Modeling ecological and evolutionary processes in spatially structured populations

Chaozhi Zheng

Department of Biological and Environmental Sciences
University of Helsinki
Finland

Academic dissertation

To be presented, with permission of the Faculty of Biosciences of the University of Helsinki, for public criticism in Auditorium 5 of Viikki Building B (Latokartanonkaari 7) on April 24th 2009 at 12 o’clock noon

Helsinki 2009
Modeling ecological and evolutionary processes in spatially structured populations

Chaozhi Zheng

The thesis is based on the following articles, which are referred to in the text by their Roman numerals:


**Table of contributions**

The following table shows the major contributions of authors to the original articles.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original idea</td>
<td>OO</td>
<td>OO, CZ</td>
<td>IH</td>
<td>WW, OO, CZ</td>
</tr>
<tr>
<td>Empirical data</td>
<td>MS, IH</td>
<td>-</td>
<td>IH</td>
<td>SH, WW</td>
</tr>
<tr>
<td>Modelling and data analysis</td>
<td>CZ, OO</td>
<td>CZ, JP, OO</td>
<td>CZ, OO, IH</td>
<td>CZ, OO</td>
</tr>
<tr>
<td>Manuscript preparation</td>
<td>CZ, OO, MS, IH</td>
<td>CZ, OO</td>
<td>CZ, OO, IH</td>
<td>CZ, OO, WW</td>
</tr>
</tbody>
</table>


---

**Supervised by**

Dr Otso Ovaskainen  
University of Helsinki  
Finland

**Reviewed by**

Dr Bob O’Hara  
University of Helsinki  
Finland

Prof Juha Alho  
University of Joensuu  
Finland

**Examined by**

Prof Steve Buckland  
University of St Andrews  
UK
# Contents

Table of contributions ........................................................................................................ 4

Table of contents .............................................................................................................. 5

Summary ............................................................................................................................ 7

1 Introduction .................................................................................................................. 7

2 Aims of the thesis ......................................................................................................... 9

3 Study systems and empirical data ................................................................................ 10
   The Glanville fritillary butterfly in the Åland Islands (I, II, III) .................................. 10
   Two specialist aphids sharing a hierarchically distributed host plant (IV) ................. 11

4 Methodologies .............................................................................................................. 12
   Bayesian inference (I, IV) ......................................................................................... 12
   Diffusion theory and animal movement (II) .............................................................. 13
   Individual based simulations (III) ............................................................................. 15

5 Main results and discussion ......................................................................................... 15
   Factors affecting the death rate in the Glanville fritillary butterfly (I) ....................... 15
   Dispersal in highly fragmented landscapes (II) ......................................................... 16
   Evolution of dispersal in the Glanville fritillary metapopulation (III) ...................... 16
   Hierarchical metapopulation dynamics of two species of aphid (IV) ....................... 17

6 Synthesis and perspectives .......................................................................................... 18

7 Acknowledgements ..................................................................................................... 19

8 References .................................................................................................................. 20

I  Age-dependent survival analyzed with Bayesian models of mark-recapture data ......................... 25

II Modelling dispersal with diffusion and habitat selection: analytical results for highly fragmented landscapes .............................................................................................................. 45

III Modelling single nucleotide effects in *phosphoglucone isomerase* on dispersal in the Glanville fritillary butterfly: coupling of ecological and evolutionary dynamics .......... 67

IV Hierarchical metapopulation dynamics of two aphid species on a shared host plant .......................................................................................................................... 93
Summary

Chaozhi Zheng
Metapopulation Research Group, Department of Biological and Environmental Sciences, PO Box 65, F-00014, University of Helsinki, Finland

1 Introduction

The ecological and evolutionary dynamics of species and populations are affected by a complex interplay between endogenous and exogenous factors (Fig. 1). Inter- and intra-specific interactions, as well as the act of natural selection, take primarily place at the level of individuals. Individuals in population biology can be considered analogous to the particles in statistical physics, where the challenge is to start from the fundamental laws describing the behavior of the particles, and to derive the emergent properties for assemblages of interacting particles (Pathria 1996). Similarly, the focus in ecological research is often at the level of populations or collections of populations, the properties of which result from the interplay between environmental factors and the individual level processes. To complete the cycle, the emerging spatio-temporal population structure feeds back to the individuals through density dependent processes or by altering the evolutionary selection pressures.

The complexity of the biological world makes research on it challenging. This thesis is about mathematical biology, i.e., the usage of mathematical methods to learn about biological phenomena. One advantage of using mathematical models is that such models have to be written down in a transparent and exact manner, forcing the researcher to be explicit about his or her assumptions. Often the most difficult part in mathematical biology is in writing down the models, which have to compromise between simplicity and complexity, as the models need to be realistic enough to capture the essential features of the system in hand, but yet simple enough to be able to bring insight into the research questions. Once a model or theory has been formulated, it can be applied in two very different ways. Using a model or theory to predict the outcome of observations corresponds to solving the ‘forward modeling problem.’ The aim of this prospective approach is to understand the causal links between the underlying assumptions and the resulting patterns. The opposite situation, in which one uses observations to infer a model structure and the values of model parameters, can be called an ‘inverse modeling problem’ (Tarantola 2006). The aim of this retrospective approach is to use indirect data to make inference about the underlying mechanisms.

In this thesis, I have applied both types of approaches to study the ecological and evolutionary dynamics of species living in fragmented landscapes, the structure of the landscape resulting either from anthropogenic or from natural processes (Riebesell 1982, Hadley 1987, Saunders et al. 1991, Benton et al. 2003). Habitat destruction, a process involving habitat deterioration, habitat loss, and habitat fragmentation, is widely regarded as one of the major causes of population and species extinction (Benton et al. 2003, Hanski 2007). There is an enormous literature on the ecological, genetic and evolutionary consequences of habitat destruction. Both empirical and theoretical results indicate that habitat loss has large and negative effects on biodiversity (Fahrig 2003), but there has been considerable debate on what are the effects of habitat fragmentation per se (Fahrig 2003, Reed 2004, Johnson and Heck 2006).

For the study of species living in highly fragmented landscapes, the metapopulation approach has become a widely applied paradigm (Hanski 1998). The metapopulation approach stems from population dynamic considerations, simplifying the structure of a landscape into a discrete set of patches of breeding habitat surrounded by unsuitable matrix. The term
metapopulation is broadly defined as a set of spatially separated local populations connected by dispersal of individuals (Hanski and Simberloff 1997). Based on dispersal rates between habitat patches and variations in local extinction risks, Harrison (1991) and Harrison and Taylor (1997) identified four categories of metapopulation dynamics: (1) classical metapopulations persisting in a balance between local extinctions and recolonizations (Levins 1970); (2) mainland-island and source-sink metapopulations where a large or high quality patch supports a permanent population, surrounded by temporary populations; (3) patchy populations where individual move frequently between habitat patches; (4) non-equilibrium populations where species undergo range expansion or decline due to unbalance between extinctions and recolonizations. These categories have provided an effective means of describing spatial dynamics, but real systems may often be intermediate between the categories (Gripenberg et al. 2008). Hanski and Gaggiotti (2004b) emphasized that the primary focus of research should be on the underlying processes, such as migration, local extinction, genetic drift, and so on, rather than directly on the spatial population structure that emerges from the processes. Metapopulation ecology, genetics, and evolution have been comprehensively reviewed by Hanski and Gaggiotti (2004a).

Much of the mathematical metapopulation theory has been developed in the context of stochastic patch occupancy models (SPOM) and their approximations. These approaches play an important role in population and community ecology, population genetics, and evolutionary biology (Ovaskainen and Hanski 2004b). SPOMs assume a network of homogeneous or heterogeneous habitat patches, which have only two possible states, occupied by the focal species or empty. SPOMs have been used to develop spatially realistic metapopulation theory (SMT; Hanski 2001, Hanski and Ovaskainen 2003), which maps colonization and extinction rates to the structure of the landscape, described by the areas, qualities and locations of the habitat fragments. The classic Levins metapopulation theory (Levins 1969, 1970) and island biogeography theory (MacArthur and Wilson 1963, 1967) can be seen as special cases of SMT (Hanski 2001). One of the most important predictions by SPOMs relate to species-specific extinction thresholds, metapopulation persistence depending not only on the total amount of habitat, but also on the spatial configuration of the habitat patch network (Hanski and Ovaskainen 2000, Ovaskainen and Hanski 2004b, Frank 2005).

One of the main limitations with SPOMs is that local population dynamics are not explicitly included. The link between individual-, population- and metapopulation-level dynamics has been in the focus of structured metapopulation models (Hastings and Wolin 1989, Gyllenberg and Hanski 1992, 1997, Lande et al. 1998, 1999, Saether et al. 1999). However, to retain mathematical tractability, these models often make otherwise very simplified assumptions such as identical habitat patches and
global migration. As a bridge between SPOMs and structured metapopulation models, Ovaskainen and Hanski (2004a) built and individual based approach for heterogeneous patch networks. In their model, dispersal was mechanistically modeled through correlated random walk and habitat selection (Ovaskainen and Cornell 2003) instead of dispersal kernels used in SMT. Ovaskainen and Hanski (2004a) derived a SPOM approximation for the individual based model, showing how the individual-level parameters link to metapopulation-level parameters.

One reason for the relatively widespread application of SPOMs is that they can be used to infer metapopulation processes from presence-absence data. Statistical approaches for conducting such analyses include nonlinear or logistic regression (Verboom et al. 1991, Hanski 1994, Eber and Brandl 1996), maximum likelihood (Molilanen 1999, 2000) and Bayesian estimation (O’Hara et al. 2002, Ter Braak and Etienne 2003, Etienne et al. 2004). As SPOMs rely on major simplifications, it is clear that statistical inference relying simply on presence-absence data has limitations and possibilities for biases (Driscoll 2007). For example, parameter estimates may be inaccurate because of poorly estimated patch sizes or unmapped patches (Molilanen 2002). Further, it is difficult to disentangle from presence-absence data whether the observed aggregated pattern of patch occupancy is because of e.g. spatially autocorrelation in extinction risk (caused e.g. by weather patterns), or because of dispersal limitation (Eber and Brandl 1996, Hanski et al. 1996). Other statistical approaches have been used to study how metapopulation processes depend in more detail on habitat quality. For example, analysis of covariance has been used to identify the relationships between patch occupancy and environmental factors (Wettstein and Schmid 1999). Münstbergova and Herben (2004) used patterns of species co-occurrence to identify unoccupied suitable and unsuitable patches by simple frequency calculations.

The metapopulation approach has been applied not only for assessing ecological dynamics, but also for bringing insight into population genetics and evolutionary biology of species inhabiting fragmented landscapes (Hanski and Gaggiotti 2004b). One topic that has received especially much attention is the evolution of dispersal and evolutionary stable dispersal rates. In a fragmented landscape, too little migration decreases colonization rates, but too much migration weakens local populations due to emigration and thus increases extinction rates. Spatially explicit dispersal evolution is difficult to study mathematically, making the use of individual-based simulation models increasingly popular in this field of research (Ronce 2007). While individual-based simulation models lack generality, they can be constructed for a particular system to study specific questions in a realistic manner, given that there is sufficient empirical information available to justify the model structure and to estimate the model parameters. As an example of such an approach, Heino and Hanski (2001) analyzed the evolution of dispersal of the Glanville fritillary butterfly, using the virtual migration model (Hanski et al. 2000) as a sub-model for dispersal.

2 Aims of the thesis

In this thesis I use mathematical modeling to analyze ecological and evolutionary processes and their interactions in spatially structured systems (Fig. 1). The main focus of the thesis is in the construction of biologically realistic models and in their parameterization with empirical data. As the focus of the thesis is in development of mathematical methods, I have not collected new data, but utilized existing data stemming from well studied model systems. The general objective of the thesis is to understand how the interactions between individuals and their spatially structured environments affect ecological and evolutionary processes in fragmented landscapes. Specifically, I address the following questions in the four chapters of the thesis.

Chapter I studies factors affecting the life-span of adult Glanville fritillary butterflies, one of the major components of fitness for short-living insects. Does the population origin of the butterfly influence its survival rate, or more specifically does it make a difference if the individual originates either from a newly-established or from an old population in a metapopulation? What are the effects of age, sex, and day vs. night on individual survival?
Chapter II analyzes dispersal in highly fragmented landscapes. How does the spatial configuration of the patch network affect various characteristics of the dispersal process, such as movement probabilities and mean times spent in patches and in the matrix? Can one overcome the need of complex simulation models by deriving approximate formulae for these quantities, starting from assumptions of random walk and habitat selection for individual movements? How can such dispersal models be used as components of population dynamic models?

Chapter III integrates over a large number of empirical studies on the Glanville fritillary butterfly to extend the dispersal model of Chapter II to cover the full life-cycle and the evolution of dispersal and correlated traits. The model developed in this chapter accounts for empirical results on a particular gene (Phosphoglucose isomerase), which affects the dispersal behaviour and other life-history traits of Glanville fritillary butterflies. Are the model predictions on spatiotemporal patterns of this gene consistent with data? How do allele frequencies relate to properties of individual populations, such as their age and connectivity? How do allele frequencies relate to the properties of entire metapopulations, such as their size? How does the strength of the coupling between ecological and evolutionary dynamics depend on the spatial scale? Are ecological dynamics driving evolutionary dynamics or vice versa, or is the coupling symmetric?

Also Chapter IV examines metapopulation dynamics, but using a different model system, and a modeling approach that is contrasting to those of Chapters I - III. This chapter studies two specialist aphid species sharing a hierarchically structured network of host plants, focusing on the following questions. Can the observed colonization and extinction rates be decomposed into intrinsic rates operating specifically at each spatial scale? How do the extinction and colonization rates depend on the scale considered, and how do the colonization rates depend on the type of the individuals (winged or unwinged aphids, or the offspring of the primary colonizer)? Are there correlations between colonization and extinction, and if so, what can be inferred about the causes of these correlations? Are there observable inter-specific interactions, and if so, what are their consequences on the dynamics and persistence of the aphid populations?

3 Study systems and empirical data

Both of the study systems are spatially structured in a hierarchical manner (Fig. 2). In the Glanville fritillary, individual butterflies constitute local populations on dry meadows, the meadows form aggregated patch networks, and at the level of the entire study systems there is a number of such patch networks. In the aphid system, local populations can be found on the ramets of the plants, there can be several plant groups on an island, and the main study system consists of several islands. What makes both study systems both interesting and challenging is that there are strong interactions between the processes operating at various spatial scales, making them excellent case studies of the overall framework depicted in Fig. 1.

The Glanville fritillary butterfly in the Åland Islands (I, II, III)

The Glanville fritillary butterfly (Melitaea cinxia) has one generation per year. In early June the adults eclose, disperse, mate, and the females lay eggs in batches. The adults die within an average of 1 to 2 weeks depending on weather conditions. The larvae remains as a group until the spring in the following year (Hanski 1999). The butterfly has a classical metapopulation structure in the Åland Islands in southwest Finland, where it persists in a balance between local extinctions and recolonizations in a network of ca 4,000 dry meadows (Hanski 1999, Nieminen et al. 2004). The entire patch network has been classified into sub-networks consisting of dynamically semi-independent metapopulations (Hanski et al. 1996). This system has been studied since 1991 by I. Hanski and his colleagues, and has become a model system of metapopulation dynamics (Ehrlich and Hanski 2004).

Chapter I is based on data that were collected for a population enclosed in a large experimental cage in the year 2003. The larvae were collected from newly-established and old populations in various parts of the metapopulation in the Åland Islands. The butterflies were individually marked and released into the cage after eclosion, and followed in an intensive study throughout their lifetime.
Chapter II is motivated by the structures of habitat patch networks typical to the Glanville fritillary, one of the patch networks used as an example in this chapter. However, the results of Chapter II are of general nature and do not utilize data specific to this or other species.

Chapter III integrates over different kinds of data collected on the Glanville fritillary butterfly metapopulation (Hanski 1999). These studies have produced much empirical data on dispersal, local population dynamics, metapopulation dynamics, and spatial genetic structure, allowing the construction of an individual-based model of dispersal evolution.

**Two specialist aphids sharing a hierarchically distributed host plant (IV)**

This study system consists of two specialist aphid species, *Metopeurum fuscoviride* Stroyan (to be called MF) and *Macrosiphoniella tanacetaria* Kaltenbach (to be called MT). The two species of aphids share the tansy (*Tanacetum vulgare* L., Asteraceae) as their host plant. This system has been intensively studied by W. Weisser and his colleagues (Weisser 2000, Massonnet et al. 2002, Weisser and Härr 2005, Rauch and Weisser 2007). Weisser and Härr (2005) have shown that both the within- and between-year dynamics of this system are characterized by extinction and recolonization events occurring at multiple spatial scales.

Compared to the Glanville fritillary metapopulation in the Åland Islands with annual generations (Chapters I and III), the life-cycles of the aphid species are more complex. Both species are cyclical parthenogens where several asexual generations are followed by a single sexual generation in autumn. As the season progresses during the summer, the number of occupied ramets and plants first increases and then decreases (Weisser 2000, Weisser and Härr 2005). One of the differences between the two aphid species is that species MF is obligatorily ant-tended while MT is not (Flatt and Weisser 2000).

Chapter IV is based on two sets of within-year data, collected in Switzerland (Basel) in 1997 (Weisser 2000) and in Finland (Tvärminne) in 2001 (Weisser and Härr 2005). The host plants in both study areas are distributed hierarchically. A single plant (genet) may consist of up to several hundred ramets (shoots). In the Tvärminne study area, the data originates from eight islands, within each of which the host plant grows in distinct groups. The Basel data originates from two study sites, within which the distribution of the host plants is relatively homogenous. The data describe the spatio-temporal dynamics of the aphids as a time series, the sizes of the aphid populations occupying the ramets being estimated weekly throughout the summer period.
4 Methodologies

The central tools used in this thesis relate to Bayesian inference (I, IV), diffusion theory (II), and individual-based simulations (III). I will next introduce these one by one.

Bayesian inference (I, IV)

Bayesian inference has become popular in many disciplines, largely due to advances in computer technology, which has made it possible to sample high dimensional probability distributions through Markov chain Monte Carlo (MCMC) methods (Gilks et al. 1996). Bayesian inference has been increasingly applied in ecological research over the last decades (Ellison 2004, Clark 2005). See Ellison (2004) for a comparison between Bayesian inference and classical frequentist inference in the context of ecology, and Golstein (2006) and Gelman (2008) for more philosophic discussions of the same topic.

Bayesian inference is a process of using observed data to update the knowledge about model parameters, through Bayes’ formula (Gelman et al. 2004),

\[ p(\theta | y) \sim p(y | \theta) p(\theta). \]  \hspace{1cm} (4-1)

In words, the posterior distribution \( p(\theta | y) \) of parameters \( \theta \) given the data \( y \) is proportional to the product of the likelihood \( p(y | \theta) \) and the prior distribution \( p(\theta) \), the latter reflecting the knowledge about \( \theta \) before the data.

The Bayesian approach can be extended to decompose the complexity of ecology problems into hierarchical levels. Letting the element \( y_{it} \) of the data \( y \) denote the observations for unit \( i \) at time \( t \), and denoting by \( x \) the unobserved variables, the hierarchical Bayes’ formula becomes (Berliner 1996, Wikle 2003, Gelman et al. 2004)

\[ p(\theta_1, \theta_2, x | y) \sim p(y | x, \theta_1) p(x | \theta_2) p(\theta_1, \theta_2). \]  \hspace{1cm} (4-2)

In the data model \( p(y | x, \theta_1) \) the parameter \( \theta_1 \) accounts for uncertainties in observations. There are two possible forms of \( x \). For the components of \( x \) corresponding to unit-specific variables such individual demographic rates, \( p(x | \theta_2) \) account for variation in units such as individuals, populations, species, or spatial locations. For those components of \( x \) that correspond to underlying variables that can’t be directly observed, \( p(x | \theta_2) \) describes the biological process model. The term \( p(\theta_1, \theta_2) \) is the joint prior distribution for \( \theta_1 \) and \( \theta_2 \), where other information than data can be incorporated. If the process \( \{x_i | t = 1, 2, \ldots\} \) is assumed to follow a Markov chain, where future states \( \{x_{t+1}, x_{t+2}, \ldots\} \) are independent of the past state \( \{\ldots, x_{t-2}, x_{t-1}\} \) given the current state \( x_t \), the hierarchical Bayesian model Eq. (4-2) is often called a state-space model (Fig. 3). State space models have recently become very popular in ecology (Jonsen et al. 2003, Rivot et al. 2004, Forester et al. 2007, Patterson et al. 2008).

![Directed graph showing the general structure of state-space models](image)

**Figure 3.** Directed graph showing the general structure of state-space models, where the \( \{y_i\} \) are the observed variables and the \( \{x_i\} \) are the hidden variables representing the unobserved biological process. Each circle represents a random variable, and the arrows denote conditional dependencies.
Given the data \( y \), one can predict an unknown observable \( \tilde{y} \) from the same process. The distribution of \( \tilde{y} \) is called a posterior predictive distribution, and it can be calculated by integrating over the posterior distribution,

\[
p(\tilde{y} | y) = \int p(\tilde{y} | \theta)p(\theta | y) d\theta. \tag{4.3}
\]

In practice, the posterior distribution in Eq. (4.2) is often high dimensional and difficult to sample directly, and thus a MCMC technique is usually applied. The integration in Eq. (4.3) is then done by sampling the posterior distribution obtained by the MCMC.

Linear regression and analysis of variance (ANOVA) are widely used in analyzing ecological data. Both of these can be applied also with Bayesian inference. Classical linear regression can be regarded as a special case of the simple Bayesian model Eq. (4.1), assuming that the likelihood is normally distributed. In simple cases, ANOVA computations can be performed by using linear regression. More generally, ANOVA can be formulated as a hierarchical regression model, to which Bayesian inference can be applied (Gelman 2005).

In Chapters I and IV I apply state-space models, where the unobserved variables have both of the forms discussed above. In Chapter I I develop a hierarchical Bayesian framework for the estimation of death rates and capture probabilities from mark-recapture data, accounting for individual heterogeneity in capture probabilities and survival. The explanatory variables in the model include the effects of age, sex, population type (new vs. old), and day vs. night on survival. In Chapter IV I model hierarchical multi-scale metapopulation dynamics of the two aphid species, decomposing the observed extinction and colonization rates into intrinsic rates operating specifically at each spatial scale. Correlations between colonization and extinction, and interspecific interactions between the two species were also incorporated. This approach is similar to many other metapopulation models (e.g. O’Hara et al. 2002) in the sense that I model separately the colonization and extinction probabilities between each time step.

**Diffusion theory and animal movement (II)**

Dispersal is crucial for all species, but especially for species inhabiting highly fragmented landscapes. In the case of the Glanville fritillary butterflies in the Åland Islands, the habitat fragments cover only a few percents of the total landscape (Chapters I & III, Nieminen et al. 2004). As the local populations are prone to extinction, the species needs to be able to colonize empty patches. Moving substantial distances and locating the small habitat patches within the unsuitable matrix is not an easy task, and may require sophisticated movement and navigation abilities. In one extreme, the individuals would have a perfect knowledge of the landscape structure, and take the least cost path leading directly to a target habitat patch. In the other extreme, the individuals would have no information of the landscape structure, and no ability to navigate, moving entirely randomly. In Chapter II I have taken an intermediate approach, in which the individuals have no large-scale information about the landscape structure, but respond locally to changes in habitat quality. More precisely, I assumed that the individuals follow otherwise a random walk, but that they display edge-mediated behavior, biasing the movement towards the habitat patch when being at the boundary zone between the patch and the unsuitable matrix. This behavior leads to decreased emigration and increased emigration rates, allowing the individuals to spend most of their time in the small habitat fragments.

Random walk models and their diffusion approximations have long been used as models of animal and plant movement (Patlak 1953, Turchin 1998, Okubo and Levin 2002, Codling et al. 2008). In a simple uncorrelated random walk model the individual jumps a distance \( \Delta x \) during a time step of length \( T \), and the process is repeated until the individual dies, death taking place with rate \( \mu \geq 0 \). Assuming that the
distribution of displacements $\Delta x$ and durations $T$ are independent, and that they are exponentially tailed, the process can be approximated with a reaction-diffusion equation. The reaction-diffusion model describes the time evolution of the probability density $p$ of the animals location $(x, y)$ at time $t$ (Schmidt et al. 2007, Fedotov and Iomin 2008),

$$\frac{\partial p(x, y, t)}{\partial t} = D \sum_{i=1}^{d} \frac{\partial^2 p}{\partial x_i^2} - \mu p,$$  \hspace{1cm} (4-4)

where $d$ is the dimension of the space, and the diffusion coefficient equals $D = \frac{1}{2d} \langle \Delta x^2 \rangle$, where $\langle X \rangle$ denotes the expectation of the random variable $X$.

Dispersal of animals is often assumed to follow a correlated random walk, as at short time scales the individuals tend to continue moving in the same direction. Diffusion approximations for correlated random walks can be derived, but they are generally rather complicated (Othmer et al. 1988, Hillen and Othmer 2000). Consider a correlation random walk where an individual moves at const speed $v$ and changes its direction by a Poisson process with intensity $\lambda$. Then the reaction-diffusion equation of Eq. (4-4) follows (Porra et al. 1997), but now $D$ is replaced by the effective diffusion coefficient $D^* = \frac{1}{d \lambda \left(1 - \cos \gamma \right) + \mu}$. Here $\gamma$ is the turning angle, i.e. the direction of the present move relative to the direction of the previous move. The dependence of the effective diffusion coefficient $D^*$ on the death rate $\mu$ is generally weak as $\mu << \lambda$.

Real animals disperse in heterogeneous landscapes, where the movement rates differ among the habitat types, and the individuals often show different kinds of behaviors at boundaries between the habitat types (Ries and Debinski 2001, Haynes and Cronin 2006). Heterogeneous landscapes can be built into the reaction-diffusion equation Eq. (4-4) either by assuming a smoothly varying diffusion coefficient (Turchin 1998), or by assuming that the landscape consists of discrete habitat types (Ovaskainen and Cornell 2003, Ovaskainen 2004a). The behavior at habitat boundaries can be incorporated by appropriate matching conditions for the reaction-diffusion equation. Cantrell and Cosner (1999) and Fagan et al. (1999) used matching conditions stemming from skew Brownian motion, predicting that organisms accumulate at habitat boundaries. Assuming that the animal has a tendency to move towards the preferred habitat type, Ovaskainen and Cornell (2003) derived another type of a matching condition, characterized by a discontinuous probability density across the boundary.

The diffusion-reaction model of Ovaskainen and Cornell (2003) can be parameterized with movement data using a likelihood-based approach, and it has been shown to fit well to observed movements of the Glanville fritillary and other butterflies (Ovaskainen 2004b, Ovaskainen et al. 2008a, Ovaskainen et al. 2008b). While the model can be simulated under arbitrary landscape structures using numerical solvers such as the finite-element method (Ovaskainen 2008), assuming a simple enough landscape structure can allow also for mathematical developments. Ovaskainen and Cornell (2003) assumed a network of circular patches, and derived analytical formulae for movement probabilities between the patches at the limit where the distances between the patches are much larger than the sizes of the patches, i.e., assuming that the landscape is highly fragmented.

In Chapter II I extend the model of Ovaskainen and Cornell (2003), allowing for spatial variation in patch quality by including patch-specific diffusion coefficients, death rates and patch preferences. Assuming that the landscape is highly fragmented (patches are small compared to inter-patch distances), I derive mathematical approximations for parameters characterizing dispersal success, such as patch-to-patch movement probabilities and the times that the animal is expected to spend in the different habitat patches. The results of Chapter II make it
Summary

It is possible to simulate the dispersal phase in metapopulation models in a computationally efficient manner.

**Individual based simulations (III)**

Individual based simulations have been used in almost all subfields of ecology and evolutionary biology. The major advantage of the simulation approach is that it allows one to include much more detailed and realistic assumptions than more mathematical approaches such as differential- or difference-equation models. For example, one may assume that the individuals differ from each other in a way or another, making it possible to simulate also evolutionary processes. DeAngelis and Mooij (2005) reviewed research based on individual based modeling (~900 papers), and described five major types of heterogeneities that were accounted for by the individual-based approach: spatial, ontogenetic, phenotypic, cognitive, and genetic variability.

One of the major drawbacks with individual-based simulations is that there is no general framework for designing, testing and analyzing the models (but see the recent book Grimm and Railsback (2005) for some guidelines). A general strategy, called pattern-oriented modeling, has recently been advocated (Wiegand et al. 2003, Grimm et al. 2005). The key idea of pattern-oriented modeling is to use multiple patterns observed in real systems, ideally at different hierarchical levels, to guide the model design and thus to reduce uncertainties in model structure and model parameters.

In Chapter III I use pattern-oriented modeling to build an individual-based evolutionary sexual model for the Glanville fritillary butterfly metapopulation. The life cycle of the butterfly was split into the dispersal phase (by adult males and females) and local dynamics (larval survival). The population dynamics of the Glanville fritillary show a strong spatial correlation structure. To mimic such dynamics, I assumed that the survival rates of the larval group over winter are spatially correlated with a multivariate normal structure, the correlation matrix depending on the patch-to-patch distances. The dispersal phase of the butterfly is modelled based on the analytic results on the diffusion model derived in Chapter II. Based on empirical data on the effect of the gene *Phosphoglucose isomerase* (*Pgi*), I assumed that the dispersal rate in the landscape matrix is affected by a locus with two alleles (A and C). The key assumption here is that the individuals with the C allele are more mobile. Based on empirical results, I further assumed that the AC heterozygote females have higher fecundity than other types of individuals, and that the CC homozygotes have a very low survival probability before they hatch as larvae. I parameterized the model by applying the pattern-oriented approach to the extensive data on the Glanville fritillary, and used the parameterized model to examine the interplay between ecological and genetic processes.

5 Main results and discussion

**Factors affecting the death rate in the Glanville fritillary butterfly (I)**

The Glanville fritillary metapopulation in the Åland Islands persists in a balance between local extinctions and recolonizations of empty habitat patches. Thus the butterflies are distributed into a set of local populations, some of which have persisted for a long time and some of which are newly established. As new and old populations are likely to have dissimilar selection pressures (as will be examined in detail in Chapter III), it is interesting to study if and how individuals originating from these types of populations differ from each other. Chapter I addresses this question, based on common garden experiment that was conducted on a population cage. In the cage individually marked butterflies were followed over their lifetime. The butterflies belonged to four groups, classified by their population type (new vs. old) and sex (female vs. male).

I modelled the death rates and capture probabilities of the individually marked butterflies using a hierarchical Bayesian model. The high search effort (in space and time) in the cage allowed for a detailed investigation on the factors affecting the probability of death. The results showed that the probability of death was much lower during the night than during the day time. In all four groups, the death rate increased with the age of the butterfly. For females originating from old populations the increase with age was much less pronounced than for the butterflies in the three other groups.
The results of Chapter I and observations that were done during the cage experiment suggest that most of the deaths were likely caused by senescence, rather than by predators, which were almost missing from the cage. Deaths due to senescence might be expected to have occurred during the day time, as during the night the butterflies remained inactive and thus had a very low rate of metabolism. It is possible that the higher death rate of old females from new than old populations represents a physiological trade-off, as the females from new populations have a superior metabolic performance (Chapter III, Hanski et al. 2004, Haag et al. 2005).

Dispersal in highly fragmented landscapes (II)

In Chapter II I derive a number of analytical results characterizing dispersal success in a network of habitat patches. These include the probabilities by which the individuals move between a given pair of patches, and the times that the individuals are expected to spend in the habitat patches and in the matrix. The results of Chapter II extend earlier results of Ovaskainen and Cornell (2003) by assuming that the habitat patches may differ in their quality, and by considering a number of additional variables characterizing the dispersal process. The results show that dispersal success, characterized e.g. by patch-to-patch movement probabilities or expected movement times, is not only affected by patch sizes and inter-patch distances, but by the entire configuration of the patch network.

The results of Chapter II can be used to construct a sub-model for the dispersal phase in metapopulation dynamics (as will be done in Chapter III). Compared with a dispersal kernel based approach, such as the virtual migration model (Hanski et al. 2000, Heino and Hanski 2001), an advantage of the movement model developed in Chapter II is that it has a mechanistic basis. Compared with most individual-based simulations with detailed movement rules (Tischendorf 1997, Vuilleumier and Metzger 2006), another advantage of the model of Chapter II is that it can be rigorously connected to data (Ovaskainen et al. 2008b). With the help of the analytical formulae derived in Chapter II, the dispersal model can be simulated in large patch networks with high computational efficiency.

Measures of dispersal successes derived in Chapter II can be viewed as measures of functional connectivity. Unlike the widely used dispersal kernel based measures of patch connectivity (Moilanen and Hanski 2002), the measures of functional connectivity that I derived account for edge effects (Haynes and Cronin 2003, 2006), for spatial variation in patch quality (Bowne et al. 2006), for the effects of patch configuration, and for dissimilar behaviours in the matrix and in the habitat patches (Ricketts 2001, Verbeelen et al. 2003). Belisle (2005) suggested to measure functional connectivity by travel cost. In the context of the model of Chapter III, travel cost relates both to between-patch movement probabilities (the probability of death during dispersal) and to the time spent in the matrix (e.g. the cost of lost reproduction opportunities).

Evolution of dispersal in the Glanville fritillary metapopulation (III)

The links between ecology and evolution provide a challenging field of research in evolutionary ecology (Van Tienderen 2000, Stockwell et al. 2003, Kokko and Lopez-Sepulcre 2007, Lion and van Baalen 2008). The feedback between demographic dynamics and evolution arises because demographic dynamics impose selection pressures upon individual traits involved in reproduction and survival, and the evolution of these traits in turn influences demographic dynamics (Fig. 1; Kokko and Lopez-Sepulcre 2007). Unlike many life-history traits such as fecundity or mortality, movement behaviour affects the fitness of an individual in a complex and often non-monotonic manner. Previous observational studies on the Glanville fritillary metapopulation in the Åland Islands (Hanski and Saccheri 2006, Saccheri and Hanski 2006) suggest that in this system there is a strong coupling between ecological dynamics and the allelic variation of a particular gene affecting the mobility of individuals. The purpose of Chapter III was to construct an evolutionary, individual-based sexual model of the dispersal for the Glanville fritillary butterfly, and to provide detailed insights on the coupling between demographic and evolutionary dynamics.
I used the pattern-oriented approach to fit the model to the observed demographic dynamics of the Glanville fritillary and the observed neutral genetic dynamics. The model was then successfully tested with two independent empirical data sets (Haag et al. 2005, Hanski and Saccheri 2006) on spatial variation in \( Pgi \) allele frequency. First, at level of local populations, the frequency of the \( C \) allele (associated with highly mobile individuals) was predicted and observed to increase with decreasing connectivity in newly-established populations, and to decrease with decreasing connectivity in old populations. This result can be explained by the selection pressures that individuals experience in different kinds of environments. Empty isolated patches are likely to be colonized with especially mobile heterozygote \( AC \) females. The offspring of these females will include both sedentary and mobile types, out of which the latter will soon emigrate out leading to a low dispersal rate in isolated old populations. Second, at level of sub-network, the frequency of \( C \) was predicted and observed to increase with decreasing metapopulation size. This result follows as small metapopulations occupying marginal patch networks show an especially high turnover rate, leading to more opportunities for colonization of empty patches and thus favouring the more mobile individuals.

I examined the causality in the coupling between demographic and evolutionary dynamics by simulation experiments, where either the genetics (the frequency of \( C \)) or the demography (the number of larval groups) was perturbed, and the induced change in demographic and evolutionary dynamics was measured. These experiments suggested that under a range of network structures, the allele frequencies in the \( Pgi \) gene have evolved to values that are close to the optimal in terms of demographic performance, as perturbing the \( Pgi \) allele frequency either upwards or downwards from the evolved value led to reduced growth rate. In highly fragmented landscapes, elevated population turnover rate selects for the \( C \) allele, especially when the metapopulation is at low density and there are opportunities for colonization. In less fragmented landscapes, the more dispersive genotype is selected for when the metapopulation reaches a high density. Thus variation in demographic dynamics creates variation in selection gradients that helps maintain \( Pgi \) polymorphism at the regional level.

Hierarchical metapopulation dynamics of two species of aphid (IV)

In Chapter IV I modelled the multi-scale dynamics of two aphid species (MF and MT, see Study systems and empirical data) by analyzing within-year data collected in two study areas. In addition to the population dynamics of the individual species, I also examined the effect of interspecific interactions.

The state-space model developed in Chapter IV is an extension of previous SPOMs (O’Hara et al. 2002, Ter Braak and Etienne 2003, Etienne et al. 2004), but the assumption of independent local extinctions is relaxed with the help of four hierarchical spatial scales (ramet, plant, group, and island). I assumed that population dynamics can be affected by a number of underlying processes, some of which are operating only very locally and some at a larger scale. These underlying processes cannot be observed directly, as the observed dynamics (e.g. colonizations and extinctions) results from their joint action. For example, an observed extinction at a given spatial scale may result from an extinction agent (e.g. predation or adverse environmental conditions) operating specifically at that spatial scale, or from an extinction agent operating at a larger scale, or from several smaller-scale extinction agents operating simultaneously. An observed colonization can be by a winged adult, by an unwinged adult, or by the offspring of a primary colonizer. The aim of Chapter IV was to use Bayesian inference to decompose the observed extinctions and colonizations into their causal components.

As expected, both colonization and extinction rates decreased with increasing spatial scales, though the decrease was less pronounced for colonization rates of winged aphids. Winged aphids were superior to non-winged aphids not only in large-distance colonization, but also they colonized at a higher per-capita rate ramets within in the same plant, and plants within the same group. The latter result is not directly supported by earlier studies (e.g. Dixon 1998, Irwin et al. 2007), which suggest that winged aphids tend to fly for a prolonged period before they settle to a new host plant. However it is still debated about how far aphids have to go before they are willing to accept a new host plant (Lodoxale et al. 1993).

The species MT had very high turnover rate at the ramet-level, whereas the species MF showed more stable dynamics. This result is in line with the
earlier observation that the species MF is obligatorily ant-tended, because of which it has a lower mortality rate and is less likely to move to another feeding location (Flatt and Weisser 2000). There was a strong indication of the two species competing with each other, leading to decreased colonization and increased extinction rates, the effect of MF to MT being stronger than vice versa. Also this result is consistent with behavioural experiments, which point to an ant-mediated asymmetry in the competition between the two species (Fischer and Völkl, unpubl.). When a MF individual colonizes a ramet occupied by MT, tending ants attack MT individuals and remove them from the plant.

The estimated model parameters were mostly consistent between the two study areas (Tvärminne and Basel, see Study systems and empirical data). As an exception, in both species the offspring of the primary colonizer were more active in generating secondary colonizations in the Basel study area than in the Tvärminne study area. A simple explanation for this result is the higher average number of ramets per plant in Basel than in Tvärminne, leading to more opportunities for local spread in Basel.

The parameterized model allows one to simulate aphid dynamics under different kinds of landscape structures, with the caveat that the model describes only within-season dynamics, there being little empirical information about winter mortality. I assumed that winter mortality is uniform across the study area, and conducted simulation experiments to study how the spatial configuration of the host plants and interspecific interactions affect metapopulation persistence. The results of these simulations showed that the species MF benefits from an aggregated distribution of the host plants, whereas for the persistence of the species MT is facilitated by a segregated distribution of the hosts. These results can be explained by the higher turnover rate of species MT, which increases the rate of local extinction and calls for a higher number of semi-independent units to ensure persistence. The simulation results also showed that the presence of species MT has almost no effect on the persistence of the species MF, and that even the dominant species MF has only a very mild effect on the species MT. These results can be explained by the low prevalence of both species, leading the competitive interactions being realized only rarely.

6 Synthesis and perspectives

In this thesis I have developed and applied various methodologies to analyze ecological and evolutionary processes in spatially structured populations. I have used analytical approaches, hierarchical Bayesian methods, and individual-based simulations. Each of these approaches has its own tradition and purpose, but they can also be used in combination to tackle complex problems from many different viewpoints.

Analytical approaches, such as systems of differential equations, are attractive for mathematicians, because they often allow for complete and mathematically beautiful solutions. However, these approaches can often be overly simplistic to describe complex biological problems, or to fit to data acquired from the field. But even if a complex problem cannot be fully treated analytically, its analysis can be greatly facilitated if analytical results are available for part of the problem (see Chapters II and III).

As illustrated in Chapters I and IV, hierarchical Bayesian methods provide a useful tool for decomposing ecological complexity into smaller components, and for generating predictions based on the estimated model parameters. Inferences and predictions from data can alternatively be done by combining individual based simulations with pattern oriented modeling (Chapter III), e.g. for problems that are so complex that it is difficult to apply the Bayesian framework. Methodology in Bayesian inference has however been steadily advancing. The method of approximated Bayesian computations (ABC) has been recently developed and applied in biological sciences (Beaumont et al. 2002, Marjoram et al. 2003, Bortot et al. 2007, Ratmann et al. 2007, Sisson et al. 2007, Joyce and Marjoram 2008), for complex probability models where the likelihood is impossible or computationally intractable. Instead of weighting a candidate parameter value by likelihood, the ABC approach assigns weights by the distance between simulated data and empirical data. Thus, the ABC approach provides a bridge between Bayesian inference and pattern oriented modeling. In the past ten years there has been a great increase in the usage of Bayesian methods in population biology and evolutionary biology, and I expect that this trend will continue also in the future.
Summary

My sincere gratitude goes to the technical support staff of MRG. I am especially grateful to Evgenily Meyke for database and programming related supports; Camilla Ekblad and Sami Ojanen for computer maintenance; Tuuli Ojala, Marika Lilja, Nina Lehtosalo and Elina Rossi for their help with administration and for translating various texts written in Finnish for me.

I would like to thank all the MRG members, who have created a nice and friendly environment for scientific and non-scientific interaction. I would like to thank the Department and LUOVA graduate school, under which this thesis was carried out. I would also like to thank all the other people who have provided direct or indirect help on this work.

I would like to express my warmest thanks to my wife Jie Zheng for her continuous support and encouragement, and to my daughter Xijin Zheng for making life worthwhile. To them this thesis is dedicated.

This work was financially supported by Academy of Finland (Grants no. 211173 and 124242 to Otso Ovaskainen).

7 Acknowledgements

First and foremost, I wish to express my deepest gratitude to Otso Ovaskainen, my supervisor, for introducing me to the field of Bayesian statistics and mathematical ecology, for sharing with me his expertise and research insights, and for valuable advises and great patience throughout this work. The writing of this thesis has been greatly improved by his help. All that I have learned from him, especially the way of scientific thinking, will definitely benefit me in my future research career.

I wish to express my respectful gratitude to Ilkka Hanski, a co-author in three out of four chapters in this thesis, for his valuable comments and guidance. I learned a lot about the biology of natural populations from discussing with him. I would like to express my sincere thanks to my co-author Wolfgang W. Weisser for providing useful discussion and for hosting me in Jena, Germany for a couple of weeks in 2007. Special thanks go to Franklin Nyabuga, a PhD student of Wolfgang W. Weisser, whose help made my staying in Jena easy and enjoyable. I would like to acknowledge Marjo Saastamoinen, Juho Pennanen, and Simone A. Härri for their contributions as co-authors. Many thanks to Marjo Saastamoinen and the cage crew in 2005, with whom I had a good time during my visit to the field cage experiment in the Åland Islands.

I would like to express my sincere thanks to Bob O’Hara and Juha Alho for reviewing and commenting this thesis.

My sincere gratitude goes to the technical support staff of MRG. I am especially grateful to Evgenily Meyke for database and programming related supports; Camilla Ekblad and Sami Ojanen for computer maintenance; Tuuli Ojala, Marika Lilja, Nina Lehtosalo and Elina Rossi for their help with administration and for translating various texts written in Finnish for me.

I would like to thank all the MRG members, who have created a nice and friendly environment for scientific and non-scientific interaction. I would like to thank the Department and LUOVA graduate school, under which this thesis was carried out. I would also like to thank all the other people who have provided direct or indirect help on this work.

I would like to express my warmest thanks to my wife Jie Zheng for her continuous support and encouragement, and to my daughter Xijin Zheng for making life worthwhile. To them this thesis is dedicated.

This work was financially supported by Academy of Finland (Grants no. 211173 and 124242 to Otso Ovaskainen).


