value d^*. The value of d^* is determined by the habitat Ω and $m(x)$.

It will be of interest to understand the precise shape of $F(d)$ due to its crucial role in the invasion of species. A natural conjecture is that $F(d)$ has a unique local maximum (and thus it must be the global maximum) in $(0, +\infty)$. However, this conjecture is false even for the case when $m(x)$ is a perturbation of positive constants.

Theorem 8.7 ([36]) *There exists a smooth function $g(x)$ with $\int_\Omega g = 0$ such that if $m = 1 + \epsilon g$, then for sufficiently small non-zero constant ϵ, the total population $F(d, m) = \int_\Omega \theta(x, d)\,dx$, as a function of d, has at least two local maxima and one local minimum in* $(0, \infty)$.

An important issue in conservation biology is to determine how resource allocation affects the population dynamics of species. As the population abundance is often a good measurement of conservation effort, it is of interest to know how resource allocation affects the total population size of species.

Assume that m is non-negative and not identically zero. Let $\theta(x)$ denote the unique positive steady state of (8.18). Given any $\delta \in (0, 1)$, define

$$U = \left\{ m \in L^\infty(\Omega) : 0 \le m \le 1, \int_\Omega m(x)\,dx = \delta|\Omega| \right\}$$

and

$$J(m) = \int_\Omega \theta(x)\,dx. \tag{8.33}$$

It is shown in [15] that there exists some $m^* \in U$ such that

$$J(m^*) = \max_{m \in U} J(m).$$

It seems that the shape of the optimal control m^* depends upon the magnitude of parameter δ. For instance, numerical simulations indicate that for rectangular domains, the optimal control m^* is concentrated at one of the corners of the rectangle when δ is small; if δ is large, the optimal control concentrates at a boundary edge of the rectangle. We refer to [15] for further discussions.

Recently Bai et al. [2] proved the following conjecture of W.-M. Ni:

Theorem 8.8 *Suppose that Ω is an interval in $\mathbb{R}^1$. Then*

$$\frac{\int_\Omega \theta(x)\,dx}{\int_\Omega m(x)\,dx} < 3. \tag{8.34}$$

Furthermore, 3 *is the optimal constant.*

Proof Without loss of generality assume that $\Omega = (0, 1)$. For simplicity we only prove (8.34) when $m_x \geq 0$. For this case, we have $\theta_x \geq 0$ (see exercise after Theorem 8.5). Multiplying (8.30), the equation of θ, by θ_x and integrating the result in $(0, x)$ we obtain

$$\frac{d}{2}\theta_x^2(x) = \int_0^x \theta_x\theta(\theta - m)\,dx \leq \int_0^x \theta_x\theta^2 = \frac{1}{3}[\theta^3(x) - \theta^3(0)] < \frac{1}{3}\theta^3(x).$$

Hence,

$$d\theta_x^2(x) < \frac{2}{3}\theta^3(x), \qquad x \in [0, 1]. \tag{8.35}$$

Next, dividing the equation of θ by θ we have

$$d\frac{\theta_{xx}}{\theta} + m - \theta = 0.$$

Integrating the above equation in $(0, 1)$ we have

$$\int_0^1 \theta\,dx - \int_0^1 m\,dx = d\int_0^1 \frac{\theta_x^2}{\theta^2}\,dx < \frac{2}{3}\int_0^1 \theta\,dx,$$

where we applied (8.35) in the last inequality. Hence, (8.34) holds. ■

Open Question (W.-M. Ni) Show that there exists some positive constant $C = C(N) > 1$, which only depends on the spatial dimension $N \geq 2$, such that for any positive solution θ,

$$\int_\Omega \theta(x)\,dx < C(N)\int_\Omega m(x)\,dx. \tag{8.36}$$

Exercise Let (u_1, u_2) be a positive solution of the two patch model

$$\begin{cases} d(u_2 - u_1) + u_1(m_1 - u_1) = 0, \\ d(u_1 - u_2) + u_2(m_2 - u_2) = 0, \end{cases} \tag{8.37}$$

where $m_1, m_2 > 0$. Show that

$$1 \leq \frac{u_1 + u_2}{m_1 + m_2} < 2.$$

8.3 Lotka–Volterra Competition Models

For the last two decades there has been tremendous interest, from both mathematicians and ecologists, in two-species Lotka–Volterra competition models in spatially heterogeneous environments; see [5–7, 9–11, 21–25, 27–29, 34, 47] and the references therein. Our main goal here is to illustrate some differences between the dynamics of Lotka–Volterra competition models in homogeneous environments and that in heterogeneous environments.

8.3.1 Homogeneous Environments

We first consider the Lotka–Volterra competition–diffusion system in homogeneous environments:

$$\begin{cases} u_t = d_1 \Delta u + u(a_1 - b_1 u - c_1 v) & \text{in } \Omega \times (0, \infty), \\ v_t = d_2 \Delta v + v(a_2 - b_2 u - c_2 v) & \text{in } \Omega \times (0, \infty), \\ \nabla u \cdot n = \nabla v \cdot n = 0 & \text{on } \partial\Omega \times (0, \infty), \\ u(x, 0) = u_0(x), \;\; v(x, 0) = v_0(x) & \text{in } \Omega. \end{cases} \tag{8.38}$$

Here u, v represent the population densities of two competing species; d_1, d_2 are their diffusion rates; a_1 and a_2 are their intrinsic growth rates; b_1 and c_2 are the intra-specific competition coefficients and b_2, c_1 are the inter-specific competition coefficients. All constants are assumed to be positive, and $u_0(x)$, $v_0(x)$ are non-negative functions that are not identically equal to zero.

Under the assumption

$$\frac{b_1}{b_2} > \frac{a_1}{a_2} > \frac{c_1}{c_2}, \tag{8.39}$$

(8.38) has a unique positive steady state, given by

$$(u^*, v^*) = \left(\frac{a_1c_2 - a_2c_1}{b_1c_2 - b_2c_1}, \frac{b_1a_2 - b_2a_1}{b_1c_2 - b_2c_1} \right). \tag{8.40}$$

It turns out that (u^*, v^*) is globally asymptotically stable:

Theorem 8.9 *Suppose that (8.39) holds. Then for any non-negative and not identically zero initial data $u(x, 0), v(x, 0) \in C(\bar{\Omega})$,*

$$\lim_{t\to\infty} (u(x, t), v(x, t)) = (u^*, v^*)$$

in $C(\bar{\Omega}) \times C(\bar{\Omega})$ norm.

Proof Consider the following system of ordinary differential equations:

$$\begin{cases} U_t = U(a_1 - b_1U - c_1V) \quad \text{in } (0, \infty), \\ V_t = V(a_2 - b_2U - c_2V) \quad \text{in } (0, \infty), \\ U(0) > 0, \quad V(0) > 0. \end{cases} \tag{8.41}$$

We claim that for any initial data $U(0) > 0$, $V(0) > 0$,

$$\lim_{t\to\infty} (U(t), V(t)) = (u^*, v^*). \tag{8.42}$$

To establish our assertion, define

$$E(t) = b_2 \left(U - u^* - u^* \ln \frac{U}{u^*} \right) + c_1 \left(V - v^* - v^* \ln \frac{V}{v^*} \right). \tag{8.43}$$

Then $dE/dt \le 0$ and $dE/dt = 0$ if and only if $(U, V) = (u^*, v^*)$. Since $E(t)$ is also bounded from below, by the LaSalle's invariance principle, (8.42) holds.

By the maximum principle, we have $u(x, t), v(x, t) > 0$ for any $x \in \bar{\Omega}$ and $t > 0$. Without loss of generality we assume that $u(x, 0) > 0$ and $v(x, 0) > 0$ in $\bar{\Omega}$.

Let $(\underline{U}, \overline{V})$ be the solution of

$$\begin{cases} \underline{U}_t = \underline{U}(a_1 - b_1\underline{U} - c_1\overline{V}) \quad \text{in } (0, \infty), \\ \overline{V}_t = \overline{V}(a_2 - b_2\underline{U} - c_2\overline{V}) \quad \text{in } (0, \infty), \\ \underline{U}(0) = \min_{x\in\bar{\Omega}} u(x, 0) > 0, \\ \overline{V}(0) = \max_{x\in\bar{\Omega}} v(x, 0) > 0; \end{cases} \tag{8.44}$$

and let $(\overline{U}, \underline{V})$ be the solution of

$$\begin{cases} \overline{U}_t = \overline{U}(a_1 - b_1\overline{U} - c_1\underline{V}) & \text{in } (0,\infty), \\ \underline{V}_t = \underline{V}(a_2 - b_2\overline{U} - c_2\underline{V}) & \text{in } (0,\infty), \\ \overline{U}(0) = \max_{x\in\bar{\Omega}} u(x,0) > 0, \\ \underline{V}(0) = \min_{x\in\bar{\Omega}} v(x,0) > 0. \end{cases} \tag{8.45}$$

Note that $(\underline{U}, \overline{V})$ and $(\overline{U}, \underline{V})$ satisfy (8.38). Since

$$\underline{U}(0) \le u(x,0) \le \overline{U}(0), \quad \underline{V}(0) \le v(x,0) \le \overline{V}(0)$$

by the comparison principle for two-species competition model (8.38) [50],

$$\underline{U}(t) \le u(x,t) \le \overline{U}(t), \quad \underline{V}(t) \le v(x,t) \le \overline{V}(t)$$

hold for all $x \in \bar{\Omega}$ and $t \ge 0$.

By (8.41) and (8.42),

$$\lim_{t\to\infty} (\underline{U}(t), \overline{V}(t)) = \lim_{t\to\infty} (\overline{U}(t), \underline{V}(t)) = (u^*, v^*). \tag{8.46}$$

Therefore, $(u(x,t), v(x,t)) \to (u^*, v^*)$ uniformly in x as $t \to \infty$. ■

8.3.2 *Competition in Heterogeneous Environment*

The semilinear parabolic system

$$\begin{cases} u_t = d_1\Delta u + u[m(x) - u - bv] & \text{in } \Omega \times (0,\infty), \\ v_t = d_2\Delta v + v[m(x) - cu - v] & \text{in } \Omega \times (0,\infty), \\ \nabla u \cdot n = \nabla v \cdot n = 0 & \text{on } \partial\Omega \times (0,\infty), \\ u(x,0) = u_0(x),\ v(x,0) = v_0(x) & \text{in } \Omega \end{cases} \tag{8.47}$$

models two species that are competing for the same resources, where $u(x,t)$ and $v(x,t)$ represent the population densities of two competing species with respective dispersal rates d_1 and d_2, the function $m(x)$ represents their common intrinsic growth rate, and b and c are inter-specific competition coefficients. We shall assume that d_1, d_2, b, and c are positive constants, and $u_0(x)$, $v_0(x)$ are non-negative functions that are not identically equal to zero.

We say that a steady state of (8.47) is a *coexistence state* if both components are positive, and it is a *semi-trivial state* if one component is positive and the other is zero. Under (A3), (8.47) has exactly two semi-trivial states, denoted by $(\theta_{d_1}, 0)$ and $(0, \theta_{d_2})$, where $\theta_d = \theta(\cdot, d)$ is the unique positive solution of (8.30).

We assume that $0 < b, c < 1$. If $m(x) \equiv \overline{m}$ for some positive constant $\overline{m}$, by Theorem 8.9, every solution (u, v) of (8.47) converges to $(\frac{1-b}{1-bc}\overline{m}, \frac{1-c}{1-bc}\overline{m})$ for all diffusion rates d_1, d_2 and any initial data. However, the dynamics of (8.47) is less transparent when m is non-constant. To this end, we start by studying the stability of the semi-trivial steady state $(\theta_{d_1}, 0)$ of (8.47). For the rest of this subsection, we focus on the case $0 < c < 1$.

Theorem 8.10 ([37]) *If (A3) holds and $m(x)$ is non-negative, then there exists some constant $c_* = c_*(m, \Omega) \in (0, 1)$ such that the followings hold:*

(a) *For $c \in (0, c_*)$, $(\theta_{d_1}, 0)$ is unstable for any $d_1, d_2 > 0$.*
(b) *For $c \in (c_*, 1)$, there exists $d_* = d_*(c, m, \Omega) > 0$ such that (i) for $d_2 \in (0, d_*)$, $(\theta_{d_1}, 0)$ is unstable for any $d_1 > 0$; (ii) for $d_2 > d_*$, $(\theta_{d_1}, 0)$ changes stability at least twice as d_1 increases from 0 to d_2.*

Note that the above theorem holds regardless of the specific value $b > 0$. The most interesting case is where $c_* < c < 1$ and $d_2 > d_*$, where we have the following implications:

(i) If $b > 1$, it is well known that without dispersal, species v always drives species u to extinction. However, with dispersal, for some ranges of dispersal rates, species v may fail to invade when rare.
(ii) If $b < 1$, it is well known that, without dispersal, species u and v always coexist. Surprisingly, for certain dispersal rates, species u is able to drive species v to extinction for arbitrary initial conditions. (See Theorem 1.9 of [37].)

Proof We sketch the main ideas in the proof of Theorem 8.10. The stability of $(\theta_{d_1}, 0)$ is determined by the sign of the smallest eigenvalue, denoted by λ_1, of the problem

$$\begin{cases} d_2\Delta\varphi + (m - c\theta_{d_1})\varphi + \lambda\varphi = 0 \text{ in } \Omega, \\ \nabla\varphi \cdot n = 0 \qquad\qquad\qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{8.48}$$

Note that $\lambda_1 = \sigma_1(d_2, m - c\theta_{d_1})$. More specifically, $(\theta_{d_1}, 0)$ is stable if $\lambda_1 > 0$ and unstable of $\lambda_1 < 0$. To determine the sign of λ_1, we observe that λ_1 is a strictly increasing function of d_2, and that

$$\lim_{d_2 \to 0} \lambda_1 = \min_{\bar{\Omega}}(c\theta_{d_1} - m) \leq \min_{\bar{\Omega}}(\theta_{d_1} - m) < 0; \tag{8.49}$$

$$\lim_{d_2 \to +\infty} \lambda_1 = \frac{\int_\Omega \theta_{d_1}}{|\Omega|}\left(c - \frac{\int_\Omega m}{\int_\Omega \theta_{d_1}}\right). \tag{8.50}$$

Set

$$c_* = \inf_{d_1 > 0} \frac{\int_\Omega m}{\int_\Omega \theta_{d_1}}.$$

By Theorem 8.6, we see that $c_* \in (0, 1)$.

For every $c \in (0, c_*)$ and any $d_1 > 0$, $\lim_{d_2 \to +\infty} \lambda_1 \le 0$. Since λ_1 is strictly increasing in d_2, we see that $\lambda_1 < 0$ for any $d_1, d_2 > 0$. This proves part (a).

For every $c \in (c_*, 1)$, for simplicity assume that there exist two positive constants $\underline{d} < \overline{d}$ such that $c - \int_\Omega m / \int_\Omega \theta_{d_1} > 0$ for $d_1 \in (\underline{d}, \overline{d})$, and $c - \int_\Omega m / \int_\Omega \theta_{d_1} < 0$ for $d_1 \in (0, \underline{d}) \cup (\overline{d}, +\infty)$. Define $d^* = d^*(d_1) := 1/\lambda_1(m - c\theta_{d_1})$, i.e.,

$$d^* = \sup_{\varphi \in \mathscr{S}} \frac{\int_\Omega (m - c\theta_{d_1})\varphi^2}{\int_\Omega |\nabla \varphi|^2},$$

where

$$\mathscr{S} = \{\varphi \in H^1(\Omega) : \int_\Omega (m - c\theta_{d_1})\varphi^2 > 0\}.$$

Now, $d^* = +\infty$ if and only if $c - \int_\Omega m / \int_\Omega \theta_{d_1} \le 0$, i.e., $d_1 \in (0, \underline{d}] \cup [\overline{d}, +\infty)$. In particular, $d^*(d_1)$ is finite when $d_1 \in (\underline{d}, \overline{d})$, and that $d^*(d_1) \to +\infty$ as $d_1 \to \underline{d}-$ or $d_1 \to \overline{d}+$. This allows us to define $d_* = \inf_{d_1 > 0} d^*(d_1)$. For $d_2 < d_*$, we have $d_2 < d^*(d_1)$ for all $d_1 > 0$. In this case, Proposition 8.3(i) says that $\lambda_1 < 0$ for all $d_1 > 0$, which implies that $(\theta_{d_1}, 0)$ is unstable for any $d_1 > 0$ and $d_2 < d_*$. For $d_2 > d_*$, we likewise have $\lambda_1 < 0$ for $d_1 \in (0, \underline{d}) \cup (\overline{d}, +\infty)$; and $\lambda_1 > 0$ in some sub-interval of $(\underline{d}, \overline{d})$. Therefore λ_1 changes sign at least twice as d_1 increases from 0 to d_2, i.e., part (b) is proved. ■

Exercise Prove that $\lambda_1 < 0$ whenever $c < 1$ and $d_1 \ge d_2$. [Hint: Observe that λ_1 is monotone increasing in c as well as in d_2, and that $\lambda_1 = 0$ when $c = 1$ and when d_2 is increased to d_1.]

For every $c > 0$, define

$$\Sigma_c = \left\{(d_1, d_2) \in \mathbb{R}^+ \times \mathbb{R}^+ : (\theta_{d_1}, 0) \text{ is linearly stable}\right\}. \tag{8.51}$$

Note that $\Sigma_c \subset \{(d_1, d_2) \in \mathbb{R}^+ \times \mathbb{R}^+ : d_1 < d_2\}$ since, by the comparison principle for principal eigenvalues, $\lambda_1 < 0$ for $d_1 \ge d_2$. Clearly, Σ_c is non-empty if and only if $c > c_*$.

In a series of important works [21–24], He and Ni classified the dynamics of a class of Lotka–Volterra competition–diffusion models which include system (8.47) as a special case. One of their results can be stated as follows:

Theorem 8.11 ([24]) *If assumption (A3) holds and $m(x)$ is non-negative, $c \in (c_*, 1)$ and $0 < b \le 1$, then $(\theta_{d_1}, 0)$ is globally asymptotically stable for any $(d_1, d_2) \in \overline{\Sigma_c}$; if $d_2 \ge d_1$ or $(d_1, d_2) \notin \overline{\Sigma_c}$, then system (8.47) has a unique positive steady state which is globally asymptotically stable.*

One key ingredient in the proof of Theorem 8.11 is the following lemma:

Lemma 8.2 *If* $bc \le 1$*, then any positive steady state of (8.47), if exists, is linearly stable.*

Proof Let (u, v) be any positive steady state of (8.47). The linear stability of (u, v) is determined by the sign of the principal eigenvalue, denoted by λ_1, of the problem

$$\begin{cases} d_1\Delta\varphi + \varphi(m - 2u - bv) - bu\psi + \lambda_1\varphi = 0 & \text{in } \Omega, \\ d_2\Delta\psi - cv\varphi + \psi(m - cu - 2v) + \lambda_1\psi = 0 & \text{in } \Omega, \\ \nabla\varphi \cdot n = \nabla\psi \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.52}$$

As φ, ψ are eigenfunctions of λ_1 and thus do not change sign in Ω, we may assume without loss that $\varphi > 0$ and $\psi < 0$ in $\bar{\Omega}$. Set $\varphi = uw$ and $\psi = -vz$. Thus $w, z > 0$ in $\bar{\Omega}$ and they satisfy

$$\begin{cases} d_1\nabla(u^2\nabla w) - u^3w + bu^2vz + \lambda_1\, u^2w = 0 & \text{in } \Omega, \\ d_2\nabla(v^2\nabla z) + cuv^2w - v^3z + \lambda_1 v^2z = 0 & \text{in } \Omega, \\ \nabla w \cdot n = \nabla z \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.53}$$

Multiplying the equation of w by w^2 and integrating the result in Ω, we have

$$-2d_1\int_\Omega u^2w|\nabla w|^2 - \int_\Omega (uw)^3 + b\int_\Omega (uw)^2(vz) + \lambda_1\int_\Omega u^2w^3 = 0.$$

If $\lambda_1 \le 0$, then we have

$$\int_\Omega (uw)^3 < b\int_\Omega (uw)^2(vz).$$

By Hölder inequality,

$$\int_\Omega (uw)^3 < b\left[\int_\Omega (uw)^3\right]^{2/3}\left[\int_\Omega (vz)^3\right]^{1/3},$$

which implies that

$$\int_\Omega (uw)^3 < b^3\int_\Omega (vz)^3. \tag{8.54}$$

Similarly, if $\lambda_1 \le 0$, by the equation of v and similar argument we have

$$\int_\Omega (vz)^3 < c^3\int_\Omega (uw)^3, \tag{8.55}$$

Clearly, (8.54) and (8.55) are in contradiction with $bc \le 1$. Hence, $\lambda_1 > 0$. This completes the proof. ■

By Theorem 8.11, the parameter region where species u wins is characterized by the closure of the set Σ_c. By a previous exercise, we have seen that, $\Sigma_c \subset \{d_1 \le d_2\}$, i.e., species u may exclude species v only if u is the slower diffuser. Furthermore, by Theorem 8.10, the set Σ_c is non-empty for every $c \in (c_*, 1)$. It is not difficult to see that $\Sigma_{c_1} \subset \Sigma_{c_2}$ for any $c_1 < c_2$ with $c_1, c_2 \in (c_*, 1)$. In fact, the set Σ_c converges to the set $\{(d_1, d_2) : 0 < d_1 < d_2\}$ as $c \to 1-$, and this gives another perspective upon why the slower diffuser wins the competition for the case when $b = c = 1$. We refer to the next section for more details on the evolution of slow dispersal.

8.4 Evolution of Dispersal

It is an important question in spatial ecology to understand which patterns of dispersal can confer some selective or evolutionary advantage. Unconditional dispersal refers to movement which does not depend on habitat quality or population density. For the evolution of unconditional dispersal in the context of reaction–diffusion models, it was shown that slower dispersal rate is selected when the environment is spatially heterogeneous but temporally constant; see [16, 20]. In contrast, for unconditional dispersal in spatially and temporally varying environments faster dispersal rates may be selected in diffusion models [26]. In this section we focus on the evolution of unconditional dispersal in spatially varying but temporally constant environments.

Consider system (8.47) for the case when $b = c = 1$:

$$\begin{cases} u_t = d_1 \Delta u + u[m(x) - u - v] & \text{in } \Omega \times (0, \infty), \\ v_t = d_2 \Delta v + v[m(x) - u - v] & \text{in } \Omega \times (0, \infty), \\ \nabla u \cdot n = \nabla v \cdot n = 0 & \text{on } \partial\Omega \times (0, \infty), \\ u(x, 0) = u_0(x), \;\; v(x, 0) = v_0(x) & \text{in } \Omega. \end{cases} \tag{8.56}$$

The following result was established in [16]:

Theorem 8.12 *Suppose that (A3) holds. If* $0 < d_1 < d_2$*, then the semi-trivial steady state* $(\theta_{d_1}, 0)$ *of (8.56) is globally asymptotically stable among all solutions with non-negative and non-trivial initial data.*

Theorem 8.12 is surprising: when $d_1 = d_2 = 0$, two species will coexist since they are identical. However, if the diffusion rates are positive for both species, the slower diffuser always outcompetes the faster one. This also shows that the PDE dynamics cannot be predicted by the ODE dynamics in this case.

Proof We first prove the instability of $(0, \theta_{d_2})$, which is determined by the sign of the smallest eigenvalue, denoted by $\lambda_1 := \lambda_1(d_1, d_2)$, of

$$\begin{cases} d_1 \Delta\varphi + (m - \theta_{d_2})\varphi + \lambda_1 \varphi = 0 & \text{in } \Omega, \\ \nabla\varphi \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.57}$$

Note that $\lambda_1(d_1, d_2) = \sigma_1(d_1, m - \theta_{d_2})$ and hence it is monotone increasing in d_1. We may normalize φ such that $\varphi > 0$ and $\int_\Omega \varphi^2 = 1$. Denote $\varphi' = \frac{\partial \varphi}{\partial d_1}$ and $\lambda_1' = \frac{\partial \lambda_1}{\partial d_1}$. Differentiating the equation of φ with respect to d_1, multiplying the result by φ and integrating in Ω, we have

$$\int_\Omega \varphi[\Delta\varphi + d_1\Delta\varphi' + (m - \theta_{d_2})\varphi' + \lambda_1'\varphi + \lambda_1\varphi'] = 0,$$

from which we obtain $\lambda_1' = \int_\Omega |\nabla\varphi|^2 > 0$. Hence, λ_1 is strictly increasing in d_1. Since $\lambda_1(d_2, d_2) = 0$ (where $\varphi = \theta_{d_2}/\|\theta_{d_2}\|_{L^2(\Omega)}$), we see that $\lambda_1 < 0$ if and only if $d_1 < d_2$.

Next, we claim that

$$\limsup_{t\to\infty} v(x,t) \le \theta_{d_2}(x). \tag{8.58}$$

To establish our assertion, note that

$$v_t = d_2\Delta v + v(m(x) - u - v) \le d_2\Delta v + v(m(x) - v).$$

Consider

$$\begin{cases} V_t = d_2\Delta V + V(m - V) & \text{in } \Omega \times (0,\infty),\\ \nabla V \cdot n = 0 & \text{on } \partial\Omega \times (0,\infty),\\ V(x,0) = v(x,0) & \text{in } \bar{\Omega}. \end{cases} \tag{8.59}$$

By the comparison principle of parabolic equations [48], $v(x,t) \le V(x,t)$. Thus

$$\limsup_{t\to\infty} v(x,t) \le \limsup_{t\to\infty} V(x,t) = \theta_{d_2}(x).$$

Therefore, for each $\varepsilon > 0$, there exists $T_1 := T_1(\varepsilon)$ such that for $t \ge T_1$ and $x \in \bar{\Omega}$,

$$v(x,t) \le (1+\varepsilon)\theta_{d_2}(x).$$

Consider next the solution $(U(x,t), V(x,t))$ of

$$\begin{cases} U_t = d_1\Delta U + U(m(x) - U - V) & \text{in } \Omega \times [T_1,\infty),\\ V_t = d_2\Delta V + V(m(x) - U - V) & \text{in } \Omega \times [T_1,\infty),\\ \nabla U \cdot n = \nabla V \cdot n = 0 & \text{on } \partial\Omega \times [T_1,\infty),\\ U(x,T_1) = \delta\varphi,\ V(x,T_1) = (1+\varepsilon)\theta_{d_2} & \text{in } \bar{\Omega}. \end{cases} \tag{8.60}$$

We check that $(\delta\varphi, (1+\varepsilon)\theta_{d_2})$ is a pair of sub-super solution of (8.60) as follows:

$$\begin{aligned} &d_2\Delta[(1+\varepsilon)\theta_{d_2}] + (1+\varepsilon)\theta_{d_2}(m - \delta\varphi - (1+\varepsilon)\theta_{d_2})\\ &= (1+\varepsilon)[d_2\Delta\theta_{d_2} + \theta_{d_2}(m - \theta_{d_2}) - (\delta\varphi + \varepsilon\theta_{d_2})\theta_{d_2}] \le 0 \end{aligned} \tag{8.61}$$

and

$$\begin{aligned} & d_1\Delta(\delta\varphi) + \delta\varphi\left[m - \delta\varphi - (1+\varepsilon)\theta_{d_2}\right] \\ & = \delta\left[d_1\Delta\varphi + \varphi(m - \theta_{d_2}) + \varphi(-\delta\varphi - \varepsilon\theta_{d_2})\right] \\ & = \delta\varphi\left[-\lambda_1 - \delta\varphi - \varepsilon\theta_{d_2}\right] \geq 0, \end{aligned} \tag{8.62}$$

since $\lambda_1 < 0$ and δ, ε are chosen small.

By the comparison principle for two-species competitive systems, we see that $U(x,t)$ is increasing in t and $V(x,t)$ is decreasing in t. Therefore $(U(x,t), V(x,t))$ converges, as $t \to \infty$, to some limit $(U^*(x), V^*(x))$. By the elliptic regularity theory we can show that $(U^*(x), V^*(x))$ is a non-negative steady state of (8.56), with $U^* > 0$.

We claim that $V^* \equiv 0$. If not, then (U^*, V^*) is a positive steady state of (8.56), i.e., they satisfy

$$\begin{cases} d_1\Delta U^* + U^*(m(x) - U^* - V^*) = 0 & \text{in } \Omega, \\ d_2\Delta V^* + V^*(m(x) - U^* - V^*) = 0 & \text{in } \Omega, \\ \nabla U^* \cdot n = \nabla V^* \cdot n = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.63}$$

Consider the smallest eigenvalue, denoted by $\lambda_1(d)$, of the eigenvalue problem

$$d\Delta\varphi + (m - U^* - V^*) + \lambda\varphi = 0 \quad \text{in } \Omega, \quad \nabla\varphi \cdot n = 0 \quad \text{in } \partial\Omega.$$

Since m is non-constant, one can show that $m - U^* - V^*$ is also non-constant. Hence, $\lambda_1(d)$ is strictly increasing in d. By the equation of U^*, we see that $\lambda_1(d_1) = 0$ with corresponding $\varphi = U^*$. Similarly from the equation of V^* we get $\lambda_1(d_2) = 0$, which is a contradiction, since $d_1 \neq d_2$.

Hence, $V^* = 0$ and $U^* = \theta_{d_1}$, i.e., $\lim_{t\to\infty}(U(x,t), V(x,t)) = (\theta_{d_1}, 0)$. Choose δ, ϵ small such that $U(x, T_1) = \delta\varphi \leq u(x, T_1)$ and $v(x, T_1) \leq (1+\varepsilon)\theta_{d_2} = V(x, T_1)$. By the comparison principle for two-species competition systems, we have $U(x,t) \leq u(x,t)$ and $v(x,t) \leq V(x,t)$. In particular, $v(x,t) \to 0$ as $t \to \infty$ and $\liminf_{t\to\infty} u(x,t) \geq \theta_{d_1}(x)$. Since, by repeating the previous argument for (8.58), one can also show that $\limsup_{t\to\infty} u(x,t) \leq \theta_{d_1}(x)$, we have $\lim_{t\to\infty} u(x,t) = \theta_{d_1}$. This completes the proof. ■

Consider k-species competition model

$$\begin{cases} u_{i,t} = d_i\Delta u_i + u_i(m - \sum_{i=1}^k u_i) & \text{in } \Omega \times (0,\infty), \\ \nabla u_i \cdot n = 0 & \text{on } \partial\Omega \times (0,\infty). \end{cases} \tag{8.64}$$

A challenging open problem is whether the slowest diffuser still wins the competition in the context of k competing species with $k \geq 3$.

Open Problem Suppose that m is positive, non-constant, and continuous in $\bar{\Omega}$. If $0 < d_1 < d_2 < \ldots < d_k$ and $k \geq 3$, is $(\theta_{d_1}, 0, \ldots, 0)$ globally asymptotically stable among all positive initial data?

The mathematical difficulty in solving this open problem is that competition models for three or more species are not monotone dynamical systems.

8.5 Persistence and Competition in Advective Environments

In this section we consider the persistence of a single species and the competition of two populations in advective environments. We will focus on the effects of advection and boundary conditions on the persistence and competition of populations.

8.5.1 Single Species in Advective Environment

How can populations persist in streams when they are constantly washed downstream? This question, termed as the "drift paradox" in the literatures, has received considerable attention. Speirs and Gurney [52] offered an explanation based upon the diffusive movement of organisms, and they considered the following reaction–diffusion model:

$$\begin{cases} u_t = du_{xx} - qu_x + u(r-u), & \text{for } 0 < x < L,\ t > 0, \\ du_x(0,t) - qu(0,t) = 0, & \text{for } t > 0, \\ u(L,t) = 0, & \text{for } t > 0, \\ u(x,0) = u_0(x), & \text{for } 0 < x < L, \end{cases} \tag{8.65}$$

where $u(x,t)$ denotes the population density at location x and time t, d is the diffusion rate, L is the size of the habitat, and in the sequel, we call $x = 0$ the upstream end and $x = L$ the downstream end. The constant q is the effective speed of the current (sometimes we also call q the advection speed/rate, and we remark here that q is positive since $x = L$ is defined to be the downstream end). The constant $r > 0$ accounts for the intrinsic growth rate, which indicates the spatial homogeneity of the environment. We assume that u_0 is non-negative and not identically zero, and $d,\ r,\ q,\ L$ are all positive constants. In other words, the spatial heterogeneity of the problem (8.65) is introduced solely by the drift and the boundary conditions.

Speirs and Gurney [52] studied the local stability of steady state $u = 0$ and concluded that it is unstable if and only if $q < \sqrt{4dr}$ and $L > L^*$, where

$$L^* = 2d \frac{\pi - \arctan\left(\frac{\sqrt{4dr-q^2}}{q}\right)}{\sqrt{4dr-q^2}}.$$

That is, the persistence is only possible when advection is slow relative to diffusion and the stream is long enough. It is natural to inquire whether such predictions still hold for other situations. To this end, Vasilyeva and Lutscher [53] considered the following single species problem with a different boundary condition at the downstream end $x = L$. Their model is given by

$$\begin{cases} u_t = du_{xx} - qu_x + u(r-u), & \text{for } 0 < x < L,\ t > 0, \\ du_x(0,t) - qu(0,t) = 0, & \text{for } t > 0, \\ u_x(L,t) = 0, & \text{for } t > 0, \\ u(x,0) = u_0(x), & \text{for } 0 < x < L. \end{cases} \tag{8.66}$$

The following result, which is similar in nature to the result of Speirs and Gurney, was proved in [53]:

Theorem 8.13 *The species can persist if and only if $q < \sqrt{4dr}$ and $L > L^{**}$, where*

$$L^{**} = \begin{cases} 2d\dfrac{\arctan\frac{q\sqrt{4dr-q^2}}{2rd-q^2}}{\sqrt{4dr-q^2}} & \text{for } 0 < q \le \sqrt{2dr}, \\ 2d\dfrac{\pi+\arctan\frac{q\sqrt{4dr-q^2}}{2rd-q^2}}{\sqrt{4dr-q^2}} & \text{for } \sqrt{2dr} < q < \sqrt{4dr}. \end{cases} \tag{8.67}$$

Proof The stability of $u = 0$ is determined by the sign of λ_1, the smallest eigenvalue of the eigenvalue problem

$$\begin{cases} d\varphi_{xx} - q\varphi_x + r\varphi + \lambda_1\varphi = 0, & \text{for } 0 < x < L, \\ d\varphi_x(0) - q\varphi(0) = \varphi_x(L) = 0. \end{cases} \tag{8.68}$$

Set $\varphi = e^{qx/(2d)}\psi$. Then

$$\begin{cases} d\psi_{xx} + \psi(-\frac{q^2}{4d} + r + \lambda_1) = 0, & \text{for } 0 < x < L, \\ \psi_x(0) - \frac{q}{2d}\psi(0) = \psi_x(L) + \frac{q}{2d}\psi(L) = 0. \end{cases} \tag{8.69}$$

Thus

$$\psi(x) = A\cos(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}x) + B\sin(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}x).$$

As a consequence of the boundary conditions of ψ, we have

$$A = B\frac{\sqrt{4d(r+\lambda_1)-q^2}}{q}$$

and

$$A\frac{-\sqrt{4d(r+\lambda_1)-q^2}}{2d}\sin(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}L)+B\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}\cos(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}L)$$
$$+\frac{q}{2d}[A\cos(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}L)+B\sin(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}L)]=0. \tag{8.70}$$

Combining these two equations and using $(A, B) \neq (0, 0)$, we obtain

$$\tan(\frac{\sqrt{4d(r+\lambda_1)-q^2}}{2d}L)\Big[\frac{q}{d}-\frac{2(r+\lambda_1)}{q}\Big]+\frac{\sqrt{4d(r+\lambda_1)-q^2}}{d}=0.$$

Set $\lambda_1 = 0$, then

$$\tan(\frac{\sqrt{4dr-q^2}}{2d}L^{**})=\frac{q\sqrt{4dr-q^2}}{2rd-q^2},$$

where L^{**} is the critical length given by (8.67) so that $\lambda_1 < 0$ when $L > L^{**}$; and $\lambda_1 > 0$ when $L < L^{**}$. This finishes the proof. ■

It is natural to consider more general boundary conditions:

$$\begin{cases} u_t = du_{xx} - qu_x + u(r-u), & \text{for } 0 < x < L,\ t > 0, \\ du_x(0,t) - qu(0,t) = 0, & \text{for } t > 0, \\ du_x(L,t) - qu(L,t) = -bqu(L,t), & \text{for } t > 0, \\ u(x,0) = u_0(x), & \text{for } 0 < x < L. \end{cases} \tag{8.71}$$

Here the (non-negative) parameter b measures the rate of population loss at the downstream end $x = L$ caused by the drift [38].

It is shown in [41] that the species can persist if and only if $q < q^*$ and $L > L^{***}$, where

$$q^* = \begin{cases} \sqrt{\frac{dr}{b(1-b)}} & 0 < b \le \frac{1}{2}; \\ \sqrt{4dr} & b \ge \frac{1}{2}, \end{cases}$$

and L^{***} is an explicit function of d, r, q, b. (See Lemmas 2.1 and 2.2 of [41] for details.) It is interesting to see that the critical value q^* depends on b only for $b \le \frac{1}{2}$, while for $b \ge \frac{1}{2}$, $q^* = \sqrt{4rd}$ is the minimal traveling wave speed for the Fisher-KPP equation in the whole real line.

Exercise

(i) Show that L^{**}, given in (8.67), is a strictly decreasing function of d.

(ii) Prove that there exists some $d_* > 0$ such that L^* is decreasing for $d < d^*$ and increasing for $d > d^*$. What is the biological interpretation of this result?

8.5.2 Evolution of Faster Dispersal

When the movement of organisms is subject to external forces such as river flow, how should species disperse to avoid the invasion of a mutant species with different movement strategies? In this subsection we consider a two-species competition model in an open advective environment: Individuals are exposed to unidirectional flow, with a net loss of individuals at the downstream end. We assumed that two species have the same advection rates but different dispersal rates. More specifically, we consider

$$\begin{cases} u_t = d_1 u_{xx} - qu_x + u(1-u-v), & 0<x<L,\ t>0,\\ v_t = d_2 v_{xx} - qv_x + v(1-u-v), & 0<x<L,\ t>0,\\ d_1 u_x(0,t) - qu(0,t) = d_2 v_x(0,t) - qv(0,t) = 0, & t>0,\\ u_x(L,t) = v_x(L,t) = 0, & t>0,\\ u(x,0)=u_0(x),\ v(x,0)=v_0(x), & 0<x<L. \end{cases} \tag{8.72}$$

Theorem 8.14 ([38]) *If $d_1 > d_2$, then the semi-trivial steady state $(u^*, 0)$, whenever it exists, is globally asymptotically stable among non-negative and non-trivial solutions of (8.72), where $u^* > 0$ is the unique positive solution of*

$$\begin{cases} d_1 u^*_{xx} - qu^*_x + u^*(1-u^*) = 0, \quad 0<x<L,\\ d_1 u^*_x(0) - qu^*(0) = u^*_x(L) = 0. \end{cases} \tag{8.73}$$

Theorem 8.14 implies that in an open advective environment, unidirectional flow can put slow dispersers at a disadvantage and higher dispersal rates are being selected. In particular, in a homogeneous advective environment with the free-flow boundary condition at the downstream end, a population with higher dispersal rate will always displace one with lower dispersal rate. We refer to [12, 13, 17, 33, 42–45, 54–56] for recent developments.

In the following we illustrate that $(u^*, 0)$ is stable for $d_1 \approx d_2$, $d_1 > d_2$, and unstable for $d_1 \approx d_2$, $d_1 < d_2$. This implies that a mutant can invade when rare if and only if it has the larger dispersal rate. In terms of the adaptive dynamics theory, the joint effects of small mutation and selection will tend to increase the average diffusion rate of the species.

The stability of $(u^*, 0)$ is determined by the sign of the smallest eigenvalue, denoted by $\lambda_1 = \lambda_1(d_1, d_2)$, of the problem

$$\begin{cases} d_2\varphi_{xx} - q\varphi_x + (1-u^*)\varphi + \lambda_1\varphi = 0, \quad 0<x<L,\\ d_2\varphi_x(0) - q\varphi(0) = \varphi_x(L) = 0,\\ \varphi > 0 \text{ in } (0, L). \end{cases} \tag{8.74}$$

Lemma 8.3 *Let λ_1 be the principal eigenvalue of (8.74). Then*

$$\frac{\partial \lambda_1}{\partial d_2}\bigg|_{d_2=d_1} = \frac{\int_0^L (e^{-\frac{q}{d_1}x}u^*)_x u_x^* dx}{\int_0^L e^{-\frac{q}{d_1}x}(u^*)^2 dx} < 0.$$

In particular, for $d_1 \approx d_2$, $(u^*, 0)$ is stable if $d_1 > d_2$, and unstable if $d_1 < d_2$.

Proof We first calculate $\frac{\partial \lambda_1}{\partial d_2}$. Denote $\varphi' = \frac{\partial \varphi}{\partial d_2}$ and $\lambda_1' = \frac{\partial \lambda_1}{\partial d_2}$, and differentiate the equation of φ with respect to d_2, we obtain

$$\begin{cases} d_2\varphi'_{xx} + \varphi_{xx} - q\varphi'_x + (1-u^*)\varphi' + \lambda_1'\varphi + \lambda_1\varphi' = 0, \quad 0 < x < L, \\ d_2\varphi'_x(0) + \varphi_x(0) - q\varphi'(0) = \varphi'_x(L) = 0. \end{cases} \tag{8.75}$$

Multiplying the first equation of (8.75) by $e^{-(q/d_2)x}\varphi$, the first equation of (8.74) by $e^{-(q/d_2)x}\varphi'$, subtracting and integrating the result in $(0, L)$, we have

$$\lambda_1' \int_0^L e^{-(q/d_2)x}\varphi^2 - \int_0^L \left(e^{-(q/d_2)x}\varphi\right)_x \varphi_x = 0.$$

When $d_2 = d_1$, we have $\lambda_1 = 0$ and $\varphi = Cu^*$ for some positive constant C. Thus

$$\frac{\partial \lambda_1}{\partial d_2}|_{d_2=d_1} = \frac{\int_0^L (e^{-\frac{q}{d_1}x}u^*)_x u_x^* dx}{\int_0^L e^{-\frac{q}{d_1}x}(u^*)^2 dx}. \tag{8.76}$$

We claim that $u^* < 1$ for $0 \le x \le L$. This follows directly from the fact that, for each constant $C \ge 1$, $\overline{u} = C$ is a strict super-solution for the equation of u^*.

Next we show that $u_x^* > 0$ for $0 \le x < L$. Since $u_x^*(L) = 0$ and $u^* < 1$, by the equation of u^* we see that $u_{xx}^*(L) < 0$. Hence, there exists some $\delta > 0$ such that $u_x^* > 0$ in $[L-\delta, L)$. To prove $u_x^* > 0$ in $[0, L)$, we argue by the contradiction. If not, we may assume that there exists some $x_1 < L - \delta$ such that $u_x^* > 0$ in $[x_1, L)$ and $u_x^*(x_1) = 0$. Set $w = u_x^*/u^*$. Then w satisfies

$$d_1 w_{xx} + w_x(2w - q) = u^* w \tag{8.77}$$

in $(0, L)$ and $w(x_1) = w(L) = 0$, $w > 0$ in (x_1, L). Therefore, there exists some $x_2 \in (x_1, L)$ such that $w(x_2) = \max_{x_1 \le x \le L} w(x)$. Hence, $w_x(x_2) = 0$ and $w_{xx}(x_2) \le 0$, which contradicts (8.77). This proves $u_x^* > 0$ for $x \in [0, L)$.

By the assertion $u^* < 1$ we have $d_1 u_{xx}^* - qu_x^* < 0$ in $(0, L)$. Hence, $d_1 u_x^* - qu^*$ is strictly decreasing. Since $d_1 u_x^*(0) - qu^*(0) = 0$, then $d_1 u_x^* - qu^* < 0$ for $0 < x \le L$. Therefore, $(e^{-(q/d_1)x}u^*)_x = e^{-(q/d_1)x}(u_x^* - \frac{q}{d_1}u^*) < 0$. This, together with $u_x^* > 0$ in $[0, L)$ and (8.76), shows that $\frac{\partial \lambda_1}{\partial d_2} < 0$ when $d_2 = d_1$. The proof is complete. ■

The boundary condition appears to play an important role in the outcome of evolution. In a homogeneous advective environment with the free-flow boundary conditions, larger dispersal rates evolve. In contrast, numerical simulations suggest that in a homogeneous advective environment with more hostile boundary conditions, there seems to evolve a unique, intermediate dispersal rate, which is evolutionarily stable. To be more specific, consider

$$\begin{cases} u_t = d_1 u_{xx} - q u_x + u(1-u-v), & \text{for } 0 < x < L,\ t > 0, \\ v_t = d_2 v_{xx} - q v_x + v(1-u-v), & \text{for } 0 < x < L,\ t > 0, \\ d_1 u_x(0,t) - qu(0,t) = d_2 v_x(0,t) - qv(0,t) = 0, & \text{for } t > 0, \\ d_1 u_x(L,t) - qu(L,t) = -bqu(L,t), & \text{for } t > 0, \\ d_2 v_x(L,t) - qv(L,t) = -bqv(L,t), & \text{for } t > 0, \\ u(x,0) = u_0(x),\ v(x,0) = v_0(x), & \text{for } 0 < x < L. \end{cases} \tag{8.78}$$

For $b \in [0, 1]$, it is shown in [38, 41] that if $d_1 > d_2$, then $(u^*, 0)$, whenever it exists, is globally asymptotically stable, i.e., the faster dispersal rate is always selected. For $b \geq 3/2$, we have the following conjecture:

Conjecture Suppose that $b \in [\frac{3}{2}, +\infty]$. Then there exists some $d_* > 0$ such that if $d_2 < d_1 \leq d_*$ or $d_* \leq d_1 < d_2$, then $(u^*, 0)$, whenever it exists, is globally asymptotically stable, where $u^* > 0$ satisfies

$$\begin{cases} d_1 u^*_{xx} - q u^*_x + u^*(1-u^*) = 0, & \text{for } 0 < x < L, \\ d_1 u^*_x(0) - qu^*(0) = 0, \\ d_1 u^*_x(L) - qu^*(L) = -bqu^*(L). \end{cases} \tag{8.79}$$

The d_* above is a special case of an evolutionarily stable strategy (ESS) in the evolution game theory, i.e., an ESS is a strategy which, if adopted by a population in a given environment, cannot be invaded by any alternative strategy that is initially rare. When $b \in [0, 1]$, we can regard $d_* = +\infty$, i.e., $+\infty$ is an ESS.

8.6 Conclusion

In this chapter we studied some reaction–diffusion models in spatial ecology. Topics covered include the logistic model for a single species and related issues, two-species Lotka–Volterra competition models in homogeneous and heterogeneous environments, the persistence and competition in advective environments, and the evolution of dispersal in heterogeneous and advective environments. We introduced some basic tools for reaction–diffusion equations and systems, including the super-solution and sub-solution method, the variational principle for principal eigenvalues, Lyapunov functionals, the comparison principles for parabolic equations and systems. We also presented some mathematical problems, ranging from elementary exercises to open research questions. In the following we discuss several recent works and point interested readers to relevant references:

In [46] Nagahara and Yanagida proved that the optimal control m^* of the functional J (see (8.33) in subsection 2.4) is of the "bang-bang" type, i.e., there exists a measurable set $E \subset \Omega$ such that $m^* = 1$ in E and $m^* = 0$ in the complement of E. This answers a conjecture of Ding et al. [15] affirmatively.

We recall that $\theta(\cdot, d)$ is the unique positive solution of (8.30). An open problem is whether $\max_{x\in\bar{\Omega}} \theta(x, d)$ is monotone decreasing in d. Such question came from the study of predator–prey systems in heterogeneous environments [39]. If Ω is an interval and m is monotone, then $\max_{x\in\bar{\Omega}} \theta(x, d)$ is monotone decreasing in d; see [35], which extended an earlier result in [39]. However, the question remains open for general Ω and m.

In [49] Perthame and Souganidis considered an integro-PDE model for a population structured by the spatial variables and a continuous trait variable which is the diffusion rate. Such model can be viewed as the extension of the competition model (8.56) from two-phenotypes to infinitely many phenotypes. It is shown in [49], and independently in [32], that in the limit of small mutation rate, the unique steady state solution forms a Dirac mass in the trait variable, supported at the smallest possible diffusion rate. This echoes the result of Dockery et al. in [16], i.e., the slowest diffusion rate is favored. We refer to [19, 31] for further development.

For system (8.78), the species can persist if and only if $q < q^*$ and $L > L^{***}$ (Lemmas 2.1 and 2.2 of [41]). We proved in [18] that if $0 < b \leq 3/2$, then L^{***} is strictly decreasing in d; if $b > 3/2$, then L^{***} decreases in d first and then increases in d. This reveals a dramatic difference between $b < 3/2$ and $b > 3/2$. Our preliminary analysis of system (8.78) suggests that the conclusion of Theorem 8.14, which states that the faster diffuser can always competitively exclude the slower diffuser, may fail for some $1 < b < 3/2$, i.e., the faster dispersal rate may not be selected. This is in strong contrast to the case $0 \leq b \leq 1$ for which the faster dispersal rate is always selected [38, 41].

In conclusion, the materials presented in this chapter illustrate some interesting questions in spatial ecology and evolution. Such questions are, on the one hand, well connected with important issues in biology, and on the other hand, deeply rooted in mathematics and bringing new and exciting challenges.

Acknowledgements We sincerely thank the referee for his comments and suggestions which help improve the presentation. KYL and YL were partially supported by the NSF grant DMS-1411476 and DMS-1853561. Part of the work was done during the visit of YL to the University of Alberta to participate in the 2016 *Séminaire de Mathématiques Supérieures: Dynamics of Biological Systems Summer School*, and he thanks the organizers for the warm hospitality.

References

1. I. Averill, K.-Y. Lam, Y. Lou, The role of advection in a two-species competition model: a bifurcation approach. Mem. Am. Math. Soc. **245**(1161), v+117 (2017)
2. X.L. Bai, X.Q. He, F. Li, An optimization problem and its application in population dynamics. Proc. Am. Math. Soc. **144**, 2161–2170 (2016)
3. K.J. Brown, S.S. Lin, On the existence of positive eigenvalue problem with indefinite weight function. J. Math. Anal. Appl. **75**, 112–120 (1980)
4. R.S. Cantrell, C. Cosner, Diffusive logistic equations with indefinite weights: population models in a disrupted environments. Proc. R. Soc. Edinb. **112A**, 293–318 (1989)
5. R.S. Cantrell, C. Cosner, The effects of spatial heterogeneity in population dynamics. J. Math. Biol. **29**, 315–338 (1991)
6. R.S. Cantrell, C. Cosner, Should a park be an island? SIAM J. Appl. Math. **53**, 219–252(1993)
7. R.S. Cantrell, C. Cosner, On the effects of spatial heterogeneity on the persistence of interacting species. J. Math. Biol. **37**, 103–145 (1998)
8. R.S. Cantrell, C. Cosner, *Spatial Ecology via Reaction-Diffusion Equations*. Series in Mathematical and Computational Biology (Wiley, Chichester, 2003)
9. R.S. Cantrell, C. Cosner, Y. Lou, Movement towards better environments and the evolution of rapid diffusion. Math Biosci. **204**, 199–214 (2006)
10. R.S. Cantrell, C. Cosner, Y. Lou, Advection mediated coexistence of competing species. Proc. R. Soc. Edinb. **137A**, 497–518 (2007)
11. C. Cosner, Reaction-diffusion-advection models for the effects and evolution of dispersal. Discr. Cont. Dyn. Syst. **34**, 1701–1745 (2014)
12. R.H. Cui, Y. Lou, Spatial SIS epidemic models in advective environments. J. Differ. Equ. **261**, 3305–3343 (2016)
13. R.H. Cui, K.-Y. Lam, Y. Lou, Dynamics and asymptotic profiles of steady states to an epidemic model in advective environments. J. Differ. Equ. **263**, 2343–2373 (2017)
14. D. DeAngelis, W.-M. Ni, B. Zhang, Dispersal and spatial heterogeneity: single species. J. Math. Biol. **72**, 239–254 (2016)
15. W. Ding, H. Finotti, S. Lenhart, Y. Lou, Q. Ye, Optimal control of growth coefficient on a steady-state population model. Nonlinear Anal. Real World Appl. **11**, 688–704 (2010)
16. J. Dockery, V. Hutson, K. Mischaikow, M. Pernarowski, The evolution of slow dispersal rates: a reaction-diffusion model. J. Math. Biol. **37**, 61–83 (1998)
17. M. Golubitsky, W. Hao, K.-Y. Lam, Y. Lou, Dimorphism by singularity theory in a model for river ecology. Bull. Math. Biol. **79**, 1051–1069 (2017)
18. M. Golubitsky, W. Hao, K.-Y. Lam, Y. Lou, Evolution of dispersal for a river species in homogeneous advective environment, in preparation
19. W. Hao, K.-Y. Lam, Y. Lou, Concentration phenomena in an integro-PDE model for evolution of conditional dispersal. Indiana Univ. Math. J. **68**, 881–923 (2019)
20. A. Hastings, Can spatial variation alone lead to selection for dispersal? Theor. Popul. Biol. **24**, 244–251 (1983)
21. X. He, W.-M. Ni, The effects of diffusion and spatial variation in Lotka-Volterra competition-diffusion system I: heterogeneity vs. homogeneity. J. Differ. Equ. **254**, 528–546 (2013)
22. X. He, W.-M. Ni, The effects of diffusion and spatial variation in Lotka-Volterra competition-diffusion system II: the general case. J. Differ. Equ. **254**, 4088–4108 (2013)
23. X. He, W.-M. Ni, Global dynamics of the Lotka-Volterra competition-diffusion system: diffusion and spatial heterogeneity I. Commun. Pure Appl. Math. **69**, 981–1014 (2016)
24. X. He, W.-M. Ni, Global dynamics of the Lotka-Volterra competition-diffusion system with equal amount of total resources, II. Calc. Var. Partial Differ. Equ. **55**, Art. 25, 20 (2016)
25. S.-B. Hsu, H. Smith and P. Waltman, Competitive exclusion and coexistence for competitive systems on ordered Banach spaces. Trans. Am. Math. Soc. **348**, 4083–4094 (1996)

26. V. Hutson, K. Mischaikow, P. Pol*áč*ik, The evolution of dispersal rates in a heterogeneous time-periodic environment. J. Math. Biol. **43**, 501–533 (2001)
27. V. Hutson, Y. Lou, K. Mischaikow, Spatial heterogeneity of resources versus Lotka-Volterra dynamics. J. Differ. Equ. **185**, 97–136 (2002)
28. V. Hutson, Y. Lou, K. Mischaikow, P. Poláčik, Competing species near the degenerate limit. SIAM J. Math. Anal. **35**, 453–491 (2003)
29. V. Hutson, Y. Lou, K. Mischaikow, Convergence in competition models with small diffusion coefficients. J. Differ. Equ. **211**, 135–161 (2005)
30. C.Y. Kao, Y. Lou, E. Yanagida, Principal eigenvalue for an elliptic problem with indefinite weight on cylindrical domains. Math. Biosci. Eng. **5**, 315–335 (2008)
31. K.-Y. Lam, Stability of Dirac concentrations in an integro-PDE Model for evolution of dispersal. Calc. Var. Partial Differ. Equ. **56**, 32 pp. (2017)
32. K.-Y. Lam, Y. Lou, An integro-PDE model for evolution of random dispersal. J. Funct. Anal. **272**, 1755–1790 (2017)
33. K.-Y. Lam, Y. Lou, F. Lutscher, Evolution of dispersal in closed advective environments. J. Biol. Dyn. **9**, 188–212 (2015)
34. K.-Y. Lam, W.-M. Ni, Uniqueness and complete dynamics of the Lotka-Volterra competition diffusion system. SIAM J. Appl. Math. **72**, 1695–1712 (2012)
35. R. Li, Y. Lou, Some monotone properties for solutions to a reaction-diffusion model. Discr. Contin. Dyn. Syst. B **24**, 4445–4455 (2019)
36. S. Liang, Y. Lou, On the dependence of population size upon random dispersal rate. Discrete Contin. Dyn. Syst. B **17**, 2771–2788 (2012)
37. Y. Lou, On the effects of migration and spatial heterogeneity on single and multiple species. J. Differ. Equ. **223**, 400–426 (2006)
38. Y. Lou, F. Lutscher, Evolution of dispersal in open advective environments. J. Math. Biol. **69**, 1319–1342 (2014)
39. Y. Lou, B. Wang, Local dynamics of a diffusive predator-prey model in spatially heterogeneous environment. J. Fixed Point Theory Appl. **19**, 755–772 (2017)
40. Y. Lou, E. Yanagida, Minimization of the principal eigenvalue with indefinite weight and applications to population dynamics. Jpn J. Indus. Appl. Math. **23**, 275–292 (2006)
41. Y. Lou, P. Zhou, Evolution of dispersal in advective homogeneous environment: the effect of boundary conditions. J. Differ. Equ. **259**, 141–171 (2015)
42. Y. Lou, D.M. Xiao, P. Zhou, Qualitative analysis for a Lotka-Volterra competition system in advective homogeneous environment. Discrete Contin. Dyn. Syst. A **36**, 953–969 (2016)
43. Y. Lou, X.-Q. Zhao, P. Zhou, Global dynamics of a Lotka-Volterra competition-diffusion-advection system in heterogeneous environments. J. Math. Pures Appl. **121**, 47–82 (2019)
44. F. Lutscher, E. Pachepsky, M.A. Lewis, The effect of dispersal patterns on stream populations. SIAM Rev. **47**, 749–772 (2005)
45. F. Lutscher, M.A. Lewis, E. McCauley, Effects of heterogeneity on spread and persistence in rivers. Bull. Math. Biol. **68**, 2129–2160 (2006)
46. K. Nagahara, E. Yanagida, Maximization of the total population in a reaction-diffusion model with logistic growth. Calc. Var. Partial Differ. Equ. **57**, Art 80, 14pp (2018)
47. W.-M. Ni, *The Mathematics of Diffusion*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 82 (SIAM, Philadelphia, 2011)
48. C.V. Pao, *Nonlinear Parabolic and Elliptic Equations* (Springer, Berlin, 2012)
49. B. Perthame, P.E. Souganidis, Rare mutations limit of a steady state dispersal evolution model. Math. Model. Nat. Phenom. **11**, 154–166 (2016)
50. M.H. Protter, H.F. Weinberger, *Maximum Principles in Differential Equations* Corrected reprint of the 1967 original (Springer, New York, 1984)
51. H.L. Smith, Monotone dynamical systems, in *An Introduction to the Theory of Competitive and Cooperative Systems*. Mathematical Surveys and Monographs, vol. 41 (American Mathematical Society, Providence, 1995)

52. D.C. Speirs, W.S. Gurney, Population persistence in rivers and estuaries. Ecology **82**, 1219–1237 (2001)
53. O. Vasilyeva, F. Lutscher, Population dynamics in rivers: analysis of steady states. Can. Appl. Math. Q. **18**, 439–469 (2011)
54. O. Vasilyeva, F. Lutscher, Competition in advective environments. Bull. Math. Biol. **74**, 2935–2958 (2012)
55. P. Zhou, On a Lotka-Volterra competition system: diffusion vs advection. Calc. Var. Partial Differ. Equ. **55**, Art. 137, 29 (2016)
56. X.-Q. Zhao, P. Zhou, On a Lotka-Volterra competition model: the effects of advection and spatial variation. Calc. Var. Partial Differ. Equ. **55**, Art. 73, 25 (2016)

Chapter 9
Kinetic Equations and Cell Motion: An Introduction

Benoît Perthame

Abstract Kinetic theory is an old subject which finds its motivations in the description of fluids at the so-called mesoscopic scale where molecules interact but are too numerous for describing the interacting particles individually. We present several examples from physics, we give some mathematical background showing that the kinetic-transport equation enjoys interesting functional analytic properties as other partial differential equations. We also describe in full generality how macroscopic models are derived from kinetic equations. This material gives us the tools to introduce models for bacterial run and tumble motion. The subject has been progressing quickly in the last decades, and a hierarchy of models are now available up to the scale of molecular pathways describing the cell decision to tumble.

Keywords Kinetic equations · Run and tumble · Keller–Segel system · Asymptotic analysis · Diffusion limit · Biochemical pathways

9.1 Introduction

Kinetic physics is an old field which aims at describing natural phenomena in the phase space, thanks to the position (denoted by x below) and velocity of particles (denoted by $\xi \in V$ below where V is the set of possible velocities). It goes back to James C. Maxwell who wrote an essay *On the Stability of the Motion of Saturn's Rings* in 1859 explaining that Saturn rings are formed of colliding rocks. A few years later (1872), Ludwig Boltzmann gave a description of a gas as the result of collisions between molecules. He wrote the famous *Boltzmann equation* for

B. Perthame (✉)
Sorbonne University, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions, Paris, France

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4_9

the density $f(x, \xi, t)$ of molecules which at position $x \in \mathbb{R}^3$ have the velocity $\xi \in V = \mathbb{R}^3$,

$$\underbrace{\frac{\partial}{\partial t} f(x, \xi, t) + \xi.\nabla_x f}_{\text{Transport with velocity } \xi} = \underbrace{Q(f, f).}_{\text{Binary collisions}}$$

We refer to the textbooks [21, 22, 76] for this rich and deep subject.

A new chapter on kinetic physics was written 60 years after Boltzmann, with the description of plasmas by Andrei Vlasov (1938). In physics, a plasma is matter made of charged particles (ions, electrons) and the equations are written, again with $v \in \mathbb{R}^3$, as

$$\begin{cases} \frac{\partial}{\partial t} f(x, \xi, t) + \xi.\nabla_x f + \overbrace{E(x, t).\nabla_\xi f}^{\text{Force of the electric field}} = 0, \\ \\ -\Delta U = n(x, t) := \int_{\mathbb{R}^3} f(x, \xi, t) d\xi, \qquad E = -\nabla_x U, \end{cases}$$

where U denotes the electric potential and E the electric field.

In nuclear physics, X-ray imaging and radiotherapy [19], scattering (neutrons, photons, radiative transfer), the equation is written for particles velocities $\xi \in V = \mathbb{S}^2$ as

$$\frac{\partial}{\partial t} f(x, \xi, t) + \xi.\nabla_x f + \underbrace{\sigma(x, t) f}_{\text{Absorbtion}} = \underbrace{\int_V K(\xi, \xi') f(x, \xi', t) d\xi'}_{\text{Emmission}}. \tag{9.1}$$

The kernel $K(\xi, \xi')$ describes the rate of change from velocity ξ' to velocity ξ (usually under the effect of interaction with atoms or molecules).

The formalism of kinetic equation is well adapted to describe any phenomenon for which the knowledge of positions is not enough to predict the dynamics. As for planets in a solar system, the knowledge of the velocities is also necessary. The same need occurs in several areas of biology. For instance, in neuron networks, the density is about voltage and gating (ionic channels opening and closing) [15, 61]. Bacterial motion is also well described, at the cell scale by kinetic equations. This is the specific subject we are going to treat here.

We organize our presentation as follows. We begin by some function analytic tools for the kinetic-transport equation in order to show that, because space and velocity interplay, several remarkable regularity estimates can be derived which give tools for the mathematical analysis. Then we turn to asymptotic methods which make the relation between kinetic equations and macroscopic equation (in the space, ignoring individual velocities). This is one of the historical benefits from kinetic theory: based on the knowledge of individual behavior, one can derive the coefficients at the macro-scale. It is only in Sect. 9.5 that we introduce the models of bacterial motion. We begin with the simplest model which allows to derive the Patlak/Keller–Segel model [47], this is the most classical system to describe chemotactic movement. This paves the way for a more evolved class of equation in Sect. 9.6 where cells decide of their tumbling rate through integration along their path. The next section treats of the recent subject of molecular pathways for tumbling decision. We conclude with some references and mentioning various subjects that we do not have time to treat here.

9.2 Functional Analysis of Kinetic Equations

A deep mathematical theory has been developed about kinetic equations, with several functional inequalities which are easy to describe on the simple kinetic-transport equation

$$\begin{cases} \frac{\partial}{\partial t} f(x,\xi,t) + \xi.\nabla_x f = 0, & x \in \mathbb{R}^d,\ \xi \in V,\ t \in \mathbb{R}, \\ f(x,\xi,0) = f^0(x,\xi). \end{cases} \tag{9.2}$$

Using the method of characteristics, the solution is

$$f(x,\xi,t) = f^0(x - \xi t, \xi), \tag{9.3}$$

from which one directly concludes many useful consequences as non-negativity, $f^0 \geq 0$ implies $f(x,\xi,t) \geq 0$ and several others, which we list, non-exhaustively, now.

L^p **Estimates** Using Fubini's theorem, we obtain from (9.3) mass conservation

$$\int_{\mathbb{R}^d \times V} f(x,\xi,t) dx d\xi = \int_{\mathbb{R}^d \times V} f^0(x,\xi) dx d\xi.$$

More generally all L^p norms are conserved for all $p \geq 1$

$$\int_{\mathbb{R}^d \times V} |f(x,\xi,t)|^p dx d\xi = \int_{\mathbb{R}^d \times V} |f^0(x,\xi)|^p dx d\xi. \tag{9.4}$$

Macroscopic Quantities More interesting conclusions can be drawn on the so-called *macroscopic quantities* defined by

$$n(x,t) := \int_V f(x,\xi,t) d\xi, \qquad \text{density,}$$

$$n(x,t)u(x,t) := \int_V \xi f(x,\xi,t) d\xi, \qquad \text{momentum,}$$

$$E(x,t) = n(x,t)\frac{|u(x,t)|^2}{2} + n(x,t)e(x,t) := \int_V \frac{|\xi|^2}{2} f(x,\xi,t) d\xi, \qquad \text{energy,}$$

where $e(x,t)$ denotes the internal energy (a function of temperature). These represent, usually speaking, the observables. One cannot easily access experimentally to the molecular distribution f of molecules or ions, but quantities as the density, velocity, temperature, electric field are measurable.

Dispersion Inequalities The simplest functional inequality is certainly the *dispersion* inequality which gives the time decay of the macroscopic density. Notice that in the phase space there is no decay but conservation according to (9.4). The dispersion inequality [5] reads

$$|n(x,t)| \leq \frac{1}{t^d} \int_{\mathbb{R}^d} \sup_{\xi \in V} |f^0(x,\xi)| dx. \tag{9.5}$$

This can be compared to the heat/Fourier equation for which the decay is as $\frac{1}{t^{d/2}}$.

Proof of (9.5). We have

$$|n(x,t)| \leq \int_{\mathbb{R}^d} |f(x,\xi,t)| d\xi = \int_{\mathbb{R}^d} |f^0(x-\xi t,\xi)| d\xi,$$

and thus

$$|n(x,t)| \leq \int_{\mathbb{R}^d} \sup_{\eta \in V} |f^0(x-\xi t,\eta)| d\xi = \int_{\mathbb{R}^d} \sup_{\eta \in V} ||f^0(x-y,\eta)| \frac{dy}{t^d},$$

and thus result follows directly. □

Strichartz Inequalities for Kinetic Equations A consequence, by duality of the above dispersive estimates, are the Strichartz inequalities for kinetic equations which were discovered in [20], see also [58], F. Catsella's PhD for the correct general numbers and [46] for the endpoints. The Strichartz inequalities are

$$\|\varrho\|_{L^q(\mathbb{R}_t; L^p(\mathbb{R}_x^d))} \leq C(d) \|f^0\|_{L^a(\mathbb{R}^{2d})},$$

for any real numbers a, p, and q such that

$$1 \leq p < \frac{d}{d-1}, \qquad \frac{2}{q} = d(1-\frac{1}{p}), \qquad 1 \leq a = \frac{2p}{p+1} < \frac{2d}{2d-1}.$$

Kinetic Averaging Lemmas We begin with a very simple result, in fact the first one obtained in the framework below in [33]:

Theorem 9.1 *Let $f,\ g \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$ satisfy the stationary kinetic equation*

$$\xi.\nabla_x f = g,$$

then, for measurable subsets $V \subset \mathbb{R}^d$ such that diam(V) is finite, we have

$$\| \int_V f(x,\xi)d\xi \|_{H^{1/2}(\mathbb{R}^d)} \leq C\|f\|_{L^2}^{1/2} \ \|g\|_{L^2}^{1/2}$$

for some constant $C(\mathrm{diam}(V), d)$.

Proof We perform Fourier transform in x and set

$$\widehat{f}(k,\xi) = \int e^{-ixk} f(x,\xi)dx.$$

We have

$$ik.\xi \ \widehat{f}(k,\xi) = \widehat{g}(k,\xi).$$

Because we cannot invert globally the symbol $ik.\xi$, we introduce a parameter $\lambda > 0$ and write

$$(\lambda + ik.\xi) \ \widehat{f}(k,\xi) = \widehat{g}(k,\xi) + \lambda \widehat{f}(k,\xi).$$

Therefore, we may compute

$$\widehat{f}(k,\xi) = \frac{\widehat{g}(k,\xi) + \lambda \widehat{f}(k,\xi)}{\lambda + ik.\xi},$$

$$\widehat{n}(k,\xi) = \int_V \frac{\widehat{g}(k,\xi)}{\lambda + ik.\xi} d\xi + \lambda \int_V \frac{\widehat{f}(k,\xi)}{\lambda + ik.\xi} d\xi.$$

Consider the first term on the right-hand side, we may estimate it as

$$\begin{aligned} \left| \int_V \frac{\widehat{g}(k,\xi)}{\lambda + ik.\xi} d\xi \right|^2 &\leq \int_V |\widehat{g}(k,\xi)|^2 d\xi \int_V \frac{1}{|\lambda + ik.\xi|^2} d\xi \\ &\leq \int_V |\widehat{g}(k,\xi)|^2 d\xi \frac{1}{\lambda^2} \int_V \frac{1}{1 + |\frac{k}{\lambda}.\xi|^2} d\xi \\ &\leq \int_V |\widehat{g}(k,\xi)|^2 d\xi \ \frac{\pi}{\lambda |k|} \ \mathrm{diam}(V)^{d-1}. \end{aligned}$$

Arguing similarly on the second term on the right-hand side, we conclude that

$$|\widehat{n}(k,\xi)|^2 \leq \pi \,\mathrm{diam}(V)^{d-1} \Big[\int_V |\widehat{g}(k,\xi)|^2 d\xi \ \frac{1}{\lambda |k|} + \int_V |\widehat{f}(k,\xi)|^2 d\xi \ \frac{\lambda}{|k|} \Big].$$

Choosing the value of λ which minimizes the right-hand side, we find

$$|k|\,|\widehat{n}(k,\xi)|^2 \le C\left(\int_V |\widehat{g}(k,\xi)|^2 d\xi \int_V |\widehat{f}(k,\xi)|^2\right)^{1/2}$$

which is exactly the announced result. □

A similar result can be proved for the evolution equation and reads:

Theorem 9.2 *Let $f,\ g \in L^2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R})$ and*

$$\frac{\partial}{\partial t} f(x,\xi,t) + \xi.\nabla_x f = g.$$

Then, for measurable subsets $V \subset \mathbb{R}^d$ such that diam(V) is finite, we have

$$\| \int_V f(x,\xi,t)d\xi \|_{H^{1/2}(\mathbb{R}^d\times\mathbb{R})} \le C \|f\|_{L^2}^{1/2}\ \|g\|_{L^2}^{1/2}.$$

A number of generalizations and improvements have been derived around averaging lemmas and regularity gain in kinetic equations. The major reason is that averaging lemmas show that macroscopic quantities are compact, and this is useful for existence theory and for asymptotic results as described later. See [62] for a general compactness result.

Further reading on these questions can be found in [30, 58].

Existence of Solutions to Nonlinear Kinetic Equations The functional analytic framework presented above has been a major step towards proving existence for solutions to the Boltzmann equation, the Vlasov-Maxwell equation, and many others. We do not present this topic which is too far from our purpose here and relies on an enormous literature.

9.3 Diffusion Limit

The derivation of macroscopic equations, that are written in the physical space using only the variable x is an important chapter of kinetic theory. Indeed, to compute numerically in the phase space is resource consuming because of the number of variables, but the derivation of macroscopic coefficient from the knowledge of particles properties is useful. We refer both to an early fundamental paper and a recent survey for this topic [4, 32, 39]. In this section, to simplify the setting, we assume that $V = \mathbb{R}^d$.

We begin with several notations and assumptions. Let us introduce a function M, like Maxwellian which refers to the special case $M(\xi) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{|\xi|^2}{2})$, with the properties

$$M(\xi) \geq 0, \quad \int_{\mathbb{R}^d} M(\xi) d\xi = 1, \qquad \int_{\mathbb{R}^d} |\xi|^2 M(\xi) d\xi < \infty, \qquad \int_{\mathbb{R}^d} \xi M(\xi) d\xi = 0, \tag{9.6}$$

and let the inhomogeneous collision intensity be denoted by $k(x)$, and we assume that

$$0 < k_- \leq k(x) \leq k_+, \qquad k(x) \in Lip(\mathbb{R}^d). \tag{9.7}$$

We consider the scattering type equation rescaled as follows

$$\begin{cases} \varepsilon \frac{\partial}{\partial t} f_\varepsilon(x, \xi, t) + \xi \cdot \nabla_x f_\varepsilon = \frac{k(x)}{\varepsilon} \big[n_\varepsilon(x, t) M(\xi) - f_\varepsilon \big], \\ n_\varepsilon(x, t) := \int_{\mathbb{R}^d} f_\varepsilon(x, \xi, t) d\xi, \\ f_\varepsilon(x, \xi, t = 0) = f^0(x, \xi). \end{cases} \tag{9.8}$$

We assume that the relative L^2 norm (see [59] for a general introduction) of f^0 is bounded, that is

$$\int_{\mathbb{R}^{2d}} \frac{|f^0(x, \xi)|^2}{M(\xi)} d\xi dx := K_2^0 < \infty. \tag{9.9}$$

Theorem 9.3 *As $\varepsilon \rightarrow$, there is a weak limit*

$$f_\varepsilon(x, \xi, t) \underset{\varepsilon \to 0}{\longrightarrow} n(x, t) M(\xi),$$

$$\begin{cases} \frac{\partial}{\partial t} n(x, t) - \text{div} \big[\frac{A}{k(x)} \nabla n(x, t) \big] = 0, \\ n(x, t = 0) = n^0(x) := \int_{\mathbb{R}^d} f^0(x, \xi) d\xi, \end{cases} \tag{9.10}$$

with the matrix $A \in M_{d \times d}$ given by

$$A_{ij} = \int_{\mathbb{R}^d} \xi_i \xi_j M(\xi) d\xi, \qquad 1 \leq i, \ j \leq d.$$

Proof We are going to use the method of moments which is based on the formula obtained integrating in ξ the Eq. (9.8)

$$\frac{\partial}{\partial t} n_\varepsilon + \text{div } J_\varepsilon = 0, \tag{9.11}$$

with

$$J_\varepsilon(x,t) = \int_{\mathbb{R}^d} \frac{\xi}{\varepsilon} f_\varepsilon(x,\xi,t)d\xi. \tag{9.12}$$

We are going to prove that both n_ε and J_ε converge and identify the relation between their limits as the Eq. (9.10). To do so, we proceed in several steps.

Step 1 (A Priori Estimates) We prove the bounds

$$\begin{cases} \int_{\mathbb{R}^{2d}} \frac{f_\varepsilon(x,\xi,t)^2}{M(\xi)} dx d\xi \leq K_2^0, \qquad \int_{\mathbb{R}^{2d}} n_\varepsilon(x,t)^2 dx \leq K_2^0, \\ \int_0^\infty \int_{\mathbb{R}^{2d}} \left|\frac{f_\varepsilon(x,\xi,t)}{M(\xi)} - n_\varepsilon\right|^2 M(\xi) dx d\xi dt \leq \frac{\varepsilon^2}{2}\frac{K_2^0}{k_-}, \\ \int_0^\infty \int_{\mathbb{R}^{2d}} |J_\varepsilon(x,t)|^2 dx dt \leq \frac{1}{2}\frac{K_2^0}{k_-} \int_{\mathbb{R}^d} |\xi|^2 M(\xi) d\xi. \end{cases} \tag{9.13}$$

These are obtained multiplying the Eq. (9.8) by $\frac{f_\varepsilon}{M}$ and integrating in x, ξ. We obtain

$$\begin{aligned} \frac{\varepsilon}{2}\frac{d}{dt} \int_{\mathbb{R}^{2d}} \frac{f_\varepsilon(x,\xi,t)^2}{M(\xi)} dx d\xi &= \tfrac{1}{\varepsilon} \int_{\mathbb{R}^{2d}} k(x) \left[n_\varepsilon(x,t) f_\varepsilon(x,\xi,t) - \frac{f_\varepsilon(x,\xi,t)^2}{M(\xi)} \right] dx d\xi \\ &= -\frac{1}{\varepsilon} \int_{\mathbb{R}^{2d}} k(x) \left| \frac{f_\varepsilon(x,\xi,t)}{M(\xi)} - n_\varepsilon \right|^2 M(\xi) dx d\xi. \end{aligned}$$

From this, we conclude the first three bounds, noticing that the Cauchy–Schwarz inequality gives

$$n_\varepsilon(x,t)^2 = \left(\int_{\mathbb{R}^d} \frac{f_\varepsilon(x,\xi,t)}{M(\xi)^{1/2}} M(\xi)^{1/2} d\xi \right)^2 \leq \int_{\mathbb{R}^d} \frac{f_\varepsilon(x,\xi,t)^2}{M(\xi)} d\xi.$$

For the last inequality, we use the assumption (9.6) to write

$$J_\varepsilon(x,t) = \int_{\mathbb{R}^d} \frac{\xi}{\varepsilon} f_\varepsilon = \frac{1}{\varepsilon} \int_{\mathbb{R}^d} \xi [\frac{f_\varepsilon}{M(\xi)} - n_\varepsilon(x,t)] M(\xi) d\xi.$$

Therefore, using the Cauchy–Schwarz inequality, we find

$$|J_\varepsilon(x,t)|^2 \leq \frac{1}{\varepsilon^2} \int_{\mathbb{R}^d} \left| \frac{f_\varepsilon}{M(\xi)} - n_\varepsilon(x,t) \right|^2 M(\xi) d\xi \int_{\mathbb{R}^d} |\xi|^2 M(\xi) d\xi.$$

Integrating in space and time, and using the second inequality in (9.13), we conclude the last bound.

Step 2 (Computing J_ε) We may rewrite the Eq. (9.8) in the form

$$\frac{\xi}{\varepsilon}f_\varepsilon = \frac{\xi}{\varepsilon}n_\varepsilon(x,t)M(\xi) - \varepsilon\frac{\xi}{k(x)}\frac{\partial}{\partial t}f_\varepsilon(x,\xi,t) - \frac{\xi}{k(x)}\xi\cdot\nabla_x f_\varepsilon,$$

$$J_\varepsilon(x,t) = \int_{\mathbb{R}^d}\frac{\xi}{\varepsilon}f_\varepsilon d\xi = -\varepsilon\int_{\mathbb{R}^d}\frac{\xi}{k(x)}\frac{\partial}{\partial t}f_\varepsilon(x,\xi,t) - \int_{\mathbb{R}^d}\frac{\xi}{k(x)}\xi\cdot\nabla_x f_\varepsilon.$$

In other words, we also have

$$J_\varepsilon(x,t) = -\frac{\varepsilon^2}{k(x)}\frac{\partial J_\varepsilon(x,t)}{\partial t} - \int_{\mathbb{R}^d}\frac{\xi\otimes\xi}{k(x)}\nabla_x[f_\varepsilon - n_\varepsilon M(\xi)]d\xi - \frac{A}{k(x)}\nabla_x n_\varepsilon.$$

And thus going back to the mass conservation equation (9.11), we find

$$\frac{\partial}{\partial t}n_\varepsilon - \text{div}\big[A.\frac{\nabla n_\varepsilon(x,t)}{k(x)}\big] = \frac{\varepsilon^2}{k(x)}\frac{\partial J_\varepsilon(x,t)}{\partial t} + \int_{\mathbb{R}^d}\frac{\xi\otimes\xi}{k(x)}\nabla_x[f_\varepsilon - n_\varepsilon M(\xi)]d\xi.$$

Step 3 (Weak Limit) From the bounds (9.13), we may extract subsequences such that, in the weak limit

$$n_\varepsilon(x,t) \underset{\varepsilon\to 0}{\rightharpoonup} n(x,t), \qquad J_\varepsilon(x,t) \underset{\varepsilon\to 0}{\rightharpoonup} J(x,t),$$

and the conclusion of step 2 is simply that, in the distributional sense (that means in multiplying by smooth test functions with compact support) we have

$$\frac{\partial}{\partial t}n + \text{div}J = 0 = \frac{\partial}{\partial t}n - \text{div}\big[A.\frac{\nabla n(x,t)}{k(x)}\big].$$

Notice that the initial condition is preserved in distributional sense (see [60] for details). Therefore, there is a unique solution to the parabolic equation (9.10) and thus the full family n_ε, J_ε converges, which also implies that the full family f_ε converges weakly thanks to the second line in (9.13). □

It is possible to prove strong convergence of f_ε and n_ε, for instance using a variant of the averaging lemma in Sect. 9.2, noticing that we may write, using the bounds (9.13)

$$\varepsilon\frac{\partial}{\partial t}f_\varepsilon + \xi.\nabla f_\varepsilon = g_\varepsilon$$

with f_ε and g_ε bounded in L^2. The method based on averaging lemma shows strong convergence even when we generalize the initial data to a family $f^0_\varepsilon(x,\xi) \underset{\varepsilon\to 0}{\rightharpoonup} f(x,\xi,t)$ in the weighted $L^2(\frac{dxd\xi}{M(\xi)}$ norm.

9.4 The Drift-Diffusion Limit

Not only the parabolic equation (9.10) can be derived as a macroscopic limit of kinetic equations but also the more general drift-diffusion equation (also called Fokker–Planck equation)

$$\begin{cases} \frac{\partial}{\partial t} n(x,t) - \mathrm{div}\big[\frac{1}{k(x)} \nabla\big(A(x) n(x,t)\big)\big] + \mathrm{div}\big[n(x,t) U(x,t)\big] = 0, \\ n(x, t=0) = n^0(x) := \int_{\mathbb{R}^d} f^0(x,\xi) d\xi, \end{cases} \tag{9.14}$$

for a given symmetric positive matrix $A(x)$ and a given drift $U(x)$.

In this section, we do not try to give a rigorous derivation with a priori estimates as we did before because the technical details are more complicated and we prefer to concentrate on the mechanism at work. For the same reason, we choose a simple kinetic equation to depart from

$$\begin{cases} \varepsilon \frac{\partial}{\partial t} f_\varepsilon(x,\xi,t) + \xi \cdot \nabla_x f_\varepsilon = \frac{k(x)}{\varepsilon}\big[n_\varepsilon(x,t) M_\varepsilon(x,\xi) - f_\varepsilon\big], \\ n_\varepsilon(x,t) := \int_{\mathbb{R}^d} f_\varepsilon(x,\xi,t) d\xi, \\ f_\varepsilon(x,\xi,t=0) = f^0(x,\xi). \end{cases} \tag{9.15}$$

The assumptions on the ε-dependent Maxwellian distribution $M_\varepsilon(x,\xi) > 0$ is

$$\int_{\mathbb{R}^d} M_\varepsilon(x,\xi) d\xi = 1, \qquad \int_{\mathbb{R}^d} \xi M_\varepsilon(x,\xi) d\xi = \varepsilon U(x), \qquad \int_{\mathbb{R}^d} \xi_i \xi_j M_\varepsilon(x,\xi) d\xi = A_{ij}(x). \tag{9.16}$$

and, as $\varepsilon \to 0$, we assume the strong convergence in L^1, with uniform bounds in L^∞,

$$M_\varepsilon(x,\xi) \to M(x,\xi) \qquad |\xi|^2 M_\varepsilon(x,\xi) \to |\xi|^2 M(x,\xi). \tag{9.17}$$

Again, we can have in mind a Gaussian.

Theorem 9.4 *With the assumptions* (9.16)–(9.17)*, the limit $n(x,t)$ of $n_\varepsilon(x,t)$, of solutions of* (9.15)*, satisfies the drift-diffusion equation* (9.14) *and*

$$f_\varepsilon(x,\xi,t) \underset{\varepsilon\to 0}{\rightharpoonup} n(x,t) M(x,\xi).$$

We explain the derivation formally. As in the proof on Theorem 9.3, we observe that

$$\frac{\partial}{\partial t} n_\varepsilon + \mathrm{div}\, J_\varepsilon = 0, \qquad J_\varepsilon(x,t) = \int_{\mathbb{R}^d} \frac{\xi}{\varepsilon} f_\varepsilon(x,\xi,t) d\xi.$$

In order to compute $J_\varepsilon(x,t)$, we rewrite the Eq. (9.15) as

$$\frac{f_\varepsilon}{\varepsilon} = \frac{1}{\varepsilon} n_\varepsilon(x,t) M_\varepsilon(x,\xi) - \frac{\varepsilon}{k(x)} \frac{\partial}{\partial t} f_\varepsilon(x,\xi,t) - \frac{1}{k(x)} \xi \cdot \nabla_x f_\varepsilon,$$

$$\begin{aligned} J_\varepsilon(x,t) &= n_\varepsilon(x,t) \int_{\mathbb{R}^d} \frac{\xi}{\varepsilon} M_\varepsilon(x,\xi) d\xi - \frac{\varepsilon}{k(x)} \frac{\partial}{\partial t} J_\varepsilon(x,t) - \frac{1}{k(x)} \nabla_x \int_{\mathbb{R}^d} \xi \otimes \xi f_\varepsilon d\xi \\ &= n_\varepsilon(x,t) U(x) - \frac{1}{k(x)} \nabla_x \int_{\mathbb{R}^d} \xi \otimes \xi n_\varepsilon M_\varepsilon(x,\xi) d\xi + O(\varepsilon) \\ &= n_\varepsilon(x,t) U(x) - \frac{1}{k(x)} \nabla_x A(x) n_\varepsilon + O(\varepsilon), \end{aligned}$$

because $f_\varepsilon = n_e M_\varepsilon + O(\varepsilon)$.

Therefore, when $\varepsilon \to 0$, we find

$$\frac{\partial}{\partial t} n(x,t) + \text{div } J(x,t) = 0,$$

$$J(x,t) = -\frac{1}{k(x)} \nabla_x A(x)[n(x,t)] + n(x,t) U(x).$$

This gives the result. □

9.5 Bacterial Movement

Bacterial motion and bacterial population self-organization is a wide and fascinating area of biology, which has generated an important literature with the progresses of experimental observations. For instance, several examples are presented in Murray's book [53] showing mechanisms which underlie the pattern formation. The first mechanism is due to interaction with the environment through nutrient or other physical effects [31]. The second communication mechanism used by bacteria is chemotaxis, i.e., the motion of cells directed by a chemical signal. It is the central mechanism for *Escherichia coli*, which has raised an enormous interest since Adler's seminal paper [1], see also [13, 14, 51] and the book [7] for all biological aspects of *E. coli*.

Since the 80's, observations at the cell scale have shown that bacteria as *E. Coli* or *B. Subtilis* move by run and tumble depending on the coordination of motors that control the flagella. To give an idea of scales, the run time is about 1 s, the run length is a few µm, and tumbling takes a much shorter time (1/10 s). To take into account this phenomenon the kinetic formalism is needed and that was proposed, early after the first observations, by Alt and his co-authors [3, 56]. Recent surveys on the subject can be found in [16, 26, 37].

The modeling goes as follows. Denote by $f(x,\xi,t)$ the number density of cells located at $x\in\mathbb{R}^d$ and moving with the velocity $\xi\in V$, usually it is admitted that tumbles are always with the same speed and thus $V=\mathbb{S}^{d-1}$. As in the physical models presented before, we set

$$\begin{cases} \frac{\partial}{\partial t} f(x,\xi,t) + \overbrace{\xi\cdot\nabla_x f}^{\textbf{run}} = \overbrace{\mathscr{K}[c,f]}^{\textbf{tumble}}, \\[2ex] f(x,\xi,t=0) = f^0(x,\xi) \geq 0, \qquad f^0\in L^1. \end{cases} \tag{9.18}$$

Compared to scattering equation (9.1), the novelty here is to take into account the rules leading cells to tumble, which is described but the term

$$\mathscr{K}[c,f] = \underbrace{\int_V K(c;\xi,\xi')f(\xi')d\xi'}_{\text{cells of velocity } \xi' \text{ turning to } \xi} \quad - \quad \underbrace{\int_V K(c;\xi',\xi)d\xi'\; f(x,\xi,t)}_{\text{cells of velocity } \xi \text{ turning to another}} \quad . \tag{9.19}$$

Here $c(x,t)$ describes a molecular environment which modulates cell responses. An important issue is that K may depend functionally on c. For instance it may depend in a non-local way on c, or on derivatives of c. For example, being given a function

$$\Phi\in C^1(\mathbb{R};\mathbb{R}), \qquad 0<\min\Phi(\cdot)<\max\Phi(\cdot)<\infty,$$

we can take a kernel with memory

$$K(c;\xi,\xi') = \Phi\big(c(x-\varepsilon\xi',t)\big), \tag{9.20}$$

which expresses that a cell responds using the average concentration during their run of duration 2ε and using a middle rule for the integration. The function $\Phi(\cdot)$ takes into account possible response modulation to the signal. One may be more accurate and use an integration rule

$$K(c;\xi,\xi') = \Phi\big(\omega * c(x,t)\big)$$

for an appropriate kernel ω and convolution along the path.

This chemical signal $c(\cdot)$ can be emitted by the cells themselves and diffused in the media, then one writes

$$\tau\frac{\partial c}{\partial t} - \Delta c(x,t) = n(x,t) := \int_V f(t,x,\xi)d\xi, \tag{9.21}$$

with $\tau\geq 0$ the diffusion time scale, usually small compared to cells dynamics. But the function $c(\cdot)$ can also be imposed from outside and then, the Eq. (9.18) is linear. Then, it is simply a variant of the scattering equation.

Existence of Solutions Some general properties of solutions are

1. *Non-negativity.* We have for all times $f(x,\xi,t) \geq 0$.
2. *Mass conservation.* We have for all times $t \geq 0$

$$\int_{\mathbb{R}^d \times V} f(x,\xi,t) dx\, d\xi = \int_{\mathbb{R}^d \times V} f^0(x,\xi) dx\, d\xi.$$

   Notice that this property follows from the symmetric form of the tumbling kernel where both $K(c;\xi,\xi')$ and $K(c;\xi',\xi)$ appear.

It is however difficult to draw more elaborate conclusions in terms of a priori bounds, in particular when the kinetic equation and the diffusion equation for chemoattractant are coupled. This difficulty opened the route to several existence results, which are still not complete in full generality and in particular when the tumbling kernel depends on ∇c, see [10, 12, 23, 41]. Blow-up of solutions, under certain large mass conditions when the tumbling kernel depends on ∇c, has been obtained in [11].

Here we state a result from [23] which proves global existence of locally (in time) bounded solutions (thus extending a result in [36] in the linear case).

The existence theory for the nonlinear system (9.18)–(9.20) was settled in [23] and yields the following:

Theorem 9.5 ([23]) *In dimension $d = 3$, assume that V is bounded and that $f_0 \in L^\infty(\mathbb{R}^d \times V)$, then there is a unique solution to the system (9.18)–(9.20), $f \in C\big([0,\infty); L^1(\mathbb{R}^d \times V)\big)$, moreover we have for all $T > 0$ and $0 \leq t \leq T$,*

$$0 \leq f(t,x,\xi) \leq C(T),$$

$$\|\nabla c(t)\|_{L^p(\mathbb{R}^d)} \leq C(T), \qquad \frac{d}{d-1} < p \leq \infty,$$

$$\|c(t)\|_{L^p(\mathbb{R}^d)} \leq C(T), \qquad d < p \leq \infty,$$

for some constant $C(T)$.

The proof is based on the dispersion estimates of Sect. 9.2.

This result provides global strong solutions and therefore shows a fundamental difference with the macroscopic model Patlak/Keller–Segel equation (see below) since the latter exhibits blow-up. This is rather counter-intuitive since we can expect that solutions to a hyperbolic equation have weaker estimates than the related parabolic equation.

The parabolic equation on c can be treated as well and several extensions of Theorem 9.5 have been obtained. Also specific dependency upon ∇c in the tumbling kernel have been used, see [10–12, 41] as well as various other extensions [38, 49, 82].

The Patlak/Keller–Segel System Using the small memory parameter ε introduced in the Eq. (9.20) for the tumbling rate, one can rescale space and time the kinetic equation while keeping the diffusion equation for the chemoattractant (this is questionable and a more complete analysis should use the parameter range in the chemoattractant equation to justify the reasoning). Under the symmetry assumption on V that

$$\int_V \xi d\xi = 0, \tag{9.22}$$

we may use the diffusive rescaling as in Sect. 9.4. We arrive at

$$\begin{cases} \varepsilon \frac{\partial}{\partial t} f_\varepsilon(x,\xi,t) + \xi \cdot \nabla_x f_\varepsilon = \frac{1}{\varepsilon}\mathscr{K}[c_\varepsilon, f_\varepsilon], \\ f_\varepsilon(x,\xi,t=0) = f^0(x,\xi) \geq 0, \qquad f^0 \in L^1. \end{cases} \tag{9.23}$$

The same method allows to derive a macroscopic equation as $\varepsilon \to 0$ and get the nonlinear Fokker–Planck equation (see again [23])

$$\frac{\partial}{\partial t} n(x,t) - \text{div}[D(c)\nabla n] + \text{div}[n\chi(c)\nabla c] = 0, \tag{9.24}$$

$$\tau \partial_t c(x,t) - \Delta c = n(x,t), \tag{9.25}$$

with the transport coefficients

$$D(c) = \frac{1}{|V|^2} \frac{\int_{V\times V} \xi \otimes \xi d\xi}{\Phi(c)}, \qquad \chi(c) = \frac{1}{|V|^2} \int_{V\times V} \xi \otimes \xi d\xi \frac{\Phi'(c)}{\Phi(c)}.$$

This system is called the Patlak/Keller–Segel system. It has been widely studied usually with D and χ constant . We refer to the Sect. 9.6 for further results around this system.

Proof Let us show the formal derivation again, assuming convergence of all functions. First, we identify

$$\int_V [\Phi(c(x-\varepsilon\xi',t)) f_\varepsilon(x,\xi',t) - \Phi(c(x-\varepsilon\xi,t)) f_\varepsilon(x,\xi,t)] d\xi' = \varepsilon^2 \frac{\partial}{\partial t} f_\varepsilon(x,\xi,t) + \varepsilon\xi \cdot \nabla_x f_\varepsilon,$$

therefore as $\varepsilon \to 0$ we find

$$\int_V [\Phi(c(x,t)) f(x,\xi',t) - \Phi(c(x,t)) f(x,\xi,t)] d\xi' = 0$$

which means that the limit f of f_ε is independent of ξ and thus satisfies

$$f(x,\xi,t) = \frac{1}{|V|} n(x,t), \qquad \xi \in V.$$

As a second step we write as usual

$$\frac{\partial}{\partial t} n_\varepsilon(x,t) + \operatorname{div} J_\varepsilon(x,t) = 0, \qquad J_\varepsilon(x,t) = \frac{1}{\varepsilon}\int_V \xi f_{:} e(x,\xi,t) d\xi.$$

The third step is to compute J_ε which we do in using the Eq. (9.23) as follows. After multiplication by a component ξ_j and integrating in ξ, we have

$$\frac{1}{\varepsilon}\int_{V\times V} [\Phi(c(x-\varepsilon\xi',t))\xi_j f_\varepsilon(x,\xi',t) - \Phi(c(x-\varepsilon\xi,t))\xi_j f_\varepsilon(x,\xi,t)] d\xi' d\xi$$
$$= \varepsilon\frac{\partial}{\partial t}\int_V \xi_j f_\varepsilon(x,\xi,t) d\xi + \int_V \xi\xi_j \cdot \nabla_x f_\varepsilon d\xi.$$

We neglect the terms in ε and get, using (9.22),

$$-\frac{1}{\varepsilon}\int_{V\times V} \Phi(c(x-\varepsilon\xi,t))\xi_j f_\varepsilon(x,\xi,t)] d\xi' d\xi = \frac{1}{|V|}\int_V \xi\xi_j \cdot \nabla_x n d\xi + O(\varepsilon)$$

and expansion of $\Phi(c(x-\varepsilon\xi,t))$ gives

$$-\Phi(c(x,t))|V|J_\varepsilon(x,t) - |V|\Phi'(c)\nabla_x c(x,t)\int_{V\times V} \xi\xi_j d\xi = \frac{1}{|V|}\int_V \xi\xi_j \cdot \nabla_x n d\xi + O(\varepsilon).$$

This yields the result. □

Blow-Up in the Keller–Segel System In order to present the major property that solutions can blow-up in finite time, we consider the particular case of the Patlak/Keller–Segel system (9.25), under the form

$$\begin{cases} \frac{\partial}{\partial t} n(x,t) - \Delta n(x,t) + \operatorname{div}(n\chi\nabla c) = 0, \quad x \in \mathbb{R}^2, \\ \\ -\Delta c(x,t) = n(x,t). \end{cases} \tag{9.26}$$

We recall the following results, see for instance [8, 59] and the references therein:

Theorem 9.6 *In dimension $d = 2$, for the Patlak/Keller–Segel system with initial data satisfying $\int_{\mathbb{R}^2} n^0[1 + |x|^2 + |\log(n^0)|]dx < \infty$, we have*

(i) for $\|n^0\|_{L^1(R^2)} < \frac{8\pi}{\chi}$ there are smooth solutions of (9.26)*,*
(ii) for $\|n^0\|_{L^1(R^2)} > \frac{8\pi}{\chi}$ solutions blow-up in finite time (as a singular measure),
(iii) for radially symmetric solutions, blow-up means

$$n(t) \approx \frac{8\pi}{\chi}\delta(x=0) + \mathit{Remainder}.$$

A very important literature is devoted to these blow-up phenomena. Among them, let us mention the blow-up result in the parabolic-parabolic case [52].

Both theoretical analysis and numerical simulations show that solutions exhibit Dirac mass singularities. These are the most common patterns exhibited by the Patlak/Keller–Segel system.

This blow-up phenomenon is compatible with observations of the amoeba *Dyctiostelium discoideum* moving on a dish. Chemotaxis leads them to form highly concentrated patterns, which we can interpret as a final stage before they change their compartment and thus the model becomes wrong. At this stage, the cells form a three-dimensional multicellular fruiting body which generates spores that can disperse. But for *E. coli*, and for many other types of cells, such a final stage is not observed and Dirac masses are not a desirable representation of the observations.

Keller–Segel System with Prevention of Overcrowding In order to circumvent this difficulty, a limitation of the drift term can be imposed in order to take into account volume effects, see [37, 67]. The system (9.26) is modified, for instance, as

$$\begin{cases} \frac{\partial}{\partial t} n(x,t) - D\Delta n(x,t) + \operatorname{div}(n\psi(n)\nabla c) = 0, \qquad x \in \mathbb{R}^2,\ t \geq 0, \\ \\ -\Delta c(x,t) + dc(x,t) = n(x,t), \end{cases}$$

with $d > 0$ a degradation rate, $\psi(n)$ a switch for n large, for instance $\psi(n) = e^{-n/n_s}$. Another example is $\psi(n) = n_S - n$, then solutions remain bounded, $n \leq n_s$ when this is true for the initial data thanks to the maximum principle. The paper [67] presents a complete analysis of the parameter range for unstability and for the dynamics of patterns formed by this system.

Adapting [67], we can simply explain why patterns are formed based on the unstability of the constant states $n \equiv \bar{n}$, $S \equiv \bar{n}/d$ when the inequality is satisfied

$$\bar{n}\psi(\bar{n}) > Dd. \tag{9.27}$$

To see that, we look for a growing perturbation $n = \bar{n} + \alpha e^{ikx} e^{\lambda t}$, $S = \frac{\bar{n}}{d} + \beta e^{ikx} e^{\lambda t}$. Inserting the expansion for $\alpha,\ \beta$ small in the equations, gives

$$\begin{cases} \lambda\alpha + D|k|^2\alpha - \bar{n}\psi(\bar{n})\beta|k|^2 = 0, \\ (|k|^2 + d)\beta = \alpha, \end{cases}$$

that is also written (for $\alpha \neq 0$),

$$\lambda = -D|k|^2 + \bar{n}\psi(\bar{n})\frac{|k|^2}{|k|^2 + d}.$$

We find a growing mode $\lambda > 0$ under the stated condition.

9.6 Modulation Along the Path

More realistic kinetic models use the modulation of signal by *E. coli*. It turns out that bacteria increase the jump length when they feel an increasing chemotactic signal and reduce their jumps when the signal decreases along their path, [25, 50]. This leads to change the tumbling rule (9.20) to

$$K(c;\xi,\xi') = \Phi\big(\frac{\partial c}{\partial t} + \xi'.\nabla c\big). \tag{9.28}$$

See [25, 28, 29]. From these papers, we borrow the stiff response case, when

$$\Phi(z) = \begin{cases} k_- & \text{for } z < 0, \\ \\ k_+ < k_- & \text{for } z > 0. \end{cases} \tag{9.29}$$

More generally, $\Phi(\cdot)$ is a (smooth) decreasing function but stiffness is definitively a correct assumption

$$\Phi(z) = \Phi\left(\frac{z}{\delta}\right), \qquad \Phi(\pm\infty) = k_\pm, \tag{9.30}$$

for a ‘small’ constant $\delta > 0$.

Macroscopic Limit Using a tumbling kernel as (9.28) leads to a new class of macroscopic limits.

In [25], the hyperbolic limit is proved, leading to the equation

$$\frac{\partial}{\partial t} n(x,t) + \operatorname{div}\big[n\,\chi(\frac{\partial c}{\partial t}, |\nabla c|)\nabla c\big] = 0,$$

where the sensitivity χ has the form (with some coefficient A and assuming V is radially symmetric)

$$\chi(\frac{\partial c}{\partial t}, |\nabla c|) = \frac{A}{|\nabla c|}\int_V \frac{\xi_1 d\xi}{\Phi\big(\frac{\partial c}{\partial t} + \xi_1|\nabla c|\big)}.$$

The non-negativity of the chemotactic sensitivity is a consequence of the assumption that Φ is decreasing.

An indetermination arises, because of the ratio $\frac{\nabla c}{|\nabla c|}$ when $\nabla c = 0$, which leads to a particularly subtle theory developed in [42, 43].

The first-order correction leads to a parabolic Fokker–Planck equation under the form

$$\frac{\partial}{\partial t} n(x,t) - \operatorname{div}\big[D\nabla n(x,t)\big] + \operatorname{div}\big[n\,\chi(\frac{\partial c}{\partial t}, |\nabla c|)\nabla c\big] = 0$$

(see also [65] for another derivation). This type of equation is called Flux-Limited Keller–Segel because U is bounded and has raised a large interest in the last few years [6, 24, 69, 70]. An important property is that solutions do not blow-up because the drift is bounded, generating patterns which are more relevant than the Dirac concentrations mentioned before.

Traveling Bands It is commonly admitted that chemotaxis is one of the key ingredients triggering the formation of traveling bands (pulses) as observed in Adler's famous experiment for *E. Coli* (1966), [1]. We refer to [74] for a complete review of experimental assays.

Recently a mathematical and quantitative explanation has been developed in [69, 70], using the Flux-Limited Keller–Segel system with nutrient S in dimension $d = 1$

$$\begin{cases} \frac{\partial}{\partial t} n(x,t) - \Delta n(x,t) + \mathrm{div}[n(U_c + U_S)] = 0, \\ U_c = \chi_c \frac{\nabla c}{|\nabla c|}, \qquad U_S = \chi_S \frac{\nabla S}{|\nabla S|}, \\ \frac{\partial c}{\partial t} - D_c \Delta c + \alpha c = \beta n, \qquad \frac{\partial S}{\partial t} - D_S \Delta S = -\gamma n S. \end{cases}$$

Traveling waves are defined as solutions of the form $n(x - \sigma t) > 0$, $c(x - \sigma t)$, $S(x - \sigma t)$ for which $n(\pm\infty = 0)$. The parameter $\sigma \in \mathbb{R}$ is called the traveling speed and is due to the movement toward fresh nutrient S. If S is ignored, then standing pulses are observed in accordance with the experimental observations in [51].

In [69], in the case with stiff response, traveling pulses to this FLKS model are built analytically and they exhibit an asymmetric profile as it is observed experimentally. For a more general response function Φ, it is difficult to assert the existence of pulses. Another important and difficult extension is to assert the existence of traveling or standing bands (pulses) for the kinetic equation; we refer to [9, 16, 18].

Instabilities Chemotaxis is a major phenomenon which produces patterns as observed in nature. Parabolic models as the Patlak/Keller–Segel system have been widely used for such purposes, cf. [37, 54, 67] and the references therein. As mentioned before, the limitation of the drift is a possible mechanism in this direction. Another direction is flux limitation as shown in [17].

There is only a limited literature on instabilities and pattern formation ability of solutions to the kinetic equations of bacterial chemotaxis. We refer to [63] where the stiffness parameter in the tumbling kernel with modulation along the path, that is δ small in (9.30), appears to be a bifurcation parameter.

Biochemical Pathways Still more elaborated and physically more relevant descriptions of bacterial run and tumble movement have been derived lately. These models aim at explaining the cell decision to tumble by an intracellular process of molecular nature [25, 28, 29, 34, 57, 64, 66, 68, 71, 73, 77, 79]. Indeed, bacteria respond to extracellular signal changes through a sophisticated signal transduction pathway. It involves a rapid response of the cell to the external signal change called 'excitation', and a 'slow adaptation' which allows the cell to subtract the background signal.

For our presentation and in order to keep simplicity, we follow closely [64] and consider a single intracellular variable, $m \in \mathbb{R}$ which represents a cell receptor methylation level. We assume a reaction rate equation for the intracellular adaptation dynamics

$$\frac{dm}{dt} = R\big(m, M(x,t)\big), \tag{9.31}$$

where R describes the chemical reaction dynamics. We call M the equilibrium of this reaction which is itself determined by a nonlinear processing of the external signal, the chemoattractant $c(x,t)$, for *E. coli* see [45]. Usually, a logarithmic rule is used

$$M(x,t) \equiv \mathscr{M}_0 \ln(c(x,t)).$$

A condition which ensures that Eq. (9.31) admits $M(x,t)$ as an attractive equilibrium point is

$$R\big(m, M\big) > 0 \quad \text{for } m < M, \qquad R\big(m, M\big) < 0 \quad \text{for } m > M. \tag{9.32}$$

In this section, we simplify again the formalism by assuming that M itself is given. Then, we can write the kinetic-transport equation which governs the dynamic of the number density function $p(x,\xi,m,t)$ of bacteria at time t, position $x \in \mathbb{R}^d$, moving at velocity $\xi \in V$ and methylation level $m \in \mathbb{R}$

$$\begin{cases} \frac{\partial}{\partial t} p + \xi \cdot \nabla_x p + \frac{\partial}{\partial m}[R(m,M)p] = \mathscr{Q}[m,M](p), \\ p(x,\xi,m,t=0) = p^0(x,\xi,m) \geq 0, \qquad p^0(x,\xi,m) \in L^1 \cap L^\infty(\mathbb{R}^d \times V \times \mathbb{R}). \end{cases} \tag{9.33}$$

The tumbling term $\mathscr{Q}[m,M](p)$ describes the velocity jump process, it is given by

$$\mathscr{Q}[m,M](p) = \int_V \big[\lambda(m,M,\xi,\xi')p(t,x,\xi',m) - \lambda(m,M,\xi',\xi)p(t,x,\xi,m)\big]\, d\xi', \tag{9.34}$$

where $\lambda(m,M,\xi,\xi')$ denotes the methylation-dependent tumbling frequency from ξ' to ξ, in other words the response of the cell depending on its environment and internal state. We borrow this formalism from [44, 72] even though this type of models, involving more general signal transduction, can be traced back to [25, 28, 29, 78].

Notice that this kind of modeling does not relate to Vlasov type of equation as in Sect. 9.1 but rather to variants of the Boltzmann equation where the particles radii can vary (clouds of droplets arising in the design of various types of engines, from diesel engines to rocket thrusters). As a consequence, the macroscopic density is defined as

$$n(x,t)=\int_{\mathbb{R}}\int_{V} p(x,\xi,m,t)d\xi dm.$$

One can also define a density in the phase space as

$$f(x,\xi,t)=\int_{\mathbb{R}} p(x,\xi,m,t)dm.$$

Fluid Limit A complete proof of existence for the nonlinear system where Eq. (9.33) is coupled to a diffusion equation for the chemoattractant produced by the cells is available in[48].

In the direction of deriving fluid equations, the authors in [25, 28, 29, 55, 72, 78, 79] developed the asymptotic theory which, departing from the kinetic level of description, allows to recover, in the diffusion and in the hyperbolic limits, macroscopic equations where the variables are only (x,t) as the Keller–Segel system which governs the dynamics of the density of cells.

From Molecular Pathways to Modulation Along the Path The present formalism based on a molecular content can also be used to derive the tumbling kernel (9.28) with modulation along the pathways. This is based on an asymptotic analysis introduced in [64] which differs from the diffusive or hyperbolic rescaling.

According to [64], we consider the equation with a fast adaptation to the external signal, which means a fast reaction rate R,

$$\frac{\partial}{\partial t}p_\varepsilon(t,x,\xi,m)+\xi\cdot\nabla_x p_\varepsilon+\frac{1}{\varepsilon}\frac{\partial}{\partial m}\Big[R\big(m,M(x,t)\big)p_\varepsilon\Big]=\mathcal{Q}_\varepsilon[m,M][p_\varepsilon].$$

The other rescaling used in [64] arises in the tumbling kernel, which we also simplify by ignoring the dependency in ξ,ξ', and writes

$$\mathcal{Q}_\varepsilon[m,M][p]=\int_{\xi'}\Big[\lambda\big(\frac{m-M(x,t)}{\varepsilon}\big)p(t,x,\xi',m)-\lambda\big(\frac{m-M(x,t)}{\varepsilon}\big)p(t,x,\xi,m)\Big]d\xi', \tag{9.35}$$

that can be interpreted as a stiff response.

The main result, in [64], is the following:

Theorem 9.7 *As $\varepsilon\to0$, in the weak sense of measures,*

$$p_\varepsilon(t,x,\xi,m)\to f(t,x,\xi)\ \delta\big(m-M(x,t)\big)$$

and f satisfies the kinetic-transport equation (9.18) *with the tumbling kernel*

$$K(c(x,t);\xi,\xi') = \lambda\big(\frac{\partial M}{\partial t} + \xi'.\nabla M\big), \tag{9.36}$$

and with initial data $f^0(x,\xi) = \int_{\mathbb{R}} p^0(x,\xi,m)dm$.

Notice that, surprisingly, even though we start from a tumbling kernel $\lambda\big(\frac{m-M(x,t)}{\varepsilon}\big)$ independent of the cell velocity, the outcome is ξ dependent.

Traditional models of kinetic theory in the introduction always undergo nice geometrical structures inherited from physical invariants. The tumbling kernel (9.36) is not of that type by opposition to the simple molecular dependent kernel (9.35). Therefore, it is satisfactory that the form (9.36) can be derived rigorously.

A last observation is a discussion in [64] about the scales for *E. coli* which do not necessarily correspond to measurements.

Proof We only indicate very roughly the method and difficulty.

In order to identify the limit, following [28, 29], we introduce a new variable

$$p_\varepsilon(t,x,\xi,m) = \varepsilon q(t,x,\xi,\frac{m-M(c)}{\varepsilon}), \qquad y = \frac{m-M(x,t)}{\varepsilon}.$$

To simplify the notations, we assume that $R(m,M) = (m-M)\ G(m-M)$. The Eq. (9.33) becomes

$$\tfrac{\partial}{\partial t}q(t,x,\xi,y) + \xi\cdot\nabla_x q + \tfrac{1}{\varepsilon}\tfrac{\partial}{\partial y}[yG(\varepsilon y)q] = \frac{1}{\epsilon}D_t M\partial_y q$$

$$+\int \lambda(y)\big[q(t,x,\xi',y) - q(t,x,\xi,y)\big]d\xi'.$$

This can be further written

$$\frac{\partial}{\partial y}\big[\big(yG(\varepsilon y) - D_t M\big)q\big] = O(\varepsilon).$$

Because $G(\cdot) > 0$, a property inherited from the assumption (9.32), we infer that q should converge to a Dirac mass at the equilibrium point, that is

$$q \longrightarrow f(x,\xi,t)\delta(y - \frac{D_t M}{G(0)}),$$

for some weight $f(x,\xi,t)$ we can identify by integrating the Eq. (9.33). However, the resulting equation, there is a difficulty to identify the limit of $\mathscr{Q}_\varepsilon[m,M][p_\varepsilon]$ which leads to another change of variables and motivates the scaling used. We refer the interested reader to [64]. □

9.7 Conclusion

This presentation has been focused on some modeling and mathematical aspects of chemotaxis at the mesoscale level with a special emphasis on the questions of asymptotic analysis. It has been reduced to bacterial movement, more precisely *E. coli*. Extensions to multispecies are treated in [2, 27]. Models for motion of cells in the extracellular tissue also use the kinetic formalism, see [35, 40].

Several important subjects are not treated or are just mentioned vaguely. Among them, the question of instabilities has been mentioned very roughly. The physical approaches are also very important also and I have decided not to enter this field because the number of references from the mathematical side was already too large. Another subject of importance, which deserves be advocated, is numerical methods. Some references are [75, 80, 81].

The understanding of detailed motion of cells and the behaviors of different cells are so rich that we can expect many additional models and questions will still arise.

References

1. J. Adler, Chemotaxis in bacteria. Science **153**, 708–716 (1966)
2. L. Almeida, C. Emako, N. Vauchelet, Existence and diffusive limit of a two-species kinetic model of chemotaxis. Kinet. Relat. Models **8**, 359 (2015)
3. W. Alt, Biased random walk models for chemotaxis and related diffusion approximations. J. Math. Biol. **9**, 147–177 (1980)
4. C. Bardos, R. Santos, R. Sentis, Diffusion approximation and computation of the critical size. Trans. Am. Math. Soc. **284**(2), 617–649 (1984)
5. C. Bardos, P. Degond, Global existence for the Vlasov-Poisson equation in 3 space variables with small initial data. Ann. Inst. H. Poincaré Anal. Non Linéaire **2**(2), 101–118 (1985)
6. N. Bellomo, M. Winkler, A degenerate chemotaxis system with flux limitation: maximally extended solutions and absence of gradient blow-up. Commun. Part. Differ. Equ. **42**, 436–473 (2017)
7. H.C. Berg, *E. coli in Motion* (Springer, Berlin, 2004)
8. A. Blanchet, J. Dolbeault, B. Perthame, Two-dimensional Keller-Segel model: optimal critical mass and qualitative properties of the solutions. Electron. J. Differ. Equ. **2006**(44), 1–32 (2006)
9. E. Bouin, V. Calvez, G. Nadin, Propagation in a kinetic reaction-transport equation: travelling waves and accelerating fronts. Arch. Ration. Mech. Anal. **217**(2), 571–617(2015)
10. N. Bournaveas, V. Calvez, Global existence for the kinetic chemotaxis model without pointwise memory effects, and including internal variables. Kinet. Relat. Models **1**(1), 29–48 (2008)
11. N. Bournaveas, V. Calvez, Critical mass phenomenon for a chemotaxis kinetic model with spherically symmetric initial data. Ann. Inst. H. Poincaré Anal. Non Linéaire **26**(5), 1871–1895 (2009)
12. N. Bournaveas, V. Calvez, S. Gutièrrez, B. Perthame, Global existence for a kinetic model of chemotaxis via dispersion and Strichartz estimates. Commun. Partial Differ. Equ. **33**, 79–95 (2008)
13. M.P. Brenner, L.S. Levitov, E.O. Budrene, Physical mechanisms for chemotactic pattern formation by bacteria. Biophys J. **74**, 1677–1693 (1998)
14. E.O. Budrene, H.C. Berg, Dynamics of formation of symmetrical patterns by chemotactic bacteria. Nature **376**, 49–53 (1995)

15. D. Cai, L. Tao, M. Shelley, D.W. McLaughlin, An effective kinetic representation of fluctuation-driven neuronal networks with application to simple and complex cells in visual cortex. PNAS **101**, 7757–7762 (2004)
16. V. Calvez, Chemotactic waves of bacteria at the mesoscale (2016), arXiv:1607.00429
17. V. Calvez, B. Perthame, S. Yasuda, Traveling wave and aggregation in a flux-limited Keller-Segel model (2017, preprint), arXiv:1709.07296
18. V. Calvez, G. Raoul, C. Schmeiser, Confinement by biased velocity jumps: aggregation of *Escherichia coli*. Kinet. Relat. Models **8**(4), 651–666 (2015)
19. J. Caron, J.-L. Feugeas, B. Dubroca, G. Kantor, C. Dejean, T. Pichard, Ph. Nicolaï, E. D'Humières, M. Frank, V. Tikhonchuk, Deterministic model for the transport of energetic particles: application in the electron radiotherapy. Phys. Med. **31**(8), 912–921 (2015)
20. F. Castella, B. Perthame, Estimations de Strichartz pour les équations de transport cinétique. (French) [Strichartz' estimates for kinetic transport equations] C. R. Acad. Sci. Paris Sér. I Math. **322**(6), 535–540 (1996)
21. C. Cercignani, *The Boltzmann Equation and Its Applications*. Applied Mathematical Sciences, vol. 67 (Springer, New York, 1988), xii+455 pp
22. C. Cercignani, R. Illner, M. Pulvirenti, *The Mathematical Theory of Dilute Gases*. Applied Mathematical Sciences, vol. 106 (Springer, New York, 1994), viii+347 pp
23. F. Chalub, P.A. Markowich, B. Perthame, C. Schmeiser, Kinetic models for chemotaxis and their drift-diffusion limits. Monatsh. Math. **142**, 123–141 (2004)
24. A. Chertock, A. Kurganov, X. Wang, Y. Wu, On a chemotaxis model with saturated chemotactic flux. Kinet. Relat. Models **5**, 51–95 (2012)
25. Y. Dolak, C. Schmeiser, Kinetic models for chemotaxis: hydrodynamic limits and spatio-temporal mechanisms. J. Math. Biol. **51**, 595–615 (2005)
26. R. Eftimie, Hyperbolic and kinetic models for self-organized biological aggregations and movement: a brief review. J. Math. Biol. **65**, 35–75 (2012)
27. C. Emako, C. Gayrard, A. Buguin, L. Almeida, N. Vauchelet, Traveling pulses for a two species chemotaxis model. PLoS Comput. Biol. **12**, e1004843 (2016)
28. R. Erban, H. Othmer, From individual to collective behaviour in bacterial chemotaxis. SIAM J. Appl. Math. **65**(2), 361–391 (2004)
29. R. Erban, H. Othmer, Taxis equations for amoeboid cells. J. Math. Biol. **54**, 847–885 (2007)
30. R. Glassey, *The Cauchy Problem in Kinetic Theory* (SIAM, Philadelphia, 1996)
31. I. Golding, Y. Kozlovski, I. Cohen, E. BenJacob, Studies of bacterial branching growth using reaction-diffusion models for colonial development. Phys. A **260**, 510–554 (1998)
32. F. Golse, Fluid dynamic limits of the kinetic theory of gases, in *From Particle Systems to Partial Differential Equations*. Springer Proceedings in Mathematics & Statistics, vol. 75 (Springer, Heidelberg, 2014), pp. 3–91
33. F. Golse, P.-L. Lions, B. Perthame, R. Sentis, Regularity of the moments of the solution of a transport equation. J. Funct. Anal. **76**(1), 110–125 (1988)
34. G.L. Hazelbauer, Bacterial chemotaxis: the early years of molecular studies. Annu. Rev. Microbiol. **66**, 285–303 (2012)
35. T. Hillen, M5 mesoscopic and macroscopic models for mesenchymal motion. J. Math. Biol. **53**, 585–616 (2006)
36. T. Hillen, H.G. Othmer, The diffusion limit of transport equations derived from velocity-jump processes. SIAM J. Appl. Math. **61**, 751–775 (2000)
37. T. Hillen, K.J. Painter, A user's guide to PDE models for chemotaxis. J. Math. Biol. **58**, 183–217 (2009)
38. T. Hillen, K. Painter, Transport and anisotropic diffusion models for movement in oriented habitats, in *Dispersal, Individual Movement and Spatial Ecology: A mathematical perspective*, ed. by M.A. Lewis, P. Maini, S. Petrowskii (Springer, Heidelberg, 2012), pp. 177–222
39. T. Hillen, A. Swan, The diffusion limit of transport equations in biology, in *Mathematical Models and Methods for Living Systems*. Lecture Notes in Mathematics, vol. 2167, Fond. CIME/CIME Foundation Subseries (Springer, Cham, 2016), pp. 73–129

40. T. Hillen, P. Hinow, Z. Wang, Mathematical analysis of a kinetic model for cell movement in network tissues. Discrete Contin. Dyn. Syst. Ser. B **14**(3), 1055–1080 (2010)
41. H.J. Hwang, K. Kang, A. Stevens, Global solutions of nonlinear transport equations for chemosensitive movement. SIAM. J. Math. Anal. **36**, 1177–1199 (2005)
42. F. James, N. Vauchelet, Chemotaxis: from kinetic equations to aggregate dynamics. Nonlinear Differ. Equ. Appl. **20**(1), 101–127 (2013)
43. F. James, N. Vauchelet, Equivalence between duality and gradient flow solutions for one-dimensional aggregation equations. Discrete Contin. Dyn. Syst. **36**(3), 1355–1382 (2016)
44. L. Jiang, Q. Ouyang, Y. Tu, Quantitative modeling of *Escherichia coli* chemotactic motion in environments varying in space and time. PLoS Comput. Biol. **6**, e1000735 (2010)
45. Y.V. Kalinin, L. Jiang, Y. Tu, M. Wu, Logarithmic sensing in *Escherichia coli* bacterial chemotaxis. Biophys. J. **96**(6), 2439–2448 (2009)
46. M. Keel, T. Tao, Endpoint Strichartz estimates. Am. J. Math. **120**(5), 955–980 (1998)
47. E.F. Keller, L.A. Segel, Traveling bands of chemotactic bacteria: a theoretical analysis. J. Theor. Biol. **30**, 235–248 (1971)
48. J. Liao, Global solution for a kinetic chemotaxis model with internal dynamics and its fast adaptation limit. J. Differ. Equ. **259**(11), 6432–6458 (2015)
49. J.T. Locsei, Persistence of direction increases the drift velocity of run and tumble chemotaxis. J. Math. Biol. **55**(1), 41–60 (2007)
50. B. Mazzag, I. Zhulin, A. Mogilner, Model of bacterial band formation in aerotaxis. Biophys. J. **85**, 3558–3574 (2003)
51. N. Mittal, E.O. Budrene, M.P. Brenner, A. Van Oudenaarden, Motility of *Escherichia coli* cells in clusters formed by chemotactic aggregation. Proc. Natl. Acad. Sci. U. S. A. **100**, 13259–13263 (2003)
52. M. Mizoguchi, M. Winkler, Blow-up in the two-dimensional parabolic Keller-Segel system. Personal communication
53. J.D. Murray, *Mathematical Biology*, vol. 2, 2nd edn. (Springer, Berlin, 2002)
54. G. Nadin, B. Perthame, L. Ryzhik, Traveling waves for the Keller-Segel system with Fisher birth terms. Interface Free Bound. **10**, 517–538 (2008)
55. H.G. Othmer, T. Hillen, The diffusion limit of transport equations II: chemotaxis equations. SIAM J. Appl. Math. **62**, 122–1250 (2002)
56. H. Othmer, S. Dunbar, W. Alt, Models of dispersal in biological systems. J. Math. Biol. **26**, 263–298 (1988)
57. H.G. Othmer, X. Xin, C. Xue, Excitation and adaptation in bacteria-a model signal transduction system that controls taxis and spatial pattern formation. Int. J. Mol. Sci. **14**(5), 9205–9248 (2013)
58. B. Perthame, Mathematics tools for kinetic equations. Bull. Am. Math. Soc. **41**(2), 205–244 (2004)
59. B. Perthame, *Transport Equations in Biology*. Frontiers in Mathematics (Birkhäuser, Basel, 2007), x+198 pp
60. B. Perthame, *Parabolic Equations in Biology. Growth, Reaction, Movement and Diffusion*. Lecture Notes on Mathematical Modelling in the Life Sciences (Springer, Cham, 2015), xii+199 pp
61. B. Perthame, D. Salort, On a voltage-conductance kinetic system for integrate & fire neural networks. Kinet. Relat. Models **6**(4), 841–864 (2013)
62. B. Perthame, P.E. Souganidis, A limiting case for velocity averaging. Ann. Sci. école Norm. Sup. (4) **31**(4), 591–598 (1998)
63. B. Perthame, S. Yasuda, Stiff-response-induced instability for chemotactic bacteria and flux-limited Keller-Segel equation (2017, preprint), arXiv:1703.08386
64. B. Perthame, M. Tang, N. Vauchelet, Derivation of the bacterial run-and-tumble kinetic equation from a model with biochemical pathway. J. Math. Biol. **73**(5), 1161–1178 (2016)
65. B. Perthame, Z. Wang, N. Vauchelet, Modulation of stiff response in *E. coli* bacterial motion. Revista Matemática Iberoamericana. In press

66. S.L. Porter, G.H. Wadhams, J.P. Armitage, Rhodobacter sphaeroides: complexity in chemotactic signalling. Trends Microbiol. **16**(6), 251–260 (2008)
67. A.B. Potapov, T. Hillen, Metastability in chemotaxis model. J. Dyn. Differ. Equ. **17**(2), 293–330 (2005)
68. C.V. Rao, J.R. Kirby, A.P. Arkin, Design and diversity in bacterial chemotaxis: a comparative study in Escherichia coli and Bacillus subtilis. PLoS Biol **2**(2), E49 (2004)
69. J. Saragosti, V. Calvez, N. Bournaveas, A. Buguin, P. Silberzan, B. Perthame, Mathematical description of bacterial traveling pulses. PLoS Comput Biol. **6**(8), e1000890 (2010)
70. J. Saragosti, V. Calvez, N. Bournaveas, B. Perthame, A. Buguin, P. Silberzan, Directional persistence of chemotactic bacteria in a traveling concentration wave. Proc. Natl. Acad. Sci. **108**(39), 16235–16240 (2011)
71. G. Si, T. Wu, Q. Ouyang, Y. Tu, A pathway-based mean-field model for *Escherichia coli* chemotaxis. Phys. Rev. Lett. **109**, 048101 (2012)
72. G. Si, M. Tang, X. Yang, A pathway-based mean-field model for E. coli chemo- taxis: mathematical derivation and Keller-Segel limit. Multiscale Model Simul. **12**(2), 907–926 (2014)
73. Y. Tu, T.S. Shimizu, H.C. Berg, Modeling the chemotactic response of *Escherichia coli* to time-varying stimuli. Proc. Natl. Acad. Sci. U. S. A. **105**(39), 14855–14860 (2008)
74. M.J. Tindall, P.K. Maini, S.L. Porter, J.P. Armitage, Overview of mathematical approaches used to model bacterial chemotaxis II: bacterial populations. Bull. Math. Biol. **70**, 1570–1607 (2008)
75. N. Vauchelet, Numerical simulation of a kinetic model for chemotaxis. Kinet. Relat. Models **3**(3), 501–528 (2010)
76. C. Villani, A review of mathematical topics in collisional kinetic theory, in *Handbook of Mathematical Fluid Dynamics*, ed. by S. Friedlander, D. Serre (Elsevier, Amsterdam, 2002)
77. X. Xin, H.G. Othmer, A trimer of dimers-based model for the chemotactic signal transduction network in bacterial chemotaxis. Bull. Math. Biol. **74**(10), 2339–2382 (2012)
78. C. Xue, H.G. Othmer, Multiscale models of taxis-driven patterning in bacterial populations. SIAM J. Appl. Math. **70**(1), 133–167 (2009)
79. C. Xue Macroscopic equations for bacterial chemotaxis: integration of detailed biochemistry of cell signaling. J. Math. Biol. **70**, 1–44 (2015)
80. C. Yang, F. Filbet, Numerical simulations of kinetic models for chemotaxis. SIAM J. Sci. Comput. **36**, B348 (2014)
81. S. Yasuda, Monte Carlo simulation for kinetic chemotaxis model: an application to the traveling population wave. J. Comput. Phys. **330**, 1022–1042 (2017)
82. X. Zhu, G. Si, N. Deng, Q. Ouyang, T. Wu, Z. He, L. Jiang, C. Luo, Y. Tu, Frequency-dependent *Escherichia coli* chemotaxis behavior. Phys. Rev. Lett. **108**, 128101 (2012)

Index

A. Bianchi et al. (eds.), *The Dynamics of Biological Systems*, Mathematics of Planet Earth 4, https://doi.org/10.1007/978-3-030-22583-4