

PhD thesis

Heterogeneity in hidden developmental processes: inference and analysis for stage-structured populations in fluctuating environments

María Soledad CASTAÑO

A thesis presented to the University of Antilles
in fulfillment of the requirements for the degree of Doctor of Philosophy in
Physiology and Biology of Organisms, Populations and Interactions

Co-PhD supervisors:

- David PLEYDELL, French National Institute for Agricultural Research (INRA), UMR117 ASTRE, Duclos Campus, Guadeloupe, France.
- Hélène GUIIS, Center of International Research for Agricultural and Development (CIRAD), UMR117 ASTRE, Antananarivo, Madagascar.

Director:

- Jean VAILLANT, Mathematics and Computation Department, University of Antilles, Pointe-à-Pitre, Guadeloupe.

Members of the PhD jury:

- Dylan Childs, Department of Animal and Plant Sciences, University of Sheffield.
- Roberto Salguero-Gomez, Department of Animal and Plant Sciences, University of Sheffield.
- Bethan Purse, Centre for Ecology & Hydrology's Disease Ecology Group, Wallingford.
- Eva De Clercq, Catholic University of Leuven, Belgium.
- Harry Archimède, URZ, INRA, Guadeloupe.

Abstract

In this work, I present a new matrix-based model for biological stage-structured populations (SSPs) that greatly improves the characterisation of variation in development times by tracking individual histories within each stage. Neglecting such heterogeneity has historically limited the realism and predictive performance of most SSP modelling approaches. The key idea of the new model is to augment a classic Lefkovitch matrix with stage-specific integral projection models (IPMs) that track within-stage dynamics.

This new “integral projection Lefkovitch matrix” (IPLM) model drastically reduces stage-duration errors; is robust to stage distribution instabilities arising from perturbations; permits parsimonious parameterisation with random variables or time-varying covariates; and can be fitted, even when within-stage development is unmeasurable, using developmental cohort data. By using maturation-time (and not size) data, our methods greatly improve the precision of stage-structured IPMs whenever size is a poor, or unavailable, predictor of stage duration. This scenario is ubiquitous in ecology: egg (e.g. fish, bird, insects) and exoskeleton (e.g. Ecdysozoa) dimensions often remain relatively constant, and more appropriate developmental metrics can be too expensive or difficult to collect routinely. Furthermore, by incorporating a combination of laboratory and field data, Bayesian methods permit the estimation of cryptic parameters *in natura*, such as the strength of regulatory density-dependent mechanisms or environmental stochasticity in vital rates. Thus, by assimilating time series data – even of incomplete life-cycles – IPLMs permit upscaling from the laboratory to the field.

Initially, the identifiability of IPLM parameters is studied with simulated data from marked cohort studies where individual qualities correlate maturation-times. Results demonstrate that accurate sojourn-time distributions are reproduced even from small samples. Next, a temperature-dependent model is fitted to *Culicoides* (biting midge) unmarked laboratory cohort data to assess the relative role of transient and asymptotic dynamics in constant and seasonal climates. Results demonstrate that the traditional negligence of individual developmental heterogeneity affects asymptotic dynamic metrics in various ways and greatly underestimates the importance (both amplitude and duration) of transient dynamics.

Three applications/extensions of the *Culicoides* IPLM are studied. First, the fitted model is used to assess the validity and robustness of linearity assumptions of classic degree-day insect development models. Results show

that linearity only provides a robust developmental model over extremely narrow temperature intervals. Secondly, projections of adult densities are used to assess transient and asymptotic dynamics in the basic reproduction ratio (R_0) of bluetongue following the initialisation of a hypothetical adulticide-based vector control program. Results show that R_0 drops suddenly following a reduction in adult survival. But this is only a transitory effect when the vector population growth rate is not brought below one. Whether or not, and for how long, a given adulticide can maintain $R_0 < 1$ is temperature dependent, a result that has implications for integrated vector management.

Finally, the *Culicoides* IPLM is used to construct a state-space model (SSM) for analysing typical multi-annual time series data from vector abundance studies. With simulated time-series of weekly adult flight-trap data, the SSM is used to explore the identifiability of key cryptic parameters *in natura*, including the level of environmental stochasticity in mortality; the strength of density dependent mortality among larvae; the initial population density; and the expected efficiency of flight-traps. Results show that when flight-trap efficiency is known, the parameters are identifiable to a high level of precision using simulated weekly trap counts over three years. However, when trying to estimate flight-trap efficiency, a very strong correlation with the density-dependence parameter is detected, suggesting that additional data sources are required to calibrate the model for epidemiological purposes.

Applications for many state-structured populations - particularly those where cryptic developmental status has to date prevented study with IPMs - are foreseen in fields including ecological forecasting, mechanistic niche modelling, demographic compensation studies or eco-evolutionary analysis. Diverse applications are expected for conservation, agricultural, epidemiological or theoretical purposes.

Contents

Abstract	1
1 Introduction	4
1.1 Framework and motivations	4
1.2 Within-stage developmental variation in single-species models	8
1.2.1 Differential equations	9
1.2.2 Stage-duration distribution models	10
1.2.3 Matrix population models	11
1.2.4 Integral projection models	14
1.3 Parameter estimation in population models	16
1.3.1 Overview	16
1.3.2 Bayesian inference: a brief description	17
1.3.3 State-space models	19
1.4 Conclusion	23
1.4.1 Motivating example	23
1.4.2 Aims and objectives	23
1.4.3 Thesis outline	24
2 A model for individual heterogeneity in hidden developmental processes	26
2.1 Introduction	26
2.2 Integral projection Lefkovitch matrix models	28
2.2.1 The classic Lefkovitch matrix (CLM) model	28

2.2.2	Incorporating within-stage development dynamics	29
2.2.3	The general integral Lefkovitch matrix (IPLM) model	31
2.2.4	Fitting with maturation time data	32
2.3	A simulation study with correlated stage-durations	33
2.3.1	Correlated stage-durations and Gaussian copulas	34
2.3.2	Generating <i>in silico</i> survival data from the quality-dependent IPLM model	34
2.3.3	Assessing estimation performance	36
2.3.4	Results	37
2.4	Discussion	40
2.5	Conclusion	41
3	An IPLM-based study of seasonal and transient dynamics of <i>Culicoides</i> biting midges	43
3.1	Introduction	43
3.2	<i>Culicoides</i> IPLM model	45
3.3	Results	47
3.4	Discussion	57
3.5	Conclusion	59
3.6	Annexe. Analysing linear assumption of degree-day approaches.	61
4	Extending the utility of the IPLM framework	67
4.1	Estimating the effects of vector control on the R_0 of bluetongue	68
4.1.1	Introduction	68
4.1.2	Methods	69
4.1.3	Results	71
4.1.4	Discussion	73
4.1.5	Conclusion and perspectives.	75
4.2	Towards a state-space IPLM framework	77
4.2.1	Extending the IPLM model	77
4.2.2	Parameter estimation with IPLM state-space models	85

	3
4.2.3 Conclusion and perspectives	92
5 General conclusion	93
Appendices	98
A Appendix to Chapter 2	98
A.1 Discretising within-stage IPLM	98
A.2 Quality-dependent development with Gaussian copulas	99
A.3 Generating sojourn-mortality probabilities with IPLMs	100
A.4 Markov chain Monte Carlo strategy	101
B Appendix to Chapter 3	103
B.1 <i>Culicoides</i> biting midges	103
B.1.1 <i>Culicoides</i> as disease vectors	103
B.1.2 Life-cycle	105
B.2 <i>Culicoides</i> life cycle data	106
B.3 Likelihood functions	107
B.3.1 Bayesian model for expected fecundity	109
B.4 Imputation of missing data	110
B.5 Posterior distributions of <i>Culicoides</i> IPLM model parameters	111
B.6 MCMC strategy	112
B.7 Unimodal cubic Hermite spline interpolation	113
C Appendix to Chapter 4	116
C.1 A brief description of the synthetic likelihood method	116
C.1.1 Choice of summary statistics	118
D Article: <i>Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes</i>	119
Bibliography	149

Chapter 1

Introduction

1.1 Framework and motivations

The transition between the second and third millenium has been characterised by the intensification of international trade and travel, and strong interrelations between economies worldwide. This interconnected world has brought new threats for humanity, among which the risk of emerging or resurgent vector-borne (VB) pathogens has become a global issue (Gubler, 2002b; Daszak *et al.*, 2000; Anderson *et al.*, 2004; Haines *et al.*, 2006; Jones *et al.*, 2008; Murray and Daszak, 2013; Watts *et al.*, 2015; Young *et al.*, 2016).

Vector-borne pathogens represent a public health issue. Most emergent VB human diseases are zoonoses, i.e. pathogens that can be transmitted between vertebrate animals and humans. Examples of human, zoonotic and animal diseases of major concern in France are shown in tables 1.1 and 1.2. In agriculture, VB diseases affecting both animals and plants are not only responsible for huge economic losses worldwide, they also exert considerable societal pressure by disrupting food production systems at local, regional and international levels, and negatively affect the livelihoods of vulnerable farmers (Battisti and Naylor, 2009; Godfray *et al.*, 2010; Keesing *et al.*, 2010; Chakraborty and Newton, 2011; Wheeler and Von Braun, 2013; Lipper *et al.*, 2014).

The dramatic resurgence and spread of some well known VB diseases which had historically been stably bounded to specific regions (e.g. dengue, chikungunya, zika, yellow fever or Lyme disease for humans; bluetongue virus and heartwater for animals, and blackheart or sharka disease for plants), has been associated to changes in a variety of factors affecting the epidemiology of VB diseases. Which, to what extent, and how, every one of these factors shape the transmission of VB pathogens, is a matter of constant research and debate, although there

is a general agreement that most of these modifications are anthropogenically induced, such as changes in host distributions or movements, socio-economy, land use, animal health systems, or climate (Taylor *et al.*, 2001; Gubler, 2002a; Purse *et al.*, 2005; Chevalier *et al.*, 2010; Mills *et al.*, 2010).

Many pathogens causing VB diseases (e.g. viruses, bacteria, nematodes or other parasites,) are transmitted among humans, animals and plants by the bites of infected arthropods such as mosquitoes, ticks, aphids, mites, triatomine bugs or leafhoppers, among others (Beaty *et al.*, 1996; Cardinale *et al.*, 2003; Alekseev, 2004; Harris and Maramorosch, 2014). Empirical evidence of climate change affecting the distributions of many species, including arthropod vectors and the pathogens they vector, has been published profusely in recent years. For example, global warming has been related to shifts in the distribution of insects and other wildlife across latitudinal and elevation gradients (Parmesan and Yohe, 2003; Franco *et al.*, 2006; Hickling *et al.*, 2006; Hill *et al.*, 2011).

Arthropods, like many other organisms, are poikilothermic ectotherms (i.e. cold-blooded, therefore sensitive to temperature changes) and progress through a series of discrete life stages, each of which is associated with a distribution of maturation times that can vary as a function of environmental conditions. In general, for these populations, describing how each stage's vital rates vary with genetic and environmental factors provides a basis for studying their dynamics (Manly, 1990). With increasing confirmation of anthropogenically induced change in climatic patterns, it has become urgent to understand how this can affect both ecological (e.g. phenology, abundance) and evolutionary (e.g. fitness) aspects of these populations' dynamics.

Development through a life stage takes time, a time that *always* varies between individuals. Such variation on individual performance is an intrinsic aspect of populations, and can arise from several biotic (e.g. genetic or density-dependence) and abiotic (e.g. environmental) mechanisms (Randolph, 1997). However, perhaps for practical or cultural reasons, most predictive models for stage-structured populations have typically either disregarded individual heterogeneity and the variance in development time, or the way it has been incorporated has lacked realism regarding the relationships between biotic/abiotic factors and developmental time distributions (De Valpine *et al.*, 2014; Vindenes and Langanen, 2015).

Despite a growing literature on the ecological consequences of variation in individual-level traits, we still lack a general framework with computational methods for the estimation of stage duration distributions and stage-specific mortalities in fluctuating environments (Hoeting *et al.*, 2003; Murtaugh *et al.*, 2012; De Valpine *et al.*, 2014). The improved realism of such a framework would facilitate studying the mechanisms by which various factors influence

ecological dynamics, i.e. it would help determine when, and to what extent, individual heterogeneity will affect both the long-term and transient dynamics of populations.

This thesis was motivated by initiatives to further develop the current predictive framework for biological stage-structured populations. Specifically, our aim has been to develop a predictive framework that incorporates within-stage developmental variation; is flexible enough to include complexities encountered by natural populations such as time-varying covariates, density-dependence or environmental stochasticity; and enables the use of different types of data for estimation, inference and simulation.

For this, we have developed a matrix-based model that, unlike other matrix-models, incorporates within-stage dynamics. This is done using integral projection models (IPMs, Easterling *et al.* (2000b)). The novelty in this approach is that (i) a hidden-state enables the tracking of within-stage developmental variation via a parsimonious stage-specific parameterisation; (ii) the projection of within-stage dynamics permits likelihood-based model fitting with developmental data at either/both the individual or cohort level; and (iii) within-stage IPMs greatly facilitate realistic incorporation of various complexities such as time-varying covariates, density-dependence, and both endogenous and exogenous stochasticity.

In this thesis, I present, for the first time, this new class of augmented matrix model (chapter 2). I test parameter identifiability with simulated data (chapter 2) and parameterise a temperature-dependent IPLM model for *Culicoides* biting midges using laboratory data (chapter 3). The fitted model is then used to (i) analyse the robustness of the linearity assumptions of classic degree-day development models (section 3.6); (ii) make inferences regarding the evolution of bluetongue's basic reproduction ratio, R_0 , over the course of a hypothetical adulticide-based control program (chapter 4, section 4.1). Finally, I investigate the use of a modern method – synthetic likelihood – for making inference from typical time series data from vector density studies (chapter 4, section 4.2).

In the remainder of this chapter, several topics are visited in order to give a general overview of the most relevant methodological aspects that sustained this thesis. More precisely, a brief state of the art of biological stage-structured models and estimation methods is presented. We conclude the chapter by listing the aims of this work.

Table 1.1. Main vector-borne human diseases and zoonoses in France and French overseas territories.

Human Diseases and Zoonoses								
Disease	Vector	Reservoir	Distribution	Status	Morbidity	Incidence	Fatality	Trends
Dengue	<i>Ae. aegypti</i> <i>Ae. albopictus</i>	Human, vectors	French Caribbean & American Departments, Reunion, Mayotte, Pacific	Endemo-epidemic	High	High	Yes, if different serotypes are in circulation	Expansion
Chikungunya	<i>Ae. aegypti</i> <i>Ae. albopictus</i>	Human, monkey, vectors	Reunion, Mayotte, French Caribbean & American Departments, Pacific, metropolitan France,	Epidemic	High	High	Low	Recurrent epidemics each 10-20 years
Toscana virus infection	Phlebotominae (sandfly)	Human, vectors	Mediterranean rim	Endemic	Moderate	Low	Zero	Stable but better recognized
Malaria	<i>Anopheles</i>	Human	French Guiana, Mayotte, Potential: French Caribbean & American Departments, Reunion, Corsica	Endemo-epidemic	High	High	High for <i>P. falciparum</i> 5-10 deaths/year	Tending to decrease in French Guiana & Mayotte
<i>Bartonella quintana</i> infection	Body lice	Human	Cosmopolitan	Endemic	High	High	Yes	Expansion among homeless and deprived people
Lymphatic filariasis	<i>Aedes</i> , <i>Anopheles</i> , <i>Culex</i>	Human	Mayotte, French Polynesia, Wallis-and-Futuna	Endemic	Potentially high	Low	No	In regression
West Nile virus infection	<i>Culex</i>	Birds	All continents including Europe, Mediterranean rim Guadeloupe.	Endemo-epidemic	Potentially high	Low	High encephalitis if develops	Expansion in North America
Lyme borreliosis	Tick: <i>Ixodes ricinus</i>	Rodents, red deer, roe deer, vector	Metropolitan France (not South-East France)	Endemic	High	High	Very low	Expansion
Tick borne rickettsiosis	Tick	Vectors	Varies with type of rickettsiosis, (mainly South-East France)	Endemic	High	Moderate	Limited	Possible expansion
Rift Valley fever	<i>Culex</i> , <i>Aedes</i>	Ruminant, vectors	Indian Ocean, Mayotte	Endemic-epidemic	High	Documentation under way in Mayotte	Low	Expansion
Leishmaniasis	Phlebotominae (sandfly)	Dogs, sylvatic reservoirs	Metropolitan France, French Guiana, Martinique	Endemic	High	Low	Potential in visceral form	Expansion in French Guiana. Climate change influence in metropolitan France?
Chagas disease	Reduviid	Wild mammals	French Guiana	Endemic	High	Uncertain	High	Expansion
Yellow fever	<i>Ae. aegypti</i>	Monkeys, vector	French Guiana	Isolated cases	High	Low	High	Disappearing with vaccination

Adapted from Fontenille *et al.* (2013).

Table 1.2. Main vector-borne animal diseases in France and French overseas territories.

Animal Diseases								
Disease	Vector	Reservoir	Distribution	Status	Morbidity	Incidence	Fatality	Trends
Bluetongue	Culicoides midges	Bovines, ovines	Metropolitan France and overseas territories	Endemic (overseas territories), Epidemic (Metropolitan France)	Limited in overseas territories, high in metropolitan France	High	Moderate (ovines) to low (bovines)	Emerging in metropolitan France, serious economic effects
Piraplasmosis, anaplasmosis	Ticks	Vectors, bovines, equines, dogs	Metropolitan France, territories	Endemic	Moderate	High	Moderate	Stable
Heartwater	Tick: <i>Amblyomma variegatum</i>	Ruminants	Guadeloupe	Endemic	High	High	Low to High	Risk of spread in Caribbean
Trypanosomiasis	Tabanids, stomoxes	Ruminants, equines	French Guiana	Endemic	Reduced economic value	High	Low to high	Stable (French Guiana), foci in metropolitan France
Besnoitiosis	Mechanical transmission (Tabanids, stomoxes, Hyppoboscidae)	Felids, bovines	Outbreak in metropolitan France	Endemic	Loss of economic value	Apparently pseudo-contagious	Death or loss of economic value	Renewed upsurge in metropolitan France
Equine infectious anaemia	Mechanical transmission (Tabanids, stomoxes)	Equids	Outbreaks in metropolitan France, French Guiana	Endemic	Moderate	Low	Low	Regression

Adapted from Fontenille *et al.* (2013).

1.2 Within-stage developmental variation in single-species models

Many classic models of population dynamics typically neglect individual level characteristics and simply attempt to describe how a scalar density, N , evolves in time (Verhulst, 1838; Malthus, 1852; McKendrick and Pai, 1912; Pearl and Reed, 1920; Lotka, 1925). However, vital rates (growth, mortality, fecundity) are rarely constant throughout a population, and structured population models attempt to stratify, or “structure”, a population into sub-populations who’s vital rates show greater homogeneity. A classic example are the age-structured models which date to Leslie (1945). A similar class of models are the stage-structured models (SSMs) (Lefkovitch, 1965) that describe populations that develop via a discrete set of stages. For example, SSMs have been used in plants (Crone *et al.*, 2013) and many taxa such as terrestrial and marine mammals (Ozgul *et al.*, 2009; Fujiwara and Caswell, 2001); fish (Pertierra *et al.*, 1997); birds (Blackwell *et al.* 2007); amphibians (Biek *et al.*, 2002) and arthropods (Coll *et al.*, 2012). The same conceptual framework is also used in other systems where “stage” can describe diverse features including habitat types (Horvitz and Schemske, 1995); animal location (Hunter and Caswell, 2005); health status (Shulgin *et al.*, 1998); patch occupancy (Hanski, 1994) and metapopulation or metacommunity status (Johnson, 2000).

Our focus here is on stage-structured models, and more particularly in (i) the implications of assumptions

regarding variance in stage durations and the processes that generate that variation; and (ii) how mechanistic models which generate this variance more realistically can be parameterised using available (or easily obtainable) data sets.

There has been a fair amount of work on modeling single-species populations with multiple stages, including discrete, continuous and stochastic models (Cushing *et al.*, 2002; Caswell, 2006; Yamanaka *et al.*, 2012; De Valpine *et al.*, 2014). At least four general classes of SSMs can be distinguished that make different assumptions regarding within-stage variation, namely differential equations, matrix population models, stage-duration distribution models, and integral projection models. Below I describe each of these approaches with particular attention paid to their limitations regarding characterising the variance of stage durations.

1.2.1 Differential equations

Systems of differential equations group individuals in stages characterised by a common set of vital rates (e.g. development, survival and reproduction). Although facilitating the treatment of non-linear dynamics (in general, non-linear feedbacks are incorporated via vital rates such as birth or mortality rates), standard differential equations do not handle the delays imposed by stage-specific maturation times or any associated sources of variance. Both maturation times and variance are well known to affect population dynamics and long-term trajectories (Blythe *et al.*, 1984; Wearing *et al.*, 2004; Clutton-Brock and Sheldon, 2010). Yet, systems of ordinary differential equations (ODEs) typically generate exponential sojourn-time (i.e. development time) distributions which can provide a poor level of realism. In systems of ODEs this can be somewhat addressed by adding sub-compartments to obtain Erlang distributed sojourn-times (Keeling and Rohani, 2008; King *et al.*, 2008). However, the approach lacks of the flexibility and generality to provide a realistic model when vital rates (and stage-duration distributions) are sensitive to environmental fluctuations. Attempts to overcome such limitations have been addressed by models including delays and stochasticity.

Delays can be incorporated into systems of differential equations to characterise stage-duration arising from maturation in processes. For an ordinary delay differential equation (DDE), for example, the general expression is $\frac{dN(t)}{dt} = f(N(t), N(t - T))$, where $N(t)$ represents the population (or sub-population of a given stage) at time t , $f(N(t), N(t - T))$ is a function describing growth, recruitment and mortality, and $T > 0$ is the delay parameter (Manetsch, 1976; Blythe *et al.*, 1984; Nisbet *et al.*, 1985; MacDonald, 1986; Aiello and Freedman,

1990; Kuang, 1993; Nisbet, 1997; Eurich *et al.*, 2005). In the simplest examples, individuals are assumed born at the same (initial) time and delay is fixed, thus mean stage-duration is well represented but variance is neglected. Alternatively, delays can be distributed according to some probability distribution that can be parameterised with covariates, but this approach neglects the processes that generate delays, and specifies that delays are pre-determined at a set point in time and cannot vary thereafter as a function of covariates. For insects in fluctuating environments, this is a large simplification that lacks realism and generality.

Delays have been included as dynamic variables of either food availability or temperature. In some of these cases, unrealistic assumptions (such as fixed duration for egg or adult state) have been made to simplify model analysis, hampering practical applications (Nisbet and Gurney, 1983; Nisbet, 1997). In some cases, temporal fluctuations (i.e. environmental or demographic stochasticity) are included (Frank and Beek, 2001; Mao *et al.*, 2005; Chen *et al.*, 2005; Cavalerie *et al.*, 2015; Wood, 2010). A small number of studies have incorporated stage-structured time series data for inference. In such cases, independently estimated fixed parameters are generally used to maintain parsimony (Severini *et al.*, 2003; Wood, 2010; Yamanaka *et al.*, 2012).

A general drawback of DDE models is they do not realistically account for sources of variance, which can lead to unrealistic projections. This limitation has been explored with time-distributed delays, in general from a theoretical/numerical perspective with limited applicability (Cooke and Grossman, 1982; MacDonald, 1986; Eurich *et al.*, 2005). In Nelson *et al.* (2013), integral delay-differential equations parameterised with individual-scale laboratory data enable studying dominant patterns of an exceptional 51-year time series of an insect (tea tortrix) population. Although a promising approach, their methods do not enable estimation parameters with field data.

1.2.2 Stage-duration distribution models

Transitions among life stages can be characterised by the distribution of time spent in each stage. This is the approach adopted by stage-duration distribution models (SDDMs) – a statistical approach that accommodates naturally the variation typically reported in development data. This class of models, also known as “survival analysis models” or “failure time models”, have been developed in other fields for similar data, such as time to machine failure, competing risk models, disease onset, or time to death (Lindsey and Ryan, 1998; Wong *et al.*, 2005; Kalbfleisch and Prentice, 2011).

Models of stage-duration distribution (SDD) assume individual stage-durations are distributed according to a parametric family of densities, including including Erlang, gamma, Weibull, log-normal, logistic, inverse Gaussian and others (Read and Ashford, 1968; Bellows Jr and Birley, 1981; Dennis *et al.*, 1986; Breteler *et al.*, 1994; Hoeting *et al.*, 2003; Knappe *et al.*, 2014). In this way, SSDMs integrate readily the effects of individual developmental variation and can be fitted to survival data for statistical comparison.

This approach has been used to model the temporal progression of multistage systems by integrating to matrix-like forms (De Valpine, 2009) or cohort-based developmental models with stochastic process (Dennis *et al.*, 1986). The approach has been used to estimate mortality and correlated SDDs (via Gaussian copulas) in stable environments (De Valpine, 2009; Knappe *et al.*, 2014; De Valpine and Knappe, 2015), and to compare phenology between populations (Murtaugh *et al.*, 2012).

Common limitations of SSDMs include a shared distribution parameter or fixed variance across all stages (Read and Ashford, 1968; Hoeting *et al.*, 2003; Manly, 1990; Aubry *et al.*, 2010), or a poor treatment of mortality: either assumed constant through all stages or just neglected (Hoeting *et al.*, 2003). Also, the fact that in data from unmarked individuals in distinct cohorts the same individuals are repeatedly assessed introduces a non-independence that is rarely incorporated to the models (Gouno *et al.*, 2011; Knappe *et al.*, 2014).

This lack of flexibility (for example, mortality assumptions and estimation methods are highly conditioned on whether data is of marked or unmarked individuals) explains that SSDMs are often based on purely empirical methods, do not handle time-varying parameters easily, and focus on estimation without reference to a wealth of knowledge. Thus, the ability of SSDMs to represent dynamical processes is strongly limited. Although SSDMs permit simulation of individual-based models (Dennis *et al.*, 1986; Régnière and Powell, 2013), the unrealistic link with time-varying parameters and heavy computational demands prevent such approaches gaining popularity for ecological studies *in natura*.

1.2.3 Matrix population models

Matrix population models (MPMs) are probably the most popular empirical tool to describe the life-cycles of structured populations. In matrix models, individuals are grouped according to a discrete range of states, and transition probabilities define the dynamics. Popular states used to characterise population are age (Leslie matrices, Leslie (1945, 1948)) and discrete developmental stages such as egg, larva, adult, among others, (Lefkovich

matrices, Lefkovitch (1965)). Although popular in demographic studies, a clear limitation of Leslie models is that the dynamics of many populations is not well described by chronological time, that discretising age can be quite arbitrary and that sensitivities and elasticities are influenced by the choice of discretisation.

Many other structures are also used to define populations by matrix methods. These can include sex, genetic aspect, physiology, developmental status, health status, among others (Caswell, 2006).

For any MPM, a population vector N is projected forward via a transition matrix M by a single time step t . Thus at any t , $N_{t+1} = MN_t$, with N_0 the starting population vector. Elements of M describe the probability of transition between stages (i.e. development and survival), and individual contributions to newborns (fecundity), which are the key processes underlying most life-cycles (Fig. 1.1).

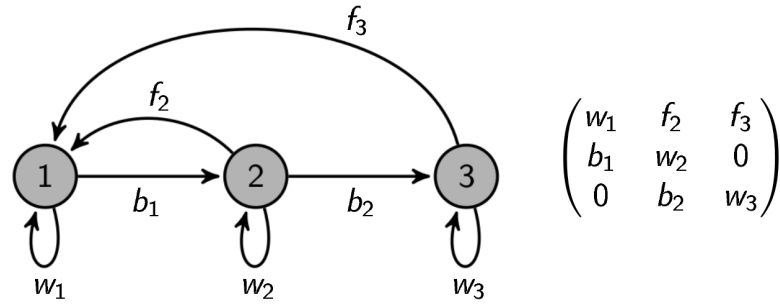


Figure 1.1. A typical life-cycle graph (left) and the corresponding projection matrix, M , of the population (right). Elements b_i and w_i describe transition probabilities of stage i , and $1 - (b_i + w_i)$ is the associated mortality probability. Elements f_i account for expected fecundity.

Long-established analytical methods exist to evaluate MPM and provide number of metrics that describe elements of the population dynamics. These include metrics of aspects of long-term dynamics (growth rate and stable distributions) (Caswell, 2006); and short-term or transient dynamics, which become especially important under fluctuating environments (Koons *et al.*, 2005; Tenhumberg *et al.*, 2009; Stott *et al.*, 2011, 2012). Moreover, techniques for studying the sensitivity of these metrics to changes in vital rates are well known (Cushing *et al.*, 2002; Caswell, 2006).

Analysis and simulation with MPMs has featured numerous levels of biological complexity including density-dependence, dispersal and time-varying parameters (Van Tienderen, 1995; Cushing *et al.*, 2002; Caswell, 2006). Detailed discussion can be found in Cushing *et al.* (2002) and Caswell (2006). In all cases, the basic matrix M is adapted to account for one or various model assumptions. For example, density-dependence can be incorporated by making at least one element of M a function of at least one element of N . This “trick” is often employed in

ecological modelling to avoid undesirable linear dynamic behaviour (i.e. populations can only explode or crash) of density-independent models. Matrix elements can also be parameterised in terms of explanatory variables allowing environmental or climatic dependence to be introduced. Similarly, stochastic effects can be introduced by perturbing some of the matrix elements by some suitably generated random numbers.

Standard MPMs rely on a Markov process assumption at the between-stage level – in other words, the probability of making the transition from one stage to another does not depend on how long an individual has been in any given stage since the model has no memory of the history of each individual within each stage. The Markov assumption produces geometric distributions for within-stage maturation times, implying that the most common stage duration for an individual is just one time step. This is a quite unrealistic assumption, giving place to poor approximations of stage-duration distributions and unrealistic projections of dynamics under natural conditions (Lefkovitch, 1965; Tuljapurkar *et al.*, 2009; Salguero-Gomez and Plotkin, 2010; Bolnick *et al.*, 2011; De Valpine *et al.*, 2014; Vindenes and Langangen, 2015). Even more, it is well known that the distributions of stage-duration impact population growth rate, sensitivities and elasticities (Caswell, 1983; Birt *et al.*, 2009; De Valpine, 2009), and that associated projections are only valid when stage distributions are stable. But, surprisingly, this reliance on the Markov assumption is quite ubiquitous in biological applications of MPMs.

Perhaps the simplest proposed solution to minimise errors arising from this approximation is to find an optimal time step Δt that minimises this bias (Cushing *et al.*, 2002). However, this approach is unlikely to work when maturation times differ greatly among stages or across time. Another commonly proposed solution is to split a compartments into r sub-compartments. This does pose the question of how to choose r since it is essentially a shape parameter of the stage-duration distribution. This has been the approach underlying age-size (Longstaff, 1984; Law, 1983; Schaalje and van der Vaart, 1989; Zuidema *et al.*, 2009) or *stage-duration-age* structured matrix models (Plant and Wilson, 1986; Caswell, 2006; Birt *et al.*, 2009). But in all these cases, different limitations appear. For example in Plant and Wilson (1986), independent stage-durations do not handle correlation between stages. More importantly, in all cases variance is tied to r , which greatly hampers the inclusion of time-varying covariates.

To summarise, the geometric stage-duration distributions of the classic Lefkovitch matrix (CLM) provides highly erroneous projections when stage-distributions are not constant. Despite several attempts to obtain more realistic stage-duration distributions by augmenting these matrices to include numerous substages, to date no

framework exists which permits a realistic incorporation of covariates without a large loss of parsimony. In the next section we show that integral projection models offer an appealing alternative to this potential drawback of augmenting CLMs.

1.2.4 Integral projection models

The integral projection model (IPM) (Easterling *et al.*, 2000b; Ellner and Rees, 2006) is as an alternative to MPMs when life cycle parameters are a function of some continuous attribute x , such as body size, mass, length or internal development. These models use individual-level data to estimate demographic functions, i.e. parametric models for demographic processes specified in the form of vital rates such as growth, maturation, survival, birth, and fertility, which are synthesised by a “kernel” function that redistributes individuals at every (discrete) time step t .

In the IPM, the population is represented by a distribution function $n(x, t)$, where $n(x, t)dx$ is the number of individuals with their state variable in the range $[x, x + dx]$ at time t . For simplicity, let's consider that x represents size. Between times t and $t + 1$, individuals can grow or die, and they can produce offspring with different sizes. An IPM updates the distribution $n(x, t)$ via

$$n(y, t + 1) = \int_U^L K(x, y)n(x, t)dx, \quad (1.1)$$

where $[L, U]$ is the range of possible sizes, and the net result of survival and reproduction is summarized by the kernel K ,

$$K(x, y) = P(x, y) + F(x, y), \quad (1.2)$$

with $P(x, y)$ representing survival and growth from state x to state y , and $F(x, y)$ accounting for offspring of size y offspring given that the parents had size x .

The dynamics is determined by the expression (1.1), a continuous-size analogue to the projection of a matrix model where the transition matrix is a size-based CLM. Thus, the kernel K is analogous to the CLM, with the advantage that in the IPM framework, by integrating the effect of $K(x, y)$ over all values of attribute x , between-individual variability is naturally accounted for. This appealing feature of IPMs, as well as the fact that these models retain much of the machinery of matrix models, has made IPMs an increasingly popular tool in ecological

and demographic modelling since the seminal work of Easterling *et al.* (2000b).

Although defined for continuous states, kernel discretisations enables treating IPMs as high resolution matrix models. Nonetheless, an intrinsic difference between MPMs and IPMs is that MPMs typically use observations to directly parameterise transition probabilities – which is their main weakness when augmenting matrix dimension to account for developmental substages – whereas IPMs infer the transition probabilities based on time-lagged regression of x giving rise to an estimation of kernel K . Thus, a major advantage of IPMs is the natural statistical treatment of variation and uncertainty they permit.

The approach has become popular in plant and animal studies where key traits are easily measured. The framework readily incorporates density-dependence, stochasticity and spatial structure (Childs *et al.*, 2003; Ellner and Rees, 2007; Jongejans *et al.*, 2011; Coulson *et al.*, 2011; Ozgul *et al.*, 2012; Merow *et al.*, 2014). An increasing literature on IPMs (or IPMS coupled to other modelling schemes) to study ecological aspects as diverse as organism development (Smallegange *et al.*, 2014, 2016); genetic traits (Coulson *et al.*, 2011; Vindenes and Langangen, 2015); infection status (Bruno *et al.*, 2011); host-parasite interactions (Metcalf *et al.*, 2015); and covariates such as abiotic environments (Metcalf *et al.*, 2009; Dalglish *et al.*, 2011) or time-lags (Kuss *et al.*, 2008), demonstrates the potential of this approach.

A strength of IPMs has been the ability to derive vital rates using regression methods to analyse observed state data (Merow *et al.*, 2014). In fact, as far as we know, all IPM-based studies rely on this regression-based approach. But IPMs can also be used when the state variable describing variation in vital rates is prohibitively difficult to be measured, i.e. when a *hidden* state such as accumulated contamination, parasitic load, physical damage or degree-days is a pertinent predictor of dynamics. To our knowledge, only De Valpine (2009) has explored IPMs with hidden states (e.g. internal or within-stage development). Although accommodating individual variation, a major drawback of the model is it does not handle fluctuating environments and that stage-duration are assigned at birth and do not change during individual lifespan. Thus, the challenge to generate methods that take advantage of IPMs to include within-stage development while enabling the incorporation of time-varying covariates and estimation from development time data has received insufficient attention. This deficit is, basically, what this thesis will address.

We adopted the Bayesian framework for estimation. In the next section a brief overview regarding the use of Bayesian estimation methods in ecology is given.

1.3 Parameter estimation in population models

1.3.1 Overview

Biologists are increasingly expected to provide estimates of responses of population at various levels . The ultimate aim of population dynamics is to give as accurate as possible predictions of the abundance of a given species in space and time. Advances in dynamic theory during the last century relied heavily on the deterministic mechanistic models and analytical techniques. By definition, deterministic models lack environmental and/or demographic stochastic components (Murray, 2002). When applied to biological systems, given their natural complexity – and, as in most of cases, irreproducibility of processes –, deterministic models generate trajectories that always diverge from the processes they seek to emulate. Such effects risk to lead to inaccurate conclusions or projections (Petchey *et al.*, 2015).

Among the new strategies developed to overcome the limitations of purely mechanistic approaches, those arising from a marriage between population dynamics and statistical theories/techniques for the analysis and interpretation of data are becoming increasingly popular. For the greater part of the 20th century these two paradigms, statistics and dynamics, developed more or less independently and rather at arms length from each other.

Dynamic models hypothesize the nature of relationship in terms of the biological processes that are thought to have given rise to the data. The parameters in the mechanistic model all have biological definitions and so they could – hypothetically – be measured independently of data sets generated by the process in question. By contrast, in statistics, the choice of data generation model was traditionally guided by mathematical convenience instead of mechanistic considerations, but the role of data for estimation and validation is central to the paradigm.

Reluctance in both camps to bridge the gap arose mainly from the huge technical difficulties involved in bringing these two disciplines together. This started to change with the revolution of personal computers in the last half century and the increasing demand for environmental forecasting for practical purposes.

The integration of data into dynamic models has been termed “data assimilation” and systems designed for such a task have been called “integrated model-data systems” or “hybrid mechanistic-statistical models” (Liu, 2008). The aims of data assimilation are to quantify the predictive performance of alternative models, identify where model predictions can be improved and to quantify errors or uncertainties in model predictions and parameter estimates.

As models become more complex, new parameters are included and augment the sensitivity of outputs to parameters values. Therefore, an accurate assessment of parameter estimates and associated uncertainties is crucial for the predictive power defining the utility of models. The Bayesian approach is a natural paradigm that permits to assess uncertainties in estimates and predictions.

1.3.2 Bayesian inference: a brief description

Bayesian inference is an approach to statistical analysis in which all forms of uncertainty are expressed in terms of probability distributions (Bernardo and Smith, 2000; Gelman *et al.*, 2003). In this method, Bayes' theorem (Bayes *et al.*, 1763) is used to update the probability for a hypothesis as more evidence or information becomes available. A Bayesian approach to a problem starts with the formulation of a data generation model – containing either deterministic and/or stochastic components – that is hoped to describe adequately the sources of variation in the data of interest. Next, a prior distribution $f(\Theta)$ over the unknown model parameters Θ is formulated, which is meant to capture our beliefs or knowledge about the situation before incorporation the data y . These beliefs are adjusted by confronting the model to data via the likelihood function, $f(y|\Theta)$, and using Bayes' rule to obtain a posterior distribution $f(\Theta|y)$. This posterior distribution provides us with an updated representation of the state of knowledge concerning the likely values that Θ might take. The analysis can be crafted such that the posterior distribution includes predictions for unobserved observations, for example, populations densities at unsampled points in space and time.

The description above is summarised in the following relationship:

$$f(\Theta|y) = \frac{f(y|\Theta) \times f(\Theta)}{\int \dots \int f(y|\Theta) \times f(\Theta) d\Theta}. \quad (1.3)$$

Although the relationship 1.3 has been known ever since the 18th century, its application was traditionally severely limited by the intractability of the integrations required to correctly normalise the posterior distribution. With the advent of digital computers over the last half century, increasingly sophisticated iterative algorithms have been developed to overcome this difficulty. The basic idea is to draw a large set of random variables from the posterior distribution from which summary statistics, such as the mean or various quantiles of interest, can be approximated.

In practice it may not be known how to directly generate such a set of random variables, but it is often

possible to build a Markov chain which has the target posterior as its stationary distribution. In other words, the Markov chain governs a random walk that explores the target distribution and provides a series of samples which, if sufficiently large, can share many of the statistical properties of the desired set of independently generated random numbers. Such methods are known as "Markov chain Monte Carlo" (MCMC) techniques (Hammersley and Handscomb, 1964; Kalos and Whitlock, 2008; Liu, 2008).

A plethora of MCMC methods have evolved from two key techniques : the Gibbs sampler; and the Metropolis-Hastings (MH) sampler (Gelman *et al.*, 2003). The MH sampler is used when the normalisation constant in equation (1.3) is analytically intractable. The required random walk through the parameter space is obtained by proposing modifications Θ' to a given parameter set Θ and accepting the proposed update with probability

$$\min \left(1, \frac{f(y|\Theta')f(\Theta')f(\Theta', \Theta)}{f(y|\Theta)f(\Theta)f(\Theta, \Theta')} \right), \quad (1.4)$$

where $f(\Theta, \Theta')$ is a proposal kernel and $f(\Theta', \Theta)$ is the likelihood of making the reverse proposal. It can be proved that under certain conditions such a random walk indeed has the target distribution as its stationary distribution. The power of the algorithm lies in the fact that the problematic normalising constant cancels in the ratio term of 1.4, so it is sufficient to know $f(y|\Theta)$ in unnormalised form.

In practice, the choice of proposal kernel $f(\Theta, \Theta')$ critically affects whether or not a good approximation can be achieved within reasonable time limits. If $f(\Theta, \Theta')$ systematically generates unlikely proposals, rejection rates become excessively high. On the other hand, if $f(\Theta, \Theta')$ systematically generates proposals Θ' that are negligibly different to Θ , acceptance rates are high but the time required to explore the target distribution can become unreasonably long. In both situations it is said that the chain exhibits "poor mixing". For complex models, how to generate an efficient proposal distribution to achieve a good compromise between these two extremes is often a non-trivial problem.

The Gibbs sampler is used when generating random samples of multiple parameters jointly from the target distribution is intractable or impracticable, but it is possible to generate random samples of subsets of parameters conditionally on other parameters being fixed. By iteratively switching which parameters are fixed / sampled, a random walk is generated that can sample unknown posterior (Smith and Roberts, 1993). The Gibbs sampler can be seen as a special case of the MH sampler where proposals are generated from subsets of the parameter set. If these proposals are generated from a conditional distribution derived from the target joint distribution then

acceptance rates will be one and it is unnecessary to calculate (eq. 1.4). However, often it is more common to use MH samplers within a Gibbs sampler to handle conditional distributions from which samples cannot be generated directly (i.e. when the normalising constant of a conditional distribution is also intractable). Such a combination is sometimes called a Metropolis-within-Gibbs sampler (Roberts and Rosenthal, 2006).

Nimble, a software for Bayesian estimation. These, and other, MCMC algorithms have become readily available to via the BUGS (Bayesian Inference Using Gibbs Sampler) (Spiegelhalter *et al.*, 1996) language. Various software implementations are now available, including WinBUGS, JAGS, openBUGS and NIMBLE (Lunn *et al.*, 2000; Plummer *et al.*, 2003; Cowles, 2013; De Valpine *et al.*, 2016). The BUGS language uses a convenient and intuitive pseudo-code to construct a directed acyclic graph that represents deterministic and stochastic dependence between various variates in a Bayesian model. A major limitation of BUGS is that users are constrained to make use of canonical distributions in the models. Users requiring to write custom functions must therefore do so in a low level language such as C.

NIMBLE offers a flexible new system for building BUGS-like models in R that automatically compiles code to C++. The main advantages of NIMBLE are that it permits generating user-defined functions and distributions in an easy, R-like language, and that C++ compilation drastically improves computational times. User-defined samplers can also be generated likewise, giving a flexibility that at the same time, remains relatively accessible for non-computational ecologists. Another key advantage of NIMBLE is it provides a suite of functions to simulate BUGS models and enables a separation of algorithm and model that, compared to writing custom MCMC samplers for a model, greatly simplifies scripts for analyses. The model and analyses of chapters 2 and 3 were first written in R, but were converted to NIMBLE for its greater clarity and efficiency.

1.3.3 State-space models

Overview

The state-space model (SSM) is a statistical framework for time-series data that allows including two sources of variability namely some dynamic process and measurement error (Newman *et al.*, 2014). A SSM combines a process model of the dynamic system; an observation model that links data to the process model; and an algorithmic component that fits the model to data, generates predictions and quantifies uncertainties. State-space models reduce bias and provide more accurate estimates of uncertainty than methods that do not fully incorporate

both sources of variation (De Valpine and Hastings, 2002).

Formally, a SSM is a Markov process with two parts: the true unobserved state process $x_{1:t}$ (with the sub-index indicating the series of discrete time-steps from 1 to t); and the observation process $y_{1:t}$ that models the way data y_t is generated given the hidden state x_t (Fig. 1.2). The SSM depends on vector of parameters θ . Calibrating the model, i.e. obtaining estimates for θ and $x_{1:t}$ that provide likely explanations for the data, is obtained via statistical inference. In the Bayesian paradigm this involves computing the posterior distributions such as $p(x_{1:t}|y_{1:t}, \theta)$, $p(x_{1:t}, \theta|y_{1:t})$, or $p(\theta|y_{1:t})$. In most of cases, the constant required to normalise these posterior distributions is analytically intractable and thus sampling-based methods are required to get estimates.

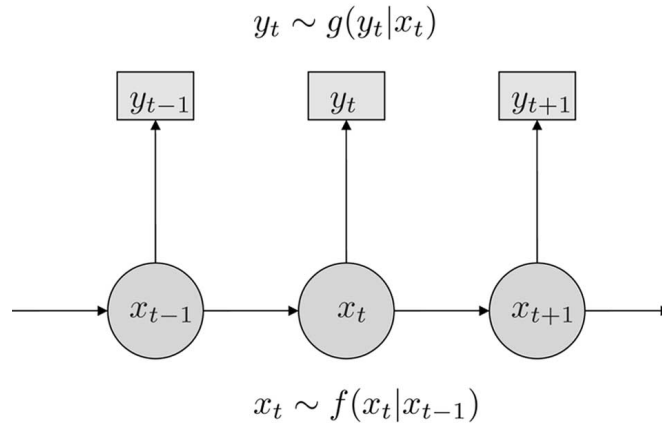


Figure 1.2. Schematic representation of a state-space model. The “state” of the hidden process x evolves conditionally given the state at the previous time step. The Markov transition rule $f(x_t|x_{t-1})$ can contain any combination of deterministic and stochastic elements. Each observation y_t is generated conditionally on the hidden state x_t . Modified from Cappé *et al.* (2007).

Methods for estimating parameters in a SSM framework.

When observations are informative, they can be used to prevent divergence between simulated trajectories and the true unknown (i.e. only partially observed) dynamic process. This uses algorithms that are collectively known as “filtering” since they filter out projections that do not fit the observed data well.

Filtering methods are popular choices for fitting SSMs. The target distribution of most filtering algorithms is $p(x_{1:t}|y_{1:t}, \theta)$. Among these algorithms, the Kalman filter (Kalman, 1960) has been a standard choice for efficiently fitting SSMs when the dynamics are linear and the observation model is Gaussian. Unfortunately, ecological dynamics rarely remain linear for more than very short durations of time and observations are more typically counts or presence/absence records implying that a Gaussian observation model can be a source of bias. Various variants to the Kalman filter have been proposed for very specific non-linear filtering problems (Routray

et al., 2002; Evensen, 2003; Lefebvre *et al.*, 2004), but these variants have failed to provide consistent reliable estimates for general non-linear and non-Gaussian systems.

More general dynamic systems, including those with chaotic or near chaotic behaviour, have been treated in the SSM framework via alternative strategies including particle filters, MCMC techniques, information reduction approaches (e.g. approximate Bayesian computation (ABC) or synthetic likelihood (SL)) and hybrid samplers (e.g. particle MCMC). These four approaches are briefly described below in order to highlight the most suitable for estimation with our IPLM models.

Particle filtering, also known as sequential Monte Carlo (SMC), offers a general recursive and computationally tractable solution for Bayesian approximation of $f(x_{1:t}|y_{1:t}, \theta)$ for nonlinear and non-Gaussian problems (Del Moral, 1996; Cappé *et al.*, 2007; Liu, 2008). The SMC sequentially approximates the marginal distribution of the latent process *on the fly*, i.e. as new observations are incorporated. These algorithms are somewhat analogous to Darwinian natural selection. At each time t , the distribution $f(x_{1:t}|y_{1:t}, \theta)$ is approximated with a set (or population) of discrete vectors $\{\hat{x}_{1:t}\}$. The elements of this set are called particles, each with an assigned weight. The weights are analogous to Darwinian fitness and are used to ensure survival of the fittest. Whether or not a proposed sequence $\hat{x}_{1:t}$ survives to the next time step depends on how likely the sequence appears to be, given θ and $y_{1:t}$, and weights (fitness) are used to remove individuals from the population via *importance sampling*.

To maintain a constant population size, a resampling step is included: particles with low importance weights are replaced by multiple copies of those with high importance weights. This resampling has a well known pitfall: if $f(x_t|x_{t-1}, \theta)$ is poorly specified then excessive loss of particle diversity at this step can result in $\{\hat{x}_{1:t}\}$ only approximating a narrow subset of $f(x_{1:t}|y_{1:t}, \theta)$ – in our Darwinian analogy this is somewhat equivalent to a loss of genetic diversity in an isolated sub-population. This problem can be tackled by increasing the number of particles N . However, in high dimensional cases as the one we are envisaging, this renders the method computationally inefficient. Moreover, a key limitation of SMC is that θ is fixed and not estimated, which hampers ecological applications.

In general Markov chain Monte Carlo (MCMC) methods outperform particle filters in high-dimensional parameter estimation problems. However, a key exception is the estimation of $f(x_{1:t}, \theta|xy_{1:t})$ where traditional MCMC algorithms rapidly become inefficient as time series length t increases. The limiting factor arises from the difficult methodological challenge of generating joint proposals $\{x'_{1:t}, \theta'\}$ that result in efficient exploration of

$f(x_{1:t}, \theta | y_{1:t})$. Recently, new hybrid strategies have been developed specifically to overcome this problem. Among these, the most celebrated are the particle MCMC (PMCMC) (Andrieu *et al.*, 2010) and the sequential Markov Chain Monte Carlo (SMC2) (Chopin *et al.*, 2013) algorithms.

The PMCMC uses particle filtering to generate state-space sequence proposals $x'_{1:t}$ within a MCMC algorithm. This has provided efficient mixing for previously intractable non-linear non-Gaussian SSM inference problems; and has been profusely used and extended since its publication (Rasmussen *et al.*, 2011; Frigola *et al.*, 2013; Donnet and Samson, 2014). The SMC2 algorithm is a sequential Monte Carlo algorithm that re-samples at every time step the parameter space via a MCMC update. Despite great flexibility in terms of the types and structure of data allowed by both PMCMC and SMC2, computational cost is still a critical factor that hampers more widespread use of these methods. A recent free software (LibBi, Murray and Daszak (2013)) for Bayesian inference of SSMs has made PMCMC and SMC2 widely available and permits parallel computing on Graphical Processing Units (GPUs) to reduce computation time.

Approximate Bayesian computation (ABC) is a popular alternative to model-based inference that is particularly powerful when likelihood functions of dynamic models are intractables (Tavaré *et al.*, 1997; Pritchard *et al.*, 1999). The ABC approach bypasses the need to evaluate likelihood functions by comparing repeated simulations with an observed data set via a set of summary statistics $S(y)$ of data y . Another recent simulation-based approach for inference in SSMs is synthetic likelihood (SL) (Wood, 2010; Fasiolo and Wood, 2015). Both methods focus on the relationship between some characteristic features of the data and the unknown parameters. For this, the observed and simulated data are transformed into a set of summary statistics that aim to capture the essential characteristics of the dynamic process and permit that subsequent inferences is based on these characteristics. Although both ABC and SL make use of a *ad hoc* vector of summary statistics $S(y)$, their main difference is that ABC uses a nonparametric-style density estimator for $f(S(y)|\theta)$ whilst SL takes a parametric form – the “synthetic likelihood” – assumed to be a Gaussian distribution $f(S(y)|\theta) \sim \mathcal{N}(\bar{S}_{\text{sim}}, \Sigma_{\text{sim}})$, where \bar{S}_{sim} and Σ_{sim} are the mean and covariance matrix of the vector of summary statistics $S(y)$. In a Bayesian context, the SL methods essentially marginalise across uncertainty in the state space x to give estimates via the posterior distribution $f(\theta|y)$.

For complex problems, the choice of the summary statistics should be based on the results of exploratory analyses. This demands iterative model checking to identify discrepancies in fit, which in turn can suggest extra statistics to incorporate in a revised SL. This drawback is largely compensated by the simplicity of the approach.

1.4 Conclusion

1.4.1 Motivating example

The models and methods I present in this thesis are generic in the sense that they could be applied to the life-cycle of many organisms with instars or eggs. Where sufficient developmental data for time-lagged regression exist (e.g. size, weight), the integral projection model is a suitable and parsimonious alternative. But for species where recording size or weight is either impracticable or prohibitively difficult, but maturation time data is available or easily obtained, the model and methods presented here are a valuable addition to the toolbox for modelling stage-structured populations. This is the case for *Culicoides*, biting midges, a genre of insects of epidemiological interest (Carpenter *et al.*, 2013) that we used in our case studies of chapters 3 and 4.

Important gaps exist in the knowledge of how life history parameters (mainly those associated to immature stages) of *Culicoides* depend on biotic and abiotic variables, under both laboratory and natural conditions (Mullens *et al.*, 2015). A life-cycle model with more demographic and biological realism than currently available models could provide a basis for exploring these interactions, and, potentially, could help analyse phenological data and identify the most influential demographic parameters, i.e. the potential control points in the *Culicoides* life-cycle.

In the appendix B.1, an overview of biological and epidemiological aspects of *Culicoides* biting midges is provided.

1.4.2 Aims and objectives

The aim of this work was to develop a predictive model for the dynamics of structured biological populations that includes the ubiquitous heterogeneity in developmental processes while enabling estimation of vital rate responses to exogenous factors – particularly temperature – for approximating seasonal variation of populations *in natura*.

Achieving this aim implies addressing the following objectives:

- Formulate a mechanistic model that can reproduce observed variance in developmental processes and enables realistic incorporation of time-varying covariates, while remaining flexible enough to incorporate other demographic (i.e. density-dependence, endogenous stochasticity, correlation in stage-durations) and environmental (i.e. multiple covariates, exogenous stochasticity) complexities.
- Develop Bayesian methods for parameter estimation with maturation time (i.e. stage-duration) and mortality

data from typical laboratory studies (i.e. that readily handle censoring and missing data).

- Analyse model performance (i.e. identifiability) with the help of simulated data prior to incorporating real data into the model.
- Fit the model to *Culicoides* maturation data at various fixed temperatures and estimate associated uncertainties.
- Develop and identify suitable functional response curves for linking *Culicoides* life history parameters to environmental variation at unsampled temperatures.
- Analyse the consequences of estimated *Culicoides* vital rate parameters on long-term and transient dynamics at fixed and time-varying temperatures, as well as in potential applications to the study of phenology and transmission risk of an animal disease vectorred by *Culicoides* midges.
- Extend the model framework to state-space models in order to upscale to natural scenarios that permit the integration of field data (e.g. time series) to reduce uncertainty and permit simulations for predicting the responses of state-structured populations to global change or local anthropogenic interventions.

1.4.3 Thesis outline

This thesis is divided into five chapters.

Chapter 1 gave a general introduction. The framework and motivations of the present work were described. Next, an overview on different modelling approaches of state-structured populations with particular emphasis on how different approaches handle developmental variation and its dependence on covariates were presented. A short description of relevant Bayesian estimation methods were outlined. Finally, a brief description of the motivating example was provided and general aims and objectives were stated.

In chapter 2, I introduce and develop in detail a novel predictive model, the “integral projection Lefkovitch matrix” (IPLM) model for stage-structured populations. I show with a simulated study (i) the flexibility of IPLMs to incorporate random effects and correlation between stage duration distributions; (ii) Bayesian estimation methods that use maturation-survival data, and handle censoring appropriately; and (iii) the estimation performance of Bayesian IPLMs. I conclude that Bayesian IPLMs show great promise as a framework for parameter estimation and model inference for more realistic stage-structured population models and suggest potential extensions, some of which are explored in the remaining chapters.

In chapter 3, the IPLM approach is extended to include covariates on demographic parameters. Real laboratory data at various fixed temperatures from two ecologically similar *Culicoides* species is used in the Bayesian framework to obtain, via spline regression, functional responses of stage-specific vital rates to temperature. Long-term and transient dynamics are analysed. I show that traditional negligence of individual developmental heterogeneity affects asymptotic growth rate estimates and greatly undermines the importance of transient oscillations in the population density. A brief assessment of the robustness of the linearity assumption of classic degree-day phenological models is performed, using the fitted *Culicoides* IPLM model. The results provide little support for linear degree-day models.

In chapter 4, two potential lines of research are explored. First, I use the *Culicoides* IPLM model to analyse temporal variation in the R_0 of bluetongue (see table 1.2) under an adult control scenario. The implications of our results for integrated vector management are outlined. Next, I explore methods to upscale to natural scenarios by using IPLMs in a state-space model in order to integrate field data (e.g. adult time series). I explore potential methods for inference and use a simulation-based approach (synthetic likelihood, Wood (2010)) in a preliminary study with simulated observation data. Methodological challenges and potential improvements are outlined. Finally, in chapter 5, the main developments and findings of this work are summarised and concluding remarks are given.

Chapter 2

A model for individual heterogeneity in hidden developmental processes

The motivating problem in this chapter ¹ is the study of organisms whose development occurs through discrete stages. Our ultimate objective is to gain insight into how variability in developmental response of individuals influences the vital rates of a stage-structured population. For this, we develop a new model that, by including within-stage development dynamics – i.e. stage-specific hidden developmental states –, provides more realistic stage-duration distributions of such populations. We demonstrate these methods by generating *in silico* maturation-time data of a cohort where correlation is included via individual qualities, and evaluate estimation performance of the new model with Bayesian methods that use stage-duration and survival data for estimating vital rates.

2.1 Introduction

A central premise of population biology is that a population's dynamics are driven by the timing of life-cycle events (Caswell, 2006). When life-cycles progress via a series of developmental stages, describing how each stage's vital rates vary with genetic and environmental factors provides a basis for studying a population's dynamics (Manly, 1990). Analysis and simulation with stage-structured models (SSMs) has featured numerous biological complexities including density-dependence, stochasticity, time-varying parameters and dispersal (Van Tienderen,

¹Most of the work presented in this chapter is part of an article submitted to the journal *Methods in Ecology and Evolution* (see Appendix D).

1995; Cushing *et al.*, 2002).

In the previous chapter (section 1.2) we have seen that assumptions of different SSM approaches (delay differential equations, stage-duration distribution models and matrix population models) regarding variation in the time required to mature through a given stage (i.e. the maturation-, development- or sojourn-time), are oversimplified, unrealistic and lack generality. We have also highlighted that such oversimplification can result in poor stage-duration distribution (SDD) approximations and inaccurate predictions of population growth rates and related quantities (Vindenes *et al.*, 2008; Bolnick *et al.*, 2011); and that models with greater generality are required if the forecast horizon (Petchey *et al.*, 2015) of SSMs is to be increased.

We have seen (section 1.2.4) that integral projection models (IPMs) enable linking individual variation (and covariates) to key population-level developmental parameters. Insufficient attention has been given to (IPMs) regarding their ability to model life-cycles with hidden developmental states where typical time-lagged regression-based parameterisation is either impracticable (e.g. parasitic load, physical damage) or a poor predictor of stage-duration (e.g. when eggshells, exoskeletons or hosts effectively hide within-stage development).

Here, we propose to formulate stage-duration distributions in terms of an internal development state whose dynamics follow an integral projection model. More precisely, we extend standard matrix models by incorporating IPM approximations that track individuals through a series of developmental substages to yield more realistic stage-duration distributions. Estimation can be based upon treating within-stage development as an unobserved state variable and by fitting model outputs to maturation-time data.

These new “integral projection Lefkovitch matrix” (IPLM) models inherit the analytical advantages of matrix models and the development heterogeneity of IPMs, allowing thus the distributional flexibility of SDDMs. Additionally, the IPLM framework enables realistic incorporation of (and parsimonious parameterisation with) both time-varying covariates and/or unmeasured local or genetic factors.

By tracking within-stage development, the new methods we develop greatly reduce errors in projected stage-duration distributions and provide valid transition probabilities for non-stable stage distributions. In consequence, errors in transient or non-linear dynamics analyses can be reduced and might improve forecast horizons of stage-structured models.

In section 2.2 we outline the IPLM framework and describe how these models can be parameterised using either marked or unmarked maturation-time data. Then, in section 2.3, a simulation-estimation study where individual

qualities correlate between-stage maturation times is developed. Bayesian inference is performed to investigate the identifiability of estimates assuming individually marked data. Finally, results are discussed in section 2.4.

2.2 Integral projection Lefkovitch matrix models

2.2.1 The classic Lefkovitch matrix (CLM) model

The following recursive formula is a popular tool for studying demographic dynamics

$$\mathbf{N}_t = \mathbf{M}_t \mathbf{N}_{t-1}, \quad (2.1)$$

where \mathbf{N}_t denotes a vector of densities for a series of k age (Leslie, 1945) or stage (Lefkovitch, 1965) classes at time t and \mathbf{M}_t is a projection matrix. While Lefkovitch (i.e. stage-structured) matrices can be constructed in many ways to match the great diversity of life-cycle strategies found in nature, here we focus on matrix models of the form

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \dots \\ n_k \end{bmatrix}_t = \begin{bmatrix} W_1 & F_2 & F_3 & \dots & F_k \\ B_1 & W_2 & 0 & \dots & 0 \\ 0 & B_2 & W_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & B_{n-1} & W_k \end{bmatrix}_t \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \dots \\ n_k \end{bmatrix}_{t-1}. \quad (2.2)$$

We call the matrix in (2.2) a ‘classic Lefkovitch matrix’ (CLM), noting that our methods can generalise to matrices for other stage-structured life cycles. A tempting misinterpretation of (2.2) is that, in time-step t , individuals in some stage $S \in \{1, \dots, k\}$ remain with probability W_S , advance one stage with probability B_S , contribute to the next generation with fecundity F_S and survive with probability $\nu_S = W_S + B_S$. However, this neglects within-stage developmental heterogeneity, assumes geometric sojourn-time distributions and only yields valid transition probabilities when stage distributions are stable (Lefkovitch, 1965; De Valpine *et al.*, 2014). Thus, such matrices can generate highly erroneous results unless vital rates are relatively constant and impervious to exogenous sources of variation.

2.2.2 Incorporating within-stage development dynamics

These limitations can be overcome by replacing scalar elements W_s , B_s and F_s of matrix \mathbf{M}_t with sub-matrices \mathbf{W}_s , \mathbf{B}_s and \mathbf{F}_s characterising within-stage development, between-stage development and fecundity respectively. Thus, every scalar n_s of equation (2.2) is replaced by \mathbf{n}_s , a vector of r_s discrete within-stage development states. Note that r_s can vary between stages.

We define sub-matrices \mathbf{W}_s , \mathbf{B}_s and \mathbf{F}_s via stage-specific IPMs. An IPM for within-stage development can be written

$$n(\delta', t) = \int_0^1 K_{\Theta}(\delta, \delta') n(\delta, t-1) d\delta, \quad (2.3)$$

where $n(\delta, t)$ is the density of individuals with developmental status δ at time t , Θ is a parameter set, and the IPM-kernel $K_{\Theta}(\delta, \delta')$ quantifies the proportion of individuals with development δ that survive and develop to δ' in one time-step. Transition to next stage occurs once $\delta \geq 1$, whereby development in the new stage is initialised with $\delta = 0$. Unlike previous models including within-stage dynamics into CLMs (Longstaff, 1984; Birt *et al.*, 2009), this approach can improve SDD approximations independently of the number of substages r_s since within-stage transition probabilities are all defined in terms of the IPM-kernel parameter set Θ .

Discretisation of a within-stage IPM. For practical purposes, we simplify the general form 2.3 by assuming that the increments by which individuals develop are drawn independently from the same distribution at each time step. Therefore, we re-write the IPM kernel $K_{\Theta}(\delta, \delta')$ as $K_{\Theta}(\Delta)$, with $\Delta = \delta' - \delta$. In the examples presented in this and the next chapter, we use for K_{Θ} a beta distribution with parameters $\{\mu, \kappa\}$ accounting for developmental rate heterogeneity, combined with survival probability ν . The beta distribution is a natural choice since δ ranges in $[0, 1)$ for each stage. This model provides the same level of parsimony as SDD models: each defines the distribution of sojourn-times and mortality with three parameters. In our model, the kernel for each stage projects individuals through a developmental process to derive the probabilities of stage completion or death in any given time interval – these probabilities provide the basis for estimating parameters from data. Consequently, some computation time is required for model fitting, but the benefit is a model formulated in discrete time-steps that can accommodate time-varying covariates.

Like other IPMs, we approximate the continuous state variable δ by a series of discrete states. Given a series of r discrete states between 0 and 1, we calculate the probability p_l of completing l discrete increments in a time-step by integrating over an interval of K_{Θ} (more details in Appendix A.1).

These transition probabilities provide, for a stage S , elements for the following $r_S \times r_S$ lower-triangular matrix:

$$\mathbf{W}_S = \nu_S \begin{bmatrix} p_{0,S} & & & & \\ p_{1,S} & p_{0,S} & & & \\ p_{2,S} & p_{1,S} & p_{0,S} & & \\ \vdots & \vdots & & \ddots & \\ p_{r-1,S} & p_{r-2,S} & & & p_{0,S} \end{bmatrix}. \quad (2.4)$$

Note, the probabilities $\{p_{0,S}, \dots, p_{r,S}\}$ depend on the stage-specific parameters $\{\mu_S, \kappa_S, \nu_S\}$. Matrix \mathbf{B}_S provides the proportion of individuals making the transition to the next stage, where development is initialised in the first substage. Thus, if \mathbf{B}_S^1 denotes the first row of matrix \mathbf{B}_S , element j of \mathbf{B}_S^1 is $\sum_{l=r+1-j}^r p_{l,S}$. Each matrix \mathbf{F}_S is constructed assuming all individuals completing stage S contribute F_S to the next generation. Thus, the first row of \mathbf{F}_S is $\mathbf{F}_S^1 = F_S \mathbf{B}_S^1$. All other elements of \mathbf{B}_S and \mathbf{F}_S are zero and $F_S = 0$ for non-reproductive stages. Alternative definitions for \mathbf{B}_S and \mathbf{F}_S are possible (e.g. transition to multiple stages or state variable allowing for processes such as shrinking), but are not explored here for simplicity.

The matrix approximation of the IPM (2.3) for stage S is therefore

$$\begin{bmatrix} \mathbf{n}_S \\ c_S \end{bmatrix}_t = \begin{bmatrix} \mathbf{W}_S & \mathbf{0} \\ \mathbf{B}_S^1 & 1 \end{bmatrix}_t \begin{bmatrix} \mathbf{n}_S \\ c_S \end{bmatrix}_{t-1}, \quad (2.5)$$

where, at each time, \mathbf{n}_S gives the distribution of population density in the r_S substages, and c_S is the cumulative density of individuals that have completed stage S . We call $\Theta_S = \{\mu_S, \kappa_S, \nu_S\}$ the parameter set of the discretised IPM-kernel.

2.2.3 The general integral Lefkovitch matrix (IPLM) model

The extension of the CLM (equation 2.2) to include within-stage dynamics is

$$N_t = \begin{bmatrix} \mathbf{W}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \dots & \mathbf{F}_k \\ \mathbf{B}_1 & \mathbf{W}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \mathbf{W}_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{B}_{m-1} & \mathbf{W}_k \end{bmatrix}_t \mathcal{N}_{t-1}, \quad (2.6)$$

where $\mathcal{N}^T = (\mathbf{n}_1^T, \dots, \mathbf{n}_k^T)$ and the model parameter set is $\Theta = \{\Theta_1, \dots, \Theta_k\}$. We call any matrix built on these principals an integral projection Lefkovitch matrix (IPLM). Because of its construction, we can note that when all $r_s = 1$, an IPLM reduces to a CLM, while when all $r_s \rightarrow \infty$, the system (2.6) becomes a canonical stage-structured IPM. In practice, we seek r_s small enough to maintain computational efficiency yet large enough to characterise sojourn-time variance for stage S . Since the dimension of Θ_s is independent of r_s , parsimony is unaffected as matrix dimension increases. Covariates or random effects can be incorporated with relative ease via Θ_s , thus IPLMs can incorporate exogenous or endogenous sources of heterogeneity. These developments can therefore greatly augment the range of scenarios studied with the powerful tools of matrix model analysis.

Beta distribution for developmental variation. The choice of a function to describe the heterogeneity in developmental rates (represented by the increments Δ) is flexible. We use beta distributions for their parsimonious flexibility on $[0, 1]$. Thus individual variation in developmental responses is modelled by the probability distribution function (PDF)

$$f(\Delta|\theta) = \frac{\Delta^{\alpha_1-1}(1-\Delta)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)}, \quad (2.7)$$

where $\theta = \{\alpha_1, \alpha_2\}$ are parameters and $B(\cdot, \cdot)$ is the beta function. Bi-modality is avoided by constraining α_1 and α_2 to be greater than one. Since α_1 and α_2 do not yield biological interpretation, we use the alternative parameterisation $\theta = \{\mu, \kappa\}$, where $\mu = E[\Delta] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$ is the expected developmental increment and $\kappa \in (0, 1)$ is a scale parameter such that $\text{Var}(\Delta) = \kappa\mu(1-\mu)$. The probability p_l (defined in Appendix A.1) of completing l

discrete increments in a time-step is thus

$$p_l = F(\frac{l+1}{r+1}|\theta) - F(\frac{l}{r+1}|\theta), \quad (2.8)$$

where $F(\Delta|\theta)$ is the cumulative distribution function (CDF) associated with $f(\Delta|\theta)$.

2.2.4 Fitting with maturation time data

We consider fitting IPLMs using either marked or unmarked cohort development data. Marked cohort data provide the time or time-interval of each stage transition for each individual. Unmarked cohort data include the number of individuals maturing from a stage in a time interval given that their development was synchronised at $t = 0$. Typically, the number dying in one or more time intervals is also reported. The harder problem of fitting an IPLM to partially observed time-series data from overlapping generations – which often arises in studies of natural insect populations – is not addressed in this section.

Since within-stage development is typically unmeasurable, the common vital rate regression strategy for IPM-kernels is unfeasible. Instead, we take a likelihood approach, which can then be used in either a Bayesian or frequentist framework. As in SDD models, the likelihood of observed stage-duration data y_s depends on the probabilities (given Θ_s) of surviving and completing stage S in each time-step. We calculate these probabilities by iterating the discretised IPM (equation 2.5). More specifically, we:

1. initialise $\mathbf{n}_s(t = 0) = (1, 0, \dots, 0)^T$ and $c_s = 0$;
2. project $\mathbf{n}_s(t)$ forward;
3. for each t , record the matured proportion c_s , and the loss of density over the vector $(\mathbf{n}_s^T, c_s)_t$, to construct a sojourn-mortality distribution (Fig. 2.1, and a more detailed description in Appendix A.3); and
4. use these probabilities to evaluate the likelihood of data y_s given Θ_s .

For marked data, individual-level covariates or random effects (e.g. individual qualities) can be included via individual-specific kernel calculations.

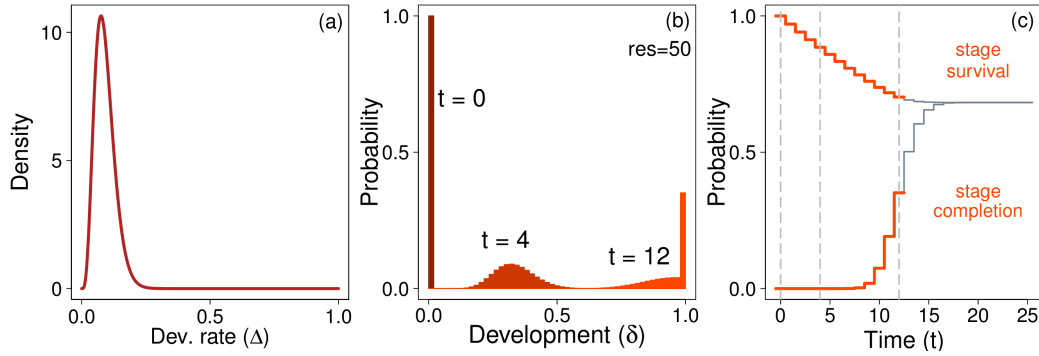


Figure 2.1. Implementing a basic integral projection model (IPM) for within-stage dynamics. The IPM-kernel K_{Θ} is defined as the product of a stage-specific survival probability ν and a probability density function (PDF) for development increments. Here, $\nu = 0.97$ and the PDF is $\text{Beta}(\alpha_1 = 5, \alpha_2 = 50)$ (a). A population, initialised in the first substage ($t = 0$), is projected forward through a series of ($r = 50$) discretised developmental increments (b). The accumulation of density in the final substage, and the loss of density over all substages, generates the “sojourn-mortality distribution” – the probabilities to complete the stage, or die in the stage, per time-step. Here, the cumulative sojourn-mortality distribution is shown with the interval $t = 0$ to $t = 12$ coloured orange (c). Dashed lines (c) correspond to the developmental distributions at $t = 0$, $t = 4$ and $t = 12$ (b).

2.3 A simulation study with correlated stage-durations

To test the identifiability of IPLM parameters, a simulation-estimation experiment was conducted. Motivated by recent directions in eco-evolution, the basic IPLM model (equation 2.6) was modified to incorporate correlated stage-durations arising from heterogeneous individual qualities.

Quality parameters are used in eco-evolution to parsimoniously quantify net effects of genetic or local factors on vital rates (Ardia, 2005; Wilson and Nussey, 2010; Vindenes and Langangen, 2015; Shyu and Caswell, 2016), and heterogeneity in individual qualities can generate correlated stage-durations (De Valpine, 2009). Despite much theoretical work, the estimation of individual qualities, associated sojourn-mortality distributions and their evolutionary consequences in real populations remains challenging. In this section, we outline how Gaussian copulas (Kruskal, 1958; Nelsen, 2006) enable individual qualities to condition IPLM kernels, and demonstrate that, even with modest sample sizes, a quality-dependent IPLM fitted to simulated marked-cohort data for just two sequential stages can accurately reproduce sojourn-mortality distributions.

2.3.1 Correlated stage-durations and Gaussian copulas

Copulas are tools for modelling correlations in arbitrary sets of random variables (Hougaard, 2012). Here, individual quality (q) and development increments (Δ) are correlated via Gaussian copulas. In this framework, we specify the marginal q - and Δ -distributions, and assume their correlation follows the same quantiles as a bivariate normal distribution with correlation parameter ρ . We assume q is fixed through an individual's lifespan which conditions implicitly (via ρ) the distribution of increments at each time-step.

Let $f_{\Delta}(\Delta|\alpha_{\Delta})$ and $F_{\Delta}(\Delta|\alpha_{\Delta})$ denote the marginal (beta) PDF and CDF of developmental increments, with parameters α_{Δ} . Let $f_q(q|\alpha_q)$ denote the marginal PDF of individual qualities, with CDF $F_q(q|\alpha_q)$ and parameters α_q . We assume q follows a standard uniform distribution, noting that any other distribution could be derived via a probability integral transform. A Gaussian copula with correlation ρ allows to establish the joint distribution $f(\Delta, q, |\alpha_{\Delta}, \alpha_q, \rho)$ while preserving the specified marginal distributions (see details in Appendix A.2), from which we obtain the conditional distribution of development increments given quality, $f_{\Delta|q}(\Delta|\alpha_{\Delta}, q, \rho)$, and its corresponding CDF, $F_{\Delta|q}(\Delta|\alpha_{\Delta}, q, \rho)$. The later provides the matrix elements p_l given q :

$$p_l|q = F_{\Delta|q}(\frac{l+1}{r+1}|\alpha_{\Delta}, q, \rho) - F_{\Delta|q}(\frac{l}{r+1}|\alpha_{\Delta}, q, \rho). \quad (2.9)$$

Since stage-durations of individuals are correlated via q , unique IPM-kernels are required for each individual at each stage.

2.3.2 Generating *in silico* survival data from the quality-dependent IPLM model

In silico survival data for two successive stages was generated as follows. Sample size was fixed as $N = 50$ individuals and the duration of the maturation experiment was set to $t_c = 21$ time steps, whereafter all data were right censored. Qualities $\{q_i\}_{i=1}^{i=N}$, correlation coefficient ρ and initial parameters $\{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2\}$ were initialised with draws from a standard uniform distribution, while model resolutions r_1 and r_2 , with draws from $\text{Uniform}(0, R_{\text{Max}})$, with $R_{\text{Max}} = 100$. Parameters for stage 1, $\{\mu_1, \kappa_1, \nu_1, r_1\}$, and stage 2, $\{\mu_2, \kappa_2, \nu_2, r_2\}$, were obtained using a rejection sampler (algorithm 2.3.1) that ran until the following constraints were satisfied:

1) all parameters α_1 and α_2 (of the alternative parameterisation $\alpha_1 = \mu \frac{(1-\kappa)}{\kappa}$ and $\alpha_2 = (1 - \mu) \frac{(1-\kappa)}{\kappa}$) for the development rate (beta) distributions f_{Δ} of each stage were greater than one, and

Algorithm 2.3.1: DATA SIMULATION WITH CONSTRAINTS (inputs= $N, N_{d_2}^{\min}$)

comment: Initialise all parameters.

$\{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, q_1, \dots, q_N, \rho\} \sim \text{Uniform}(0, 1)$

comment: Determine alternative parameters.

$\{\mu_1, \kappa_1\} \rightarrow \{\alpha_{11}, \alpha_{12}\}$

$\{\mu_2, \kappa_2\} \rightarrow \{\alpha_{21}, \alpha_{22}\}$

comment: Simulate cohort data $n_{c_1}, n_{c_2}, n_{m_1}, n_{m_2}, n_{d_2}$

while $n_{d_2} < N_{d_2}^{\min}$ **or** $\min\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}\} < 1$

***comment:** Use rejection until constraints are satisfied.

if $(n_{c_1} + n_{c_2} > n_{m_1} + n_{m_2})$ **or** $\min\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}\} < 1$

$\{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, \rho, q_1, \dots, q_N\} \sim \text{Uniform}(0, 1)$

do $\left\{ \begin{array}{l} \{\mu_1, \kappa_1\} \rightarrow \{\alpha_{11}, \alpha_{12}\} \\ \{\mu_2, \kappa_2\} \rightarrow \{\alpha_{21}, \alpha_{22}\} \\ \text{Simulate cohort data} \end{array} \right.$

else if $(n_{c_1} + n_{c_2}) \leq (n_{m_1} + n_{m_2})$

comment: Stepping-in avoids high rejection rates.

if $\nu_1 = \min\{\nu_1, \nu_2\}$

do $\nu_1 \sim \text{Uniform}(\nu_1, 1)$

else

do $\nu_2 \sim \text{Uniform}(\nu_2, 1)$

Simulate cohort data

do $\left\{ \begin{array}{l} \text{*comment: Use rejection until constraints are satisfied.} \\ \text{if } (n_{c_1} + n_{c_2} > n_{m_1} + n_{m_2}) \text{ or } \min\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}\} < 1 \\ \quad \{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, \rho, q_1, \dots, q_N\} \sim \text{Uniform}(0, 1) \\ \quad \text{do } \left\{ \begin{array}{l} \{\mu_1, \kappa_1\} \rightarrow \{\alpha_{11}, \alpha_{12}\} \\ \{\mu_2, \kappa_2\} \rightarrow \{\alpha_{21}, \alpha_{22}\} \\ \text{Simulate cohort data} \end{array} \right. \\ \text{else if } (n_{c_1} + n_{c_2}) \leq (n_{m_1} + n_{m_2}) \\ \quad \text{comment: Stepping-in avoids high rejection rates.} \\ \quad \text{if } \nu_1 = \min\{\nu_1, \nu_2\} \\ \quad \quad \text{do } \nu_1 \sim \text{Uniform}(\nu_1, 1) \\ \quad \quad \text{else} \\ \quad \quad \text{do } \nu_2 \sim \text{Uniform}(\nu_2, 1) \\ \quad \quad \text{Simulate cohort data} \end{array} \right.$

return $(n_{c_1}, n_{c_2}, n_{m_1}, n_{m_2}, n_{d_2}, \mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, q_1, \dots, q_N, \rho)$

2) daily survival parameters $\{\nu_1, \nu_2\}$ were sufficiently large that the number of individuals completing both stages (n_{d_2} , see below) was at least $N_{\min} = 35$.

After the simulation, the number of individuals that got censored (n_{c_1}, n_{c_2}), died (n_{m_1}, n_{m_2}) or developed (n_{d_1}, n_{d_2}) in each stage was recorded. If $n_{d_2} < N_{d_2}^{\min}$ and $n_{c_1} + n_{c_2} > n_{m_1} + n_{m_2}$, all parameters were resampled from their priors. Otherwise, if $n_{d_2} < N_{d_2}^{\min}$ and $n_{c_1} + n_{c_2} \leq n_{m_1} + n_{m_2}$, the lowest of the two survival probabilities was resampled from a uniform prior truncated at the current value of ν . The rejection sampler was stopped once $n_{d_2} \geq N_{d_2}^{\min}$ (see Algorithm 2.3.1).

The fate of each individual i in each stage was described by two pieces of information:

- 1) the time-to-event, $y_{iA} \in \{1, \dots, t_{\max}\}$; and
- 2) the event-type, $y_{iB} \in \{\text{stage completion, mortality, censored}\}$.

Probabilities associated with combinations of y_{iA} and y_{iB} were obtained by adapting the description given in Appendix A.3 to the two-stage quality-dependent case as follows. The probabilities $p_l|q_i$ permitted conditional construction of the IPM-approximation (2.5). The unit pulse vector $(\mathbf{n}^T, c_s)_0 = (1, 0, \dots, 0)^T$ was projected to give, for each time-step $t \in \{0, \dots, t_{\max}\}$, the probabilities to complete a stage, $p_d(t|q_i)$, or to die, $p_m(t|q_i)$. The probability of right-censor beyond t_{\max} is $p_{ci} = 1 - \sum_{t=1}^{t_{\max}} (p_d(t|q_i) + p_m(t|q_i))$. For stage 1, t_{\max} is just t_c , and death or right-censor in stage 1 happen (if it happens) at t_c , which imposes that $p_{ci} = 1$ for stage 2. Otherwise, for stage 2, t_{\max} is $t_{\max}^{(i)} = t_c - t_{s_2}^{(i)}$, where $t_{s_2}^{(i)}$ is the time-step at which individual i enters stage 2. The probabilities $p_d(t|q_i)$, $p_m(t|q_i)$ and p_{ci} define the right-censored sojourn-mortality time distribution for every of the N individuals in a given stage.

Individual-level data (i.e. time-to-event and event-type) can be obtained by sampling the categorical distribution,

$$(y_{iA}, y_{iB}) \sim \text{Categorical}(p_d(1|q_i), p_m(1|q_i), \dots, p_d(t_{\max}|q_i), p_m(t_{\max}|q_i), p_{ci}). \quad (2.10)$$

Probabilities $p_d(t|q_i)$ and $p_m(t|q_i)$ we obtained using the marginal $F_{\Delta|q}(\Delta|\boldsymbol{\alpha}_{\Delta}, q, \rho)$ to evaluate the expression 2.9. These probabilities then were used to calculate, for each stage, the "true" mean ($\tilde{\mu}$) and standard deviation ($\tilde{\sigma}$) of maturation-times and the probability to survive to maturation ($\tilde{\nu}$) (Appendix A.3).

2.3.3 Assessing estimation performance

To asses the identifiability of estimates from the quality-dependent IPLM model, 500 simulations were performed by applying the procedure described in section 2.3.2.

Estimation. We adopt a Bayesian approach to estimate stage-specific parameters $\{\mu, \kappa, \nu, r\}$. Throughout, we use standard uniform priors for $\{\mu, \kappa, \nu\}$ and the prior $\text{Uniform}(0, R_{\max})$ for r , where R_{\max} is a maximum resolution chosen to be large enough to optimise model fit but small enough to maintain computational efficiency. We use Markov chain Monte Carlo (MCMC) (Gelman *et al.*, 2003) to sample posteriors of the form

$$f(\mu, \kappa, \nu, r|\mathbf{y}) \propto f(\mu)f(\kappa)f(\nu)f(r)f(\mathbf{y}|\mu, \kappa, \nu, r), \quad (2.11)$$

where \mathbf{y} represents a set of independent data sets.

For each simulation, MCMC was used to approximate the posterior distribution

$$f(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\nu}, \mathbf{r}, \mathbf{q}, \rho | \mathbf{y}_{S_1}, \mathbf{y}_{S_2}) \propto f(\rho) f(\mathbf{q}) \prod_{s \in \{S_1, S_2\}} f(\mu_s) f(\kappa_s) f(\nu_s) f(r_s) \prod_{i=1}^N f(y_{is} | \mu_s, \kappa_s, \nu_s, r_s, \rho, q_i). \quad (2.12)$$

This was achieved using the default block Metropolis-Hastings sampler in NIMBLE (NIMBLE Development Team, 2016). Thinning was set to twice the minimum expected sample size (Plummer *et al.*, 2006) obtained from pre-runs (Appendix A.4). Thereafter, 10^4 thinned MCMC samples were generated and convergence diagnostics were performed using CODA (Plummer *et al.*, 2006).

2.3.4 Results

Posterior medians and 95% credibility intervals (CI_{95}) of the means ($\hat{\mu}$), standard deviations ($\hat{\sigma}$) and total survivals ($\hat{\nu}$) of the joint sojourn-mortality distribution were plotted against true values (Fig. 2.2). Medians were distributed evenly around the 1:1 line and uncertainty was sufficiently small to suggest that the sojourn-mortality distribution approximations were accurate given the sample size. Precision was greatest when $\tilde{\mu}$ and $\tilde{\sigma}$ were small. The CI_{95} s of estimated parameters enveloped true values in approximately 95% of simulations. Banding was evident in the posteriors for $\tilde{\nu}$; this arose from the limited set of possibilities regarding the number of individuals completing both stages.

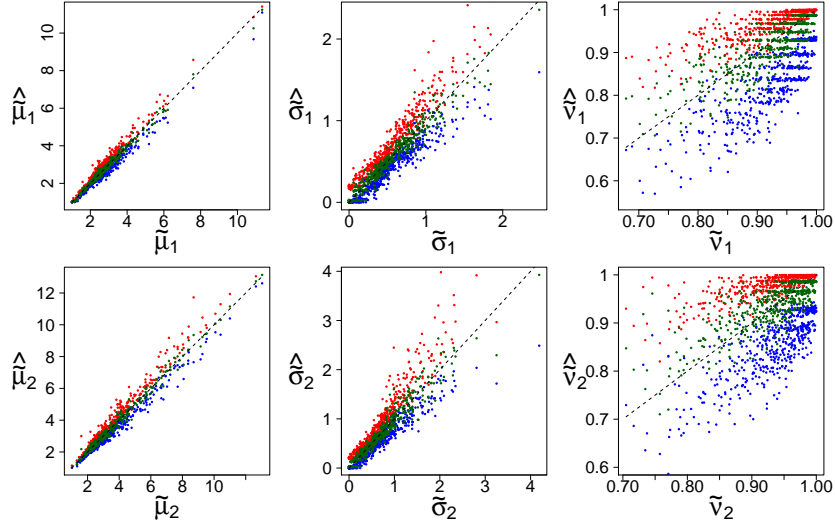


Figure 2.2. Estimated (y-axes) vs. true (x-axes) mean ($\tilde{\mu}$), standard deviation ($\tilde{\sigma}$) and total survival probability ($\tilde{\nu}$) of sojourn-mortality distributions generated from 500 simulations of $N = 50$ individuals with a two-stage quality-conditioned IPLM model. For each simulation, the median (green) and the upper (red) and lower (blue) bounds of the 95% credibility interval obtained from MCMC sampling of the posterior distribution are shown. The one-one (dashed) line is shown for reference. The number of outliers where CI_{95} s failed to envelope the true parameters were: 25 for $\tilde{\mu}_1$, 32 for $\tilde{\mu}_2$, 20 for $\tilde{\sigma}_1$, 29 for $\tilde{\sigma}_2$, 20 for $\tilde{\nu}_1$ and 28 for $\tilde{\nu}_2$.

In general, the CI_{95} s of estimated values for r_1 , r_2 , ρ and $\{q_1, \dots, q_N\}$ enveloped the true values (Figs. 2.3 and 2.4). Uncertainty was larger for these parameters than for $\tilde{\mu}$, $\tilde{\sigma}$ and $\tilde{\nu}$. The largest CI_{95} s for qualities q_i were associated with individuals that died in stage 1, and the greatest precision was achieved when sojourn-times were right censored (Fig. 2.4). For fully developed individuals, quality estimates ranged greatly in precision. True vs. fitted values of ρ showed that, despite uncertainties in the q_i , relatively accurate estimates for ρ were obtained (Fig. 2.3). The posterior median and CI_{95} s for resolution were clustered in horizontal bands suggesting that model fit was not sensitive to resolution so long as resolution was not too small (Fig. 2.3). This implies that very large values of r can be computationally superfluous since even relatively low resolutions can yield sojourn-mortality distributions as accurate as can be supported by the data.

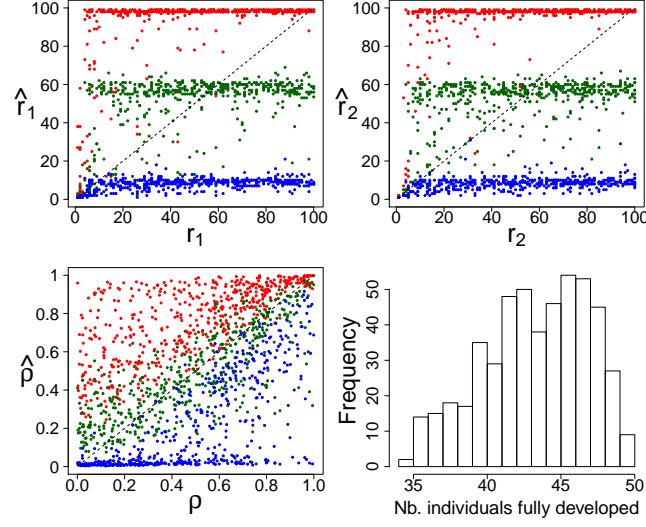


Figure 2.3. Estimated (y-axes) vs. true (x-axes) values for resolutions r_1 , r_2 and correlation ρ parameters, and frequency distribution of individuals completing both stages (bottom right), from 500 simulations with a two-stage quality-conditioned IPLM model with a sample size of $N = 50$. For each simulation, the median (green) and the upper (red) and lower (blue) bounds of the 95% credibility interval obtained from MCMC sampling of the posterior distribution are shown. The one-one line (dashed) is shown for reference.

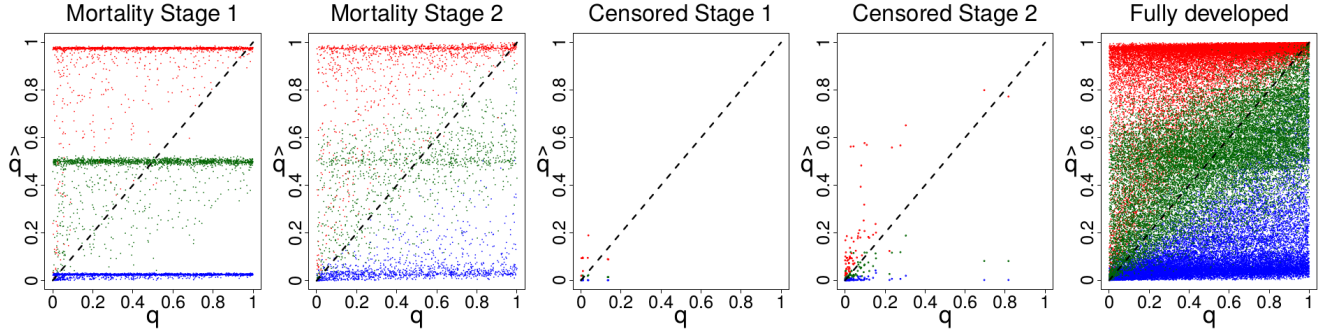


Figure 2.4. Estimated individual qualities \hat{q} versus their corresponding true value q , from 500 simulations of a two-stage quality-conditioned IPLM model with a sample size of $N = 50$. For each simulation, the median (green) and the upper (red) and lower (blue) bounds of the 95% credibility interval obtained from MCMC sampling of the posterior distribution are shown for five classes of individuals: died in stage 1; died in stage 2; censored in stage 1; censored in stage 2; and completed both stages. The one-one line (dashed) is shown for reference.

These results highlight that quality-conditioned IPLMs can successfully model development and survival in marked cohort studies. A two-stage study with $N = 50$ was used here to indicate what can be possible with a typical data set from a small experiment. Naturally, a greater number of individuals or stages would increase precision – an important consideration regarding the design of experiments to parameterise eco-evolutionary models. Most importantly, we show that even when within-stage development is unmeasurable, realistic IPM-

based matrix models can be fit to maturation-time data.

2.4 Discussion

The lack of a flexible treatment of stage-duration is a recognised limitation of many stage-structured (including Lefkovitch) models, which in turn affects dynamics and associated demographic indices (e.g. common short-term and long-term dynamics indicators) (De Valpine *et al.*, 2014). Here we have introduced the integral projection Lefkovitch matrices (IPLMs), a new tool for modelling the dynamics of stage-structured populations that, by including within-stage dynamics via integral projection models (IPMs), overcomes this limitation. By tracking within-stage development, the IPLM greatly reduces stage-duration errors and can provide valid transition probabilities when stage distributions are non-stable. In spite of the technicalities underlying IPLM models, the outputs draw on and rest on known methods of classical matrix models and algebra. This, coupled with Bayesian inference, makes implementation and analysis straightforward and particularly suitable for empirical ecologists seeking to increase realism of their models, or theoretical ecologists seeking to challenge process models with data.

The kernel of any IPM-based model must synthesise the net effects of interacting endogenous and exogenous processes on vital rate heterogeneity. We have shown that IPLMs enable realistic incorporation of unmeasured local or genetic factors represented by a *quality* variable. In a similar way, time-varying covariates could be incorporated to describe exogenous forcing on development (see chapter 3). While the popularity of IPMs has relied, to a large extent, on the strength and relative ease of using regressions models to parameterise mechanistic models, we have shown that IPM-kernels can also be estimated via the survival analysis of maturation-time data. Moreover, by using maturation-time (and not size) data, our methods can greatly improve the precision of stage-structured IPMs whenever size is a poor, or unavailable, predictor of stage duration. This scenario is ubiquitous in ecology: egg and exoskeleton dimensions often remain relatively constant, and more appropriate developmental metrics can be too expensive or difficult to collect routinely. Because many studies of marked individuals provide not only information of mean stage-duration but also on its variation, we expect many ecological studies will be able to parameterise realistic variations of the model described here in diverse applications.

Variation in traits characterising development in populations can correlate sojourn-times, affecting growth rates and other demographic indices. Trait heterogeneity is fundamental to eco-evolutionary models, and static

traits, such as *quality*, are often used to condition vital rates (Vindenes and Langangen, 2015). Yet, ecologists lack tools for tracking many important traits in natural populations (Caswell, 1983; De Valpine *et al.*, 2014). We have shown how, using Gaussian copulas, unmeasured individual traits can condition IPLM kernels to model correlated stage-durations. Although we call these individual traits *quality*, these parameters can be used and interpreted in various ways: in spatial analyses, *quality* could be used as a random effect conditioning phenotypic responses of sub-populations to unmeasured local factors; similarly, *quality* could provide a group-level random effect when modelling laboratory data; in eco-evolutionary models, *quality* can provide a synthetic index of genotypic traits that affect vital rates and fitness.

We have tested, with a simulation case study, the estimation performance of our methods, showing their ability to extract valuable information from relatively small sample sizes typical from developmental studies on emergence or maturation. Moreover, our simulation results show that even when within-stage development is unmeasurable, quality-dependent IPLMs can be used to incorporate unmeasured local or genetic factors into population models. While correlations in development violate the Markov process assumption of matrix model projections, IPLMs, given their nature, overcome this serious limitation, and can thus reduce projection error in studies, such as eco-evolutionary studies, where variation and developmental correlation are key.

The general IPLM described in this chapter could be adapted to many arthropods (e.g. mosquitoes, biting midges, tse-tse flies, etc.) as well as to other biological stage-structured population, such as fish or within-host parasites. For states other than development that structure populations and influence vital rates (e.g. population attributes such as biomarkers, health states, physiology, morphology, behavior or location, among others (Caswell, 2006)), the IPLM can represent an adequate framework to study key biological features of demography, although for certain cases (such as populations where multiple transitions or shrinking are to be considered) some effort would be required to include these complexities. Because many studies of marked individuals provide not only information on mean stage-duration but also on its variation, we expect many ecological studies will be able to parameterise realistic variations of the model described here in diverse applications.

2.5 Conclusion

In this chapter, a new matrix model for stage-structured populations (SSPs), the “integral projection Lefkovitch matrix” (IPLM), has been presented. The IPLM augments a classic Lefkovitch matrix (CLM) with stage- specific

integral projection models (IPMs) that track within-stage dynamics, in order to improve the characterisation of variation in development times by tracking individual histories within each stage. The resulting model drastically reduces stage-duration errors; is robust to stage distribution instabilities arising from perturbations; is flexible enough to permit parsimonious parameterisation while augmenting significantly the ability to incorporate diverse sources of variation (such as random variables); and can be fitted, even when within-stage development is unmeasurable, using developmental cohort data, in a way that gives access to much of the powerful matrix model toolbox.

Due to this increased resolution, it is expected that IPLM methods give more accurate estimates and projections of populations where size is not a good predictor of stage-duration but maturation-time data – even from small samples – is available or easily obtainable (e.g. egg stage in many organisms or exoskeleton dimensions in Ecdysozoa). Thus, the IPLM approach is expected to improve the forecast horizon of stage-structured models in many branches of ecology and evolution. To further investigate this conjecture, we study in the next chapter a temperature-dependent IPLM that is fitted to laboratory data of our motivating organism, *Culicoides* biting midges.

Chapter 3

An IPLM-based study of seasonal and transient dynamics of *Culicoides* biting midges

In this chapter ¹ we explore the impact of temperature changes on the life-history response of a model organism, biting midges of the genus *Culicoides*, when within-stage variation is taken into account. For this, we extend the IPLM model developed in chapter 2 – which provides valid transition probabilities for populations with non-stable stage distributions (as it is often the case in natural populations) – to include temperature dependence. We extend the Bayesian inference framework and show how laboratory data on *Culicoides* development at different temperatures can be used to fit an IPLM model for studying population dynamics that demonstrates more realistic dynamics than when developmental heterogeneity is neglected.

3.1 Introduction

Changes in environmental conditions translate into changes in demographic rates, which in turn affect the dynamics of a population (Coulson *et al.*, 2011). Temperature is recognised as a main – if not the most important – driver of development in poikilothermic species (i.e. those species whose body temperature follows that of their environment) of plants and invertebrate animals (Cloudsley-Thompson, 1962; Easterling *et al.*, 2000a). Among them, many poikilothermic Ecdysozoa (a large and diverse clade of animals that shed exoskeletons including

¹Most of the work presented in this chapter is part of an article submitted to the journal *Methods in Ecology and Evolution* (see Appendix D).

Arthropoda, Nematoda, and six smaller phyla, totalling over an estimated 4-5 million of species (Aguinaldo *et al.*, 1997; Telford *et al.*, 2008)) are of economical importance, for different reasons ranging from harvesting for human consumption to them being agricultural pests or vectors of diseases. Moreover, many of these species are of importance regarding conservation, and improved dynamic models can help guide conservation strategy (McElderry, in press) Examples of them are mosquito, nematodes, crustacean, midges and ticks.

Current methods in ecology fail to scale up realistically from laboratory development studies to field predictions of population dynamics. This is mainly because they do not provide valid transition probabilities for populations with vital rates sensitive to exogenous sources of variation. For example, arthropods, mites or nematodes exposed to varying temperatures are rarely modelled using Lefkovitch matrices because these models do not track stages accurately under time-varying vital rates. Furthermore, while IPM is regularly used to model the dynamics of wild vertebrate or plant populations, it is rarely used to model Ecdysozoan populations subject to time-varying parameters. Indeed, for many Ecdysozoa, it can be prohibitively difficult to obtain appropriate within-stage development data to fit IPMs with time-lagged regression.

Classically, the dynamics of poikilotherms are modelled using degree-day accumulation (DDA), a physiological unit capturing cumulative metabolic responses to temperature (De Reaumur, 1735; Belehradek, 1935; Russelle *et al.*, 1984). Maximum likelihood estimators are available for stochastic DDA models (Osawa *et al.*, 1983; Kemp *et al.*, 1986; Dennis *et al.*, 1986). However, these models neglect mortality, do not yield stage-specific parameterisation, require developmental homogeneity at time zero, use non-monotonic DDA and, as for most DDA models, assume a linear temperature–development relationship. Although linearity works over small temperature ranges, non-linearity becomes important when temperature fluctuations gain amplitude (McMaster and Wilhelm, 1997; Bonhomme, 2000).

Many other approaches with different levels of complexity use DD to describe developmental responses to temperature. The simplest approaches consist of purely shape descriptive functions with different numbers of parameters, such as generalised linear models (Manel and Debouzie, 1997), semi Markov processes (Munholland and Kalbfleisch, 1991), exponential expansion (Logan *et al.*, 1976), circular statistics (Morellato *et al.*, 2010), non-linear regression models (Briere *et al.*, 1999) and generalised additive models (Hodgson *et al.*, 2011). More complex non-linear degree-day models have been proposed based on temperature reaction curves of enzymatic activity (Sharpe and DeMichele, 1977; Schoolfield *et al.*, 1981). This approach has been used

to include developmental variation and time-varying environments by parameterising individual-based models (IBMs) (Régnière and Powell, 2013). This framework emphasises the fitting of non-linear expected response curves and treats variance as a nuisance parameter. Often, proportionality between SDD mean and standard deviation is assumed (Sharpe and DeMichele, 1977), and the covariates or stochastic processes that generate variance are neglected. Moreover, proponents neglect that mortality modifies SDDs, and either estimate survival by neglecting SDD shape and variance (Régnière *et al.*, 2012) or neglect mortality entirely (Yurk and Powell, 2010). In addition, computational costs can prevent IBMs from scaling well and more general solutions are required.

With IPLM models, SDD variance arises naturally from a stochastic development-mortality process. Furthermore, the assumptions used for estimation and simulation are identical, thereby eliminating potential bias arising from model mismatch. It is worth noting that despite similarity with DDA models (Plant and Wilson, 1986; Dennis *et al.*, 1986; Régnière *et al.*, 2012), the powerful potential of IPMs remains largely untapped in poikilotherm studies.

Here, we fit a temperature-dependent IPLM to unmarked maturation-time data for biting midges of the genus *Culicoides* at fixed temperatures. We model IPM-kernel parameters as a function of temperature using non-parametric regression – the model is fit using biologically justified unimodal constraints, and unimodal spline interpolation determines parameters at unmeasured temperatures. The interpolated model is used to analyse asymptotic and transient dynamics under fixed and seasonal temperatures.

3.2 *Culicoides* IPLM model

Culicoides biting midges attract considerable interest as vectors of numerous viral diseases (Tabachnick, 1996; Mellor *et al.*, 2000; Guis *et al.*, 2012; Mardulyn *et al.*, 2013; Guichard *et al.*, 2014). Modelling has provided empirical descriptions of flight-trap data for phenology, bio-geography or epidemiological risk studies (Purse *et al.*, 2004; Sanders *et al.*, 2011; Searle *et al.*, 2012; Carpenter *et al.*, 2013; Hartemink *et al.*, 2015; Diarra *et al.*, 2015). But these approaches cannot provide all the vital rates required for incorporating vector life-cycle dynamics in mechanistic epidemiological models. While insufficient *Culicoides* within-stage trait data (i.e. size, weight) exists for time-lagged regression, sufficient maturation-time data exist for fitting a temperature-dependent IPLM.

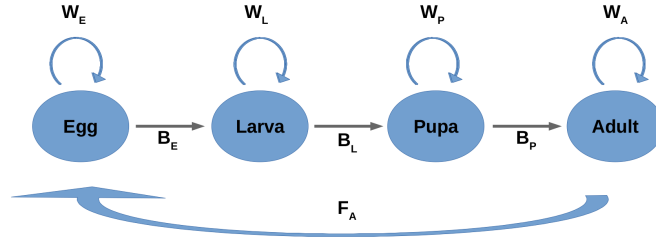


Figure 3.1. Schematic representation of a simplified *Culicoides* life cycle that provides a basis for developing an IPLM. At each time step: the values P_E , P_L , P_P and P_A represent the proportion of individuals remaining in a given stage-class; the values T_E , T_L and T_P represent the proportion of individuals passing to the next stage-class; and F_A represents the per capita fecundity.

Figure 3.1 schematises a typical *Culicoides* life cycle with nodes representing stage-classes and arrows denoting transition between stages. This scheme provides the female only egg-larva-pupa-adult (ELPA) IPLM (adapted from 2.6) that can be written as

$$\begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_t = \begin{bmatrix} \mathbf{W}_E & 0 & 0 & \mathbf{F}_A \\ \mathbf{B}_E & \mathbf{W}_L & 0 & 0 \\ 0 & \mathbf{B}_L & \mathbf{W}_P & 0 \\ 0 & 0 & \mathbf{B}_P & \mathbf{G}_A \end{bmatrix}_t \begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_{t-1}, \quad (3.1)$$

where $\mathbf{G}_A = \mathbf{W}_A + \mathbf{B}_A$ models multiple gonotrophic cycles. Although *Culicoides* develop through five larval instars, stage identification is labourious and studies typically only report sojourn time data for the ensemble of larval substages. However, this loss of resolution is probably negligible since our methods can provide sojourn time distribution estimates for the ensemble of larval stages.

For species with one reproductive stage that undergo multiple reproductive cycles (e.g. mosquitoes, biting midges, tse-tse flies), \mathbf{W}_A can be replaced by $\mathbf{G}_A = \mathbf{W}_A + \mathbf{B}_A$ where \mathbf{B}_A provides transitions between cycles. This gives a parsimonious approach that assumes vital rates do not differ significantly between reproductive cycles. Alternatively, \mathbf{W}_A could be constructed as a series of multiple sub-matrices and \mathbf{F}_A adjusted accordingly.

We fitted this model to cohort data from *C. variipennis* egg, pupae and combined larvae-pupae development studies (Mullens and Rutz, 1983; Vaughan and Turner, 1987), and to individual-level data from *C. nubeculosus* fecundity, gonotrophic cycle and egg stage-duration studies (Balenghien *et al.*, 2016). Details of each data set are given in Appendix B.2. Note, these species share similar developmental responses across the 15°C-35°C range (Purse *et al.*, 2015).

Here we use the same notation for stage-specific IPM-kernels developed in chapter 2. For each stage, temperature-dependence was modelled using unimodal splines on survival $\nu(T)$, the 1st and 99th percentiles ($\mathcal{P}_1(T)$ and $\mathcal{P}_{99}(T)$) of the developmental rate distribution $f_{\Delta}(\Delta|\alpha)$, and (for adults) fecundity $F_A(T)$. For this, unimodality constraints on the responses to temperature of these parameters were incorporated into the MCMC. The unimodality constraints ensure an optimal temperature for each stage (Sharpe and DeMichele, 1977; Knies and Kingsolver, 2010; Régnière *et al.*, 2012) and permit shape-constrained interpolation at unsampled temperatures. Interpolation was performed, for each line of MCMC output, using unimodal cubic Hermite splines (Appendix B.7). This non-parameteric regression produces a smoothed unimodal curve analogous to (the piece-wise linear) multivariate adaptive regression splines (Friedman, 1991). Note, the spline modelling for Δ was performed on percentiles rather than μ and κ in order to enforce a unimodal response to temperature. Only two percentiles were needed to identify μ and κ and the 1st and 99th proved a practical choice.

Estimation. Details of likelihoods used for model fitting, including fecundity, are given in Appendices B.3 and B.3.1, while missing-value imputation steps are given in Appendix B.4. Posteriors were sampled using parallel tempering (Swendsen and Wang, 1986; Łacki and Miasojedow, 2015) (Appendix B.5), and performed in NIMBLE (NIMBLE and R scripts used in these analyses are available on github https://github.com/scastano/IPLM_code). Ten thousand thinned post-adaption MCMC samples were generated and convergence diagnostics were performed using CODA (details in Appendix B.6). Estimates of $\mu(T)$ and $\kappa(T)$ were obtained via back-transformation of the interpolated $\mathcal{P}_{01}(T)$ and $\mathcal{P}_{99S}(T)$ (Appendix B.7).

3.3 Results

Resolution.

Posterior likelihoods were consistently poor at $r_s = 1$, whereafter small resolution increases provided large likelihood gains, rendering the CLM's posterior probability negligible (Fig. 3.2). Maximum *a posteriori* (MAP) estimates for resolution were $r_E^{(\text{MAP})} = 6$, $r_L^{(\text{MAP})} = 9$, $r_P^{(\text{MAP})} = 31$ and $r_A^{(\text{MAP})} = 8$, with variable levels of uncertainty. The reason for these large likelihood gains is clear when comparing the model fit to maturation time data: the CLM's geometric sojourn-time distributions do not fit the data well while IPLM improves fitted sojourn-time distributions in all cases (Figs. 3.3, 3.4).

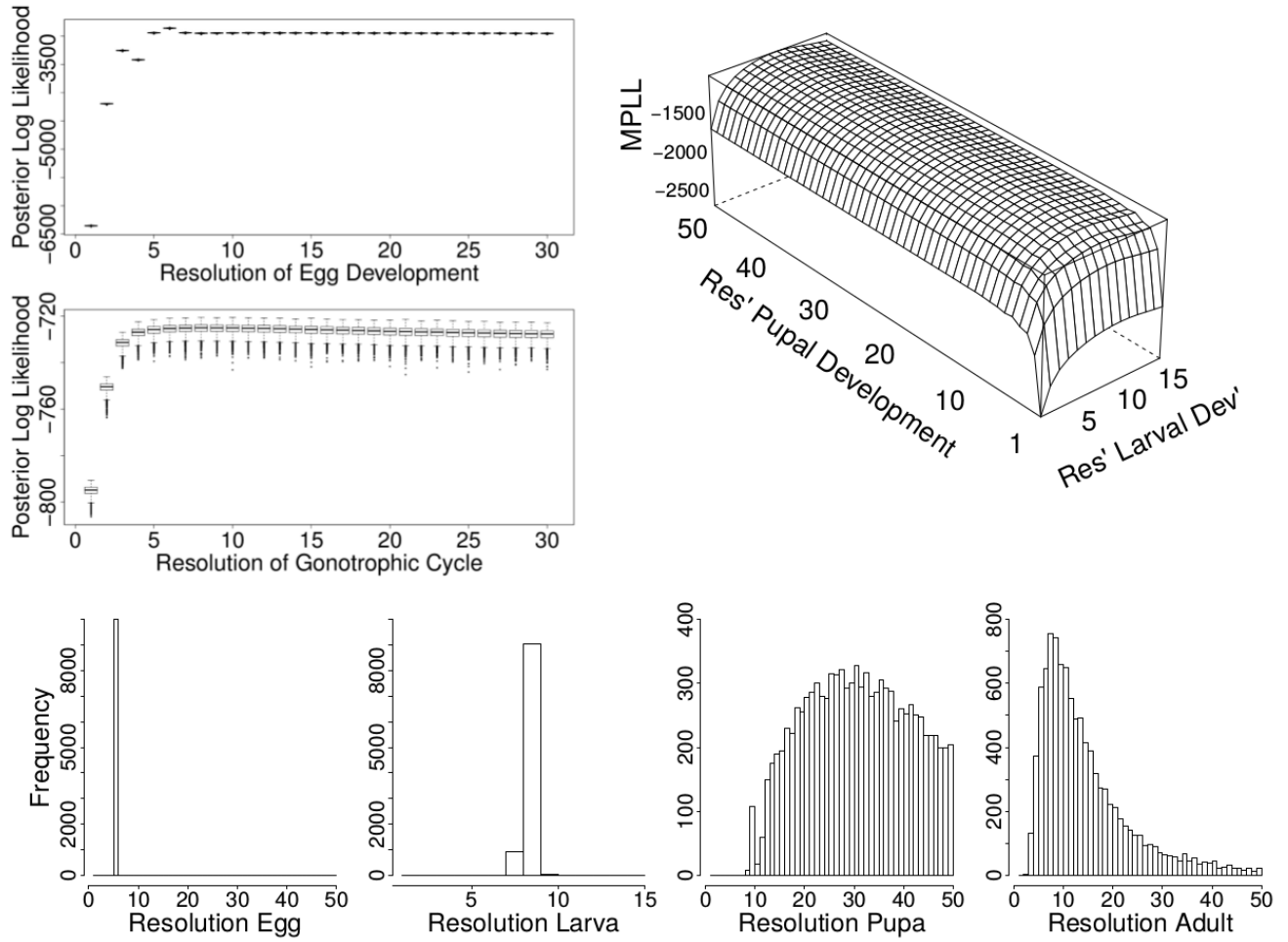


Figure 3.2. Posterior log likelihood profiles with respect to resolution (top) and the distribution of estimates for each resolution (bottom) for an IPLM model with egg, larvae, pupae and adult stages fitted to *Culicoides* biting midge data. Boxplots (top left) summarise the distribution of posterior log likelihoods and the wireplot (right) shows mean posterior log likelihoods (MPLL) calculated from 10^4 MCMC samples per resolution or combination. Maximum *a posteriori* (MAP) estimates were found at $r_E = 6$, $r_L = 9$, $r_P = 31$, $r_G = 8$ (subscript *G* indicates the gonotrophic cycle that followed by adult females) and posterior probabilities associated with the lowest resolutions were negligible.

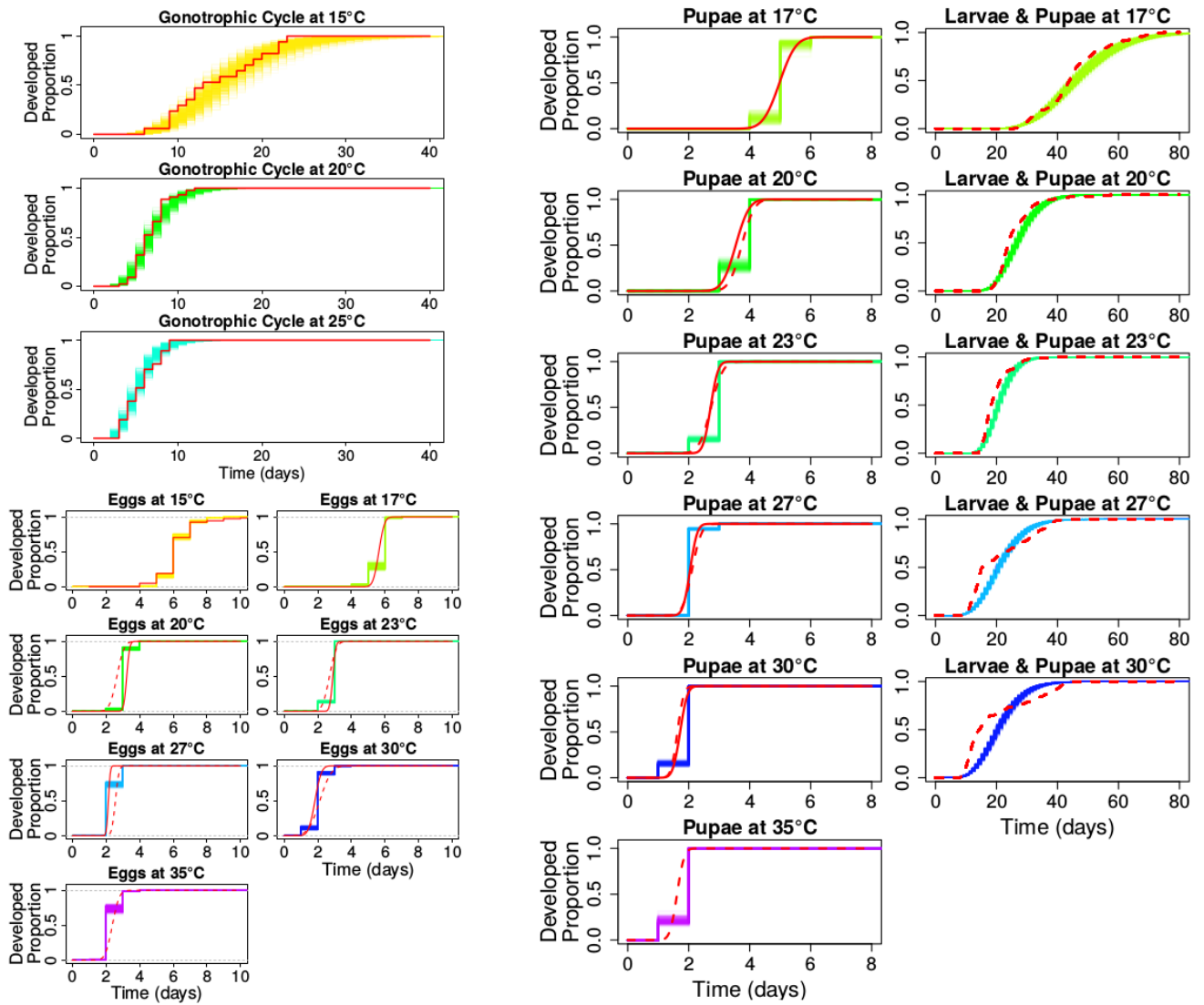


Figure 3.3. Posterior cumulative distributions of within-stage sojourn times of a temperature-dependent IPLM plotted with empirical *Culicoides* data (red lines). Where two data sets are shown, continuous red lines indicate Mullens *et al.* data, and dashed red lines indicate Vaughan *et al.* data. The fitted sojourn time distributions contrast markedly to those obtained with CLM (see Fig. 3.4).

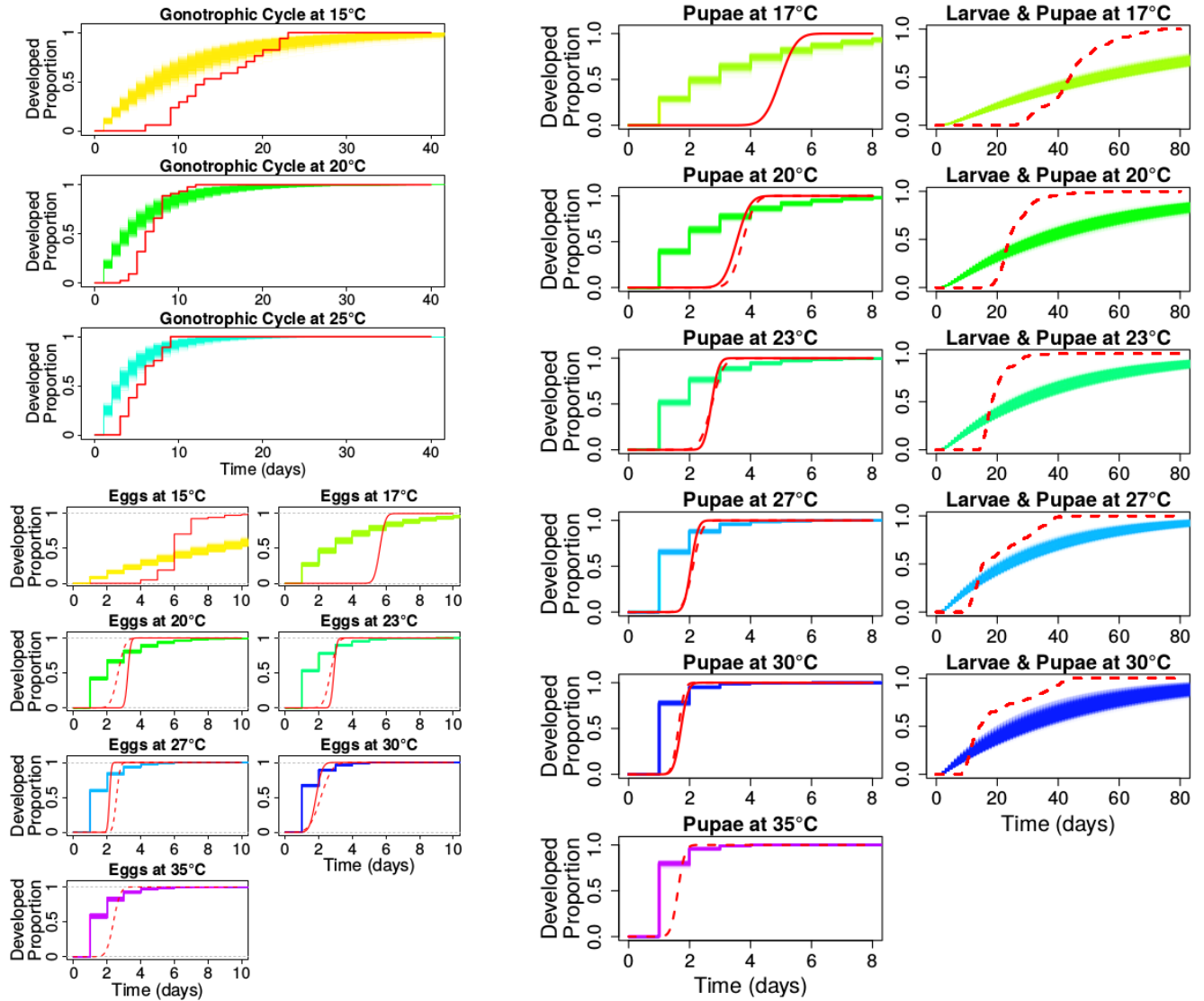


Figure 3.4. Posterior cumulative distributions of within-stage sojourn times of a temperature-dependent CLM plotted with empirical *Culicoides* data (red lines). Where two data sets are shown, continuous red lines indicate Mullens *et al.* data, and dashed red lines indicate Vaughan *et al.* data. The fitted sojourn time distributions contrast markedly to those obtained with IPLM (see Fig.3.3).

Differential Responses to Temperature.

Stages differed in their developmental responses to temperature (Figs. 3.5, 3.6). Generally, the mean and variance of developmental rates increased with temperature. However, eggs, and to a lesser extent larvae and pupae, exhibited impaired development. Larvae showed a more flat developmental rate, with values substantially lower than the other stages at most of temperatures considered.

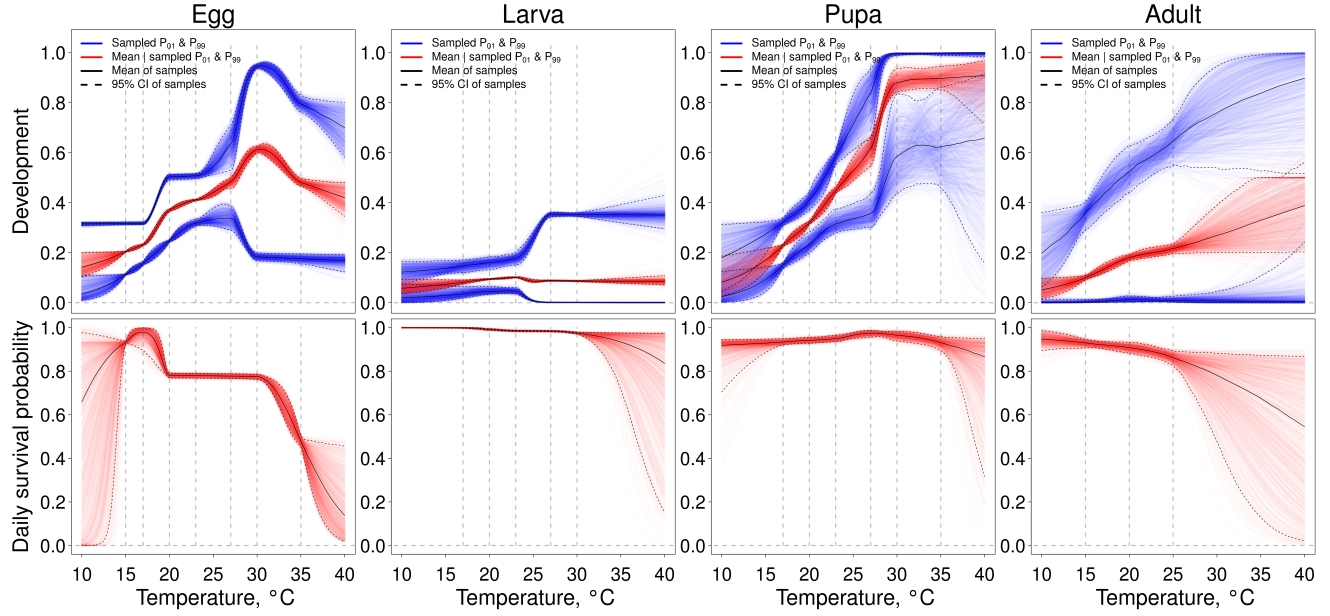


Figure 3.5. Non-linear responses to temperature of development (top) and survival (bottom) at most likely resolutions ($r_E^{MAP} = 6$, $r_L^{MAP} = 9$, $r_P^{MAP} = 31$, $r_G^{MAP} = 8$) for egg, larvae, pupae and adult midges. Results from 1000 MCMC samples are plotted with unimodal spline interpolation. Experimental temperatures are indicated with dashed vertical grey lines. Red and blue lines show median and 1% tail percentiles (P_{01} , P_{99}) of development kernel $f(\Delta|\theta)$ (top row). Expected values (black line) and 95% credibility intervals for each parameter are shown (dashed lines).

Survival was low at the highest temperatures for eggs, pupae and adults. Larvae experienced relatively high survival at all temperatures and were the most resistant stage to cold – this concurs with field reports of overwintering success being greatest for larvae (Kettle, 1962; De Liberato *et al.*, 2003). Despite these differential temperature effects, the stable stage distribution (dominant eigenvector) did not exhibit clear visual evidence of strong temperature-dependence within the range of experimental temperatures (Fig. 3.7).

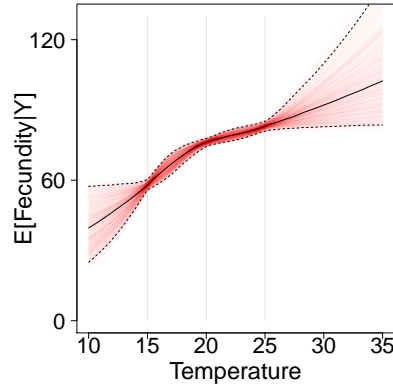


Figure 3.6. Expected fecundity from a Poisson-Jeffreys model fitted to *Culicoides* oviposition data collected at three temperatures (vertical grey lines). Posterior estimates from 1000 MCMC samples are shown with extrapolation over the range $10^{\circ} - 35^{\circ}\text{C}$. The data suggest a non-linear response of fecundity to temperature.

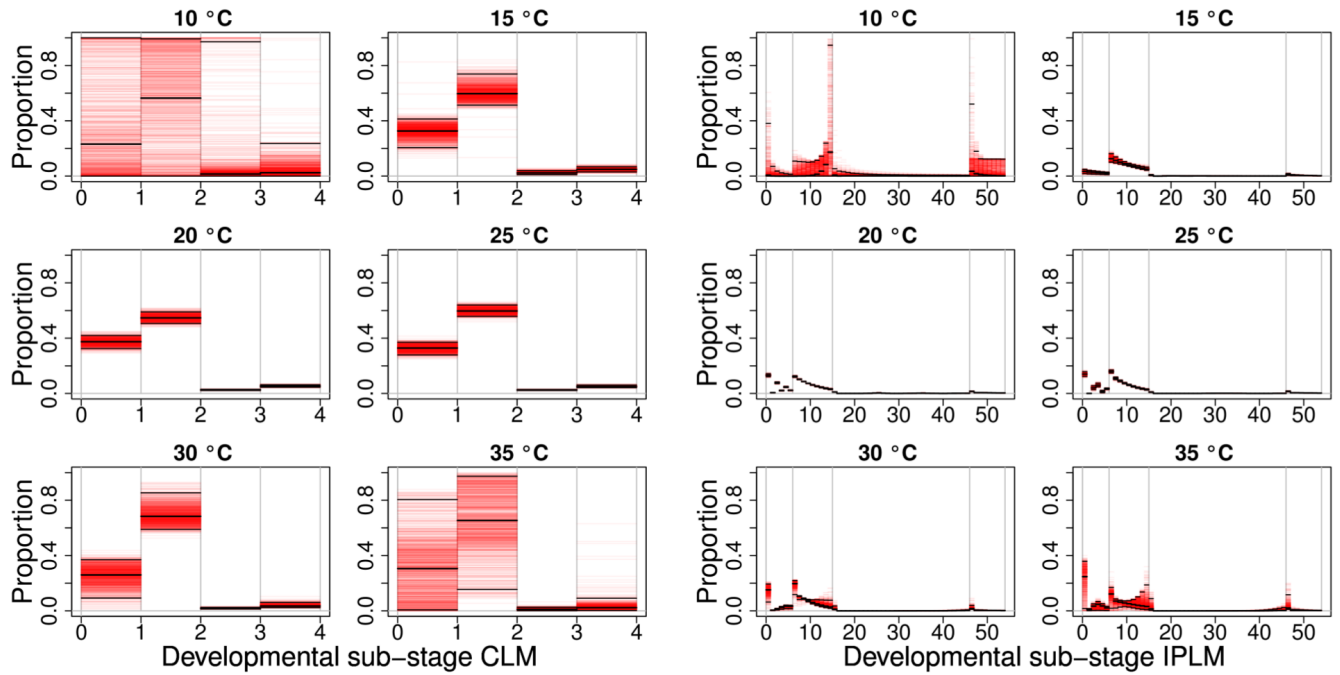


Figure 3.7. Stable stage distributions at several temperatures for *Culicoides* CLM (left) and IPLM (right) models. Results from 1000 MCMC samples (red lines) with posterior means and 95% CIs (black lines) are shown. For the IPLM model, resolutions were set to their maximum *a posteriori* (MAP) estimates: $r_E^{MAP} = 6$, $r_L^{MAP} = 9$, $r_P^{MAP} = 31$ and $r_A^{MAP} = 8$. Vertical lines separate stages.

Asymptotic dynamics at fixed temperatures.

The asymptotic growth rate (dominant eigenvalue λ_1) over a 10°C - 40°C range was similar for CLM and IPLM: both suggested temperatures in the mid-twenties optimise growth, although CLM systematically predicted slightly higher growth rates than IPLM (Fig. 3.8). Both models predicted population decline ($\lambda_1 < 1$) at high

temperatures. Both the range of temperatures yielding $\lambda_1 > 1$, and uncertainties regarding growth–decline threshold temperatures, were greater for CLM than for IPLM.

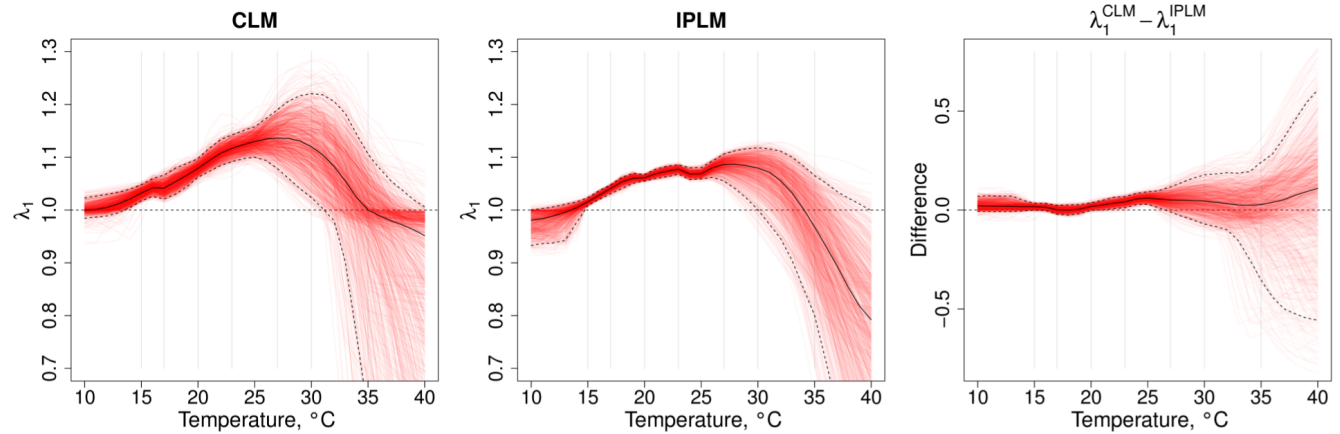


Figure 3.8. Temperature responses of asymptotic growth rate (dominant eigenvalue, λ_1) of CLM (left) and IPLM (center). Expected growth rates were higher for CLM than for IPLM over much of the temperature range, and the 95% CIs for this difference excluded zero over a range of approximately $22^\circ\text{C} - 26^\circ\text{C}$ (right). Both models predict population decline ($\lambda_1 < 1$) at higher temperatures. However, CLM predicted $\lambda_1 > 1$ over a greater range of temperatures than IPLM. Expected values (black line) with 95% CIs (dashed lines) from 1000 MCMC samples (red lines) are shown.

Transient dynamics.

To investigate potential effects of temperature perturbations, various indices of transient dynamics were quantified (Caswell, 2006; Koons *et al.*, 2007; Stott *et al.*, 2011). The duration of transient dynamics is largely determined by the *damping ratio* of first and second eigenvalues $\rho = \lambda_1/|\lambda_2|$, which represents the rate at which population approaches to the stable distribution. Plots of ρ^{-t} indicated slower convergence for IPLM than for CLM at all temperatures (Fig. 3.9). Thus, the relative importance of λ_2 increased when within-stage developmental heterogeneity was included, and CLM underestimated the duration of transients at every temperature.

The relative density ($\text{density}(t)/\lambda_1$) exhibited stronger evidence of transient oscillations for IPLM than for CLM. Maximum amplification (amp_{\max}) and maximum attenuation (att_{\max}) provide the largest and smallest possible values of relative density following perturbation at $t = 0$, while associated inertias provide the asymptotic values of amp_{\max} and att_{\max} . Each of these indices was affected when within-stage developmental heterogeneity was excluded, and CLM consistently underestimated the amplitude of transient oscillations (Fig. 3.9).

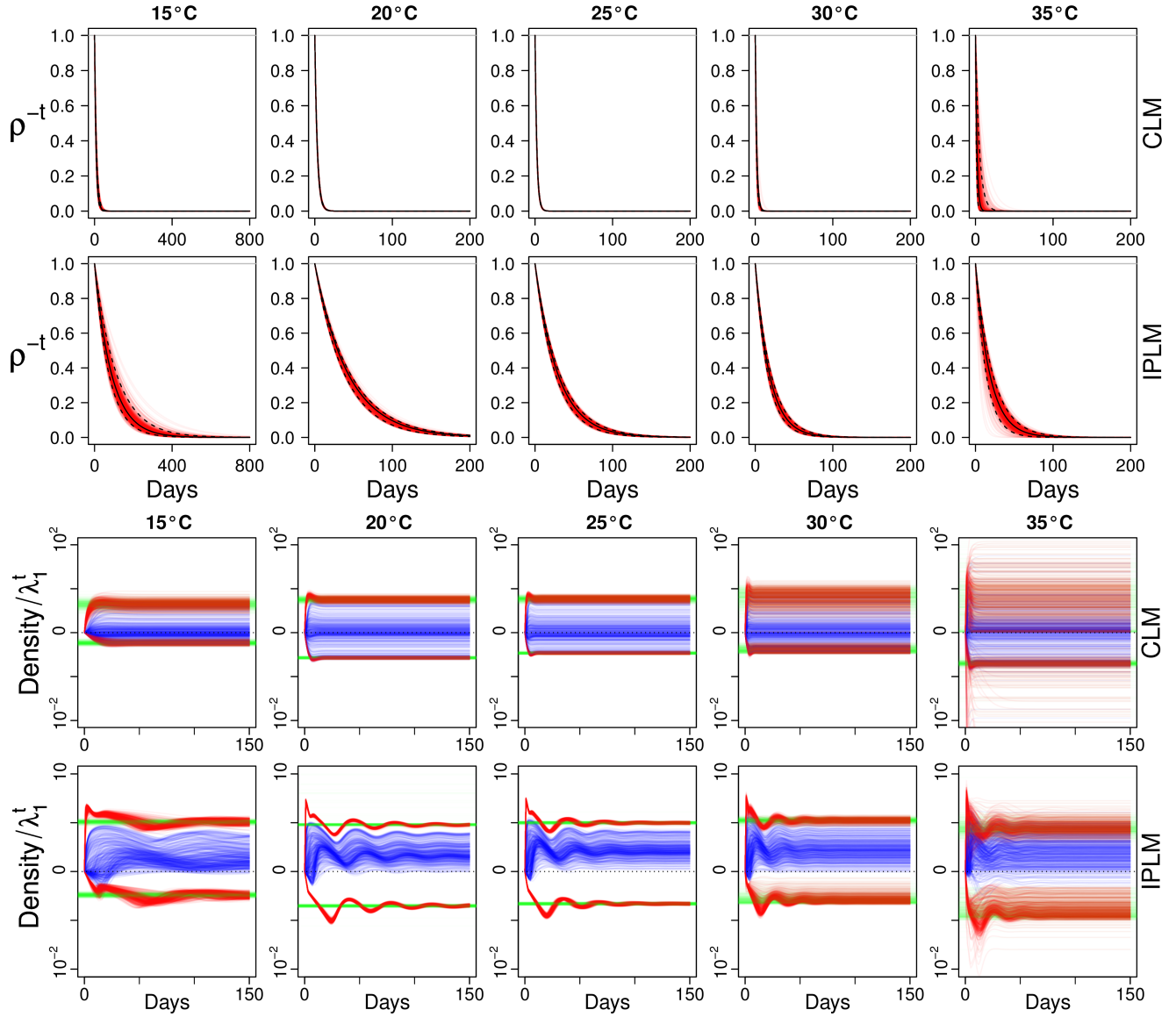


Figure 3.9. Indices of transient dynamics for *Culicoides* CLM and IPLM models at fixed temperatures. Geometric projections of inverted damping ratios ($\rho = \lambda_1/|\lambda_2|$) from 1000 MCMC samples (red lines), their means (black line) and 95% credibility intervals (dashed lines) show CLM consistently underestimated the potential duration of transient oscillations (first and second rows). These effects are clear in projected trajectories of relative densities (blue lines, third and forth rows), in all cases initial values were set to the 10°C stable stage distribution. Maximum amplification, maximum attenuation (upper and lower red lines respectively) and associated inertias (green lines) are shown.

Seasonal dynamics.

The effects of within-stage variation on seasonal dynamics were explored by plotting daily growth rates λ_t , projected relative densities (density divided by annual growth λ_Y), amp_{\max} and att_{\max} over two years for both the

CLM and IPLM. Two similar seasonal temperature profiles were modelled as

$$T_t = v + w \cos\left(t \frac{2\pi}{365}\right), \quad (3.2)$$

where v and w were set such that $\min(T_t) = \{15^\circ C\}$ and $\max(T_t) = \{25^\circ, 30^\circ C, \}$, with $t = 0$ the coldest day of the year. For projections of relative density, the initial population was set to the stable distribution associated with $15^\circ C$.

For both temperature regimes, the amplitude of annual oscillations in λ_1 and relative density were greater for CLM than IPLM (Fig. 3.10). Raising $\max(T_t)$ to 30° reduced both precision in λ_1 and the probability of $\lambda_1 > 1$ in mid-summer (Fig. 3.10). This switch from summertime growth to decline arose from high uncertainty in adult survival at $30^\circ C$. Trajectories of amp_{\max} and att_{\max} were more complex for IPLM in both seasonal temperature scenarios, and exhibited spring-time oscillations in the first year that became relatively damped by the second year. This damping suggests that, with the chosen temperature profile, spring-time flux in the stable stage distribution is mild.

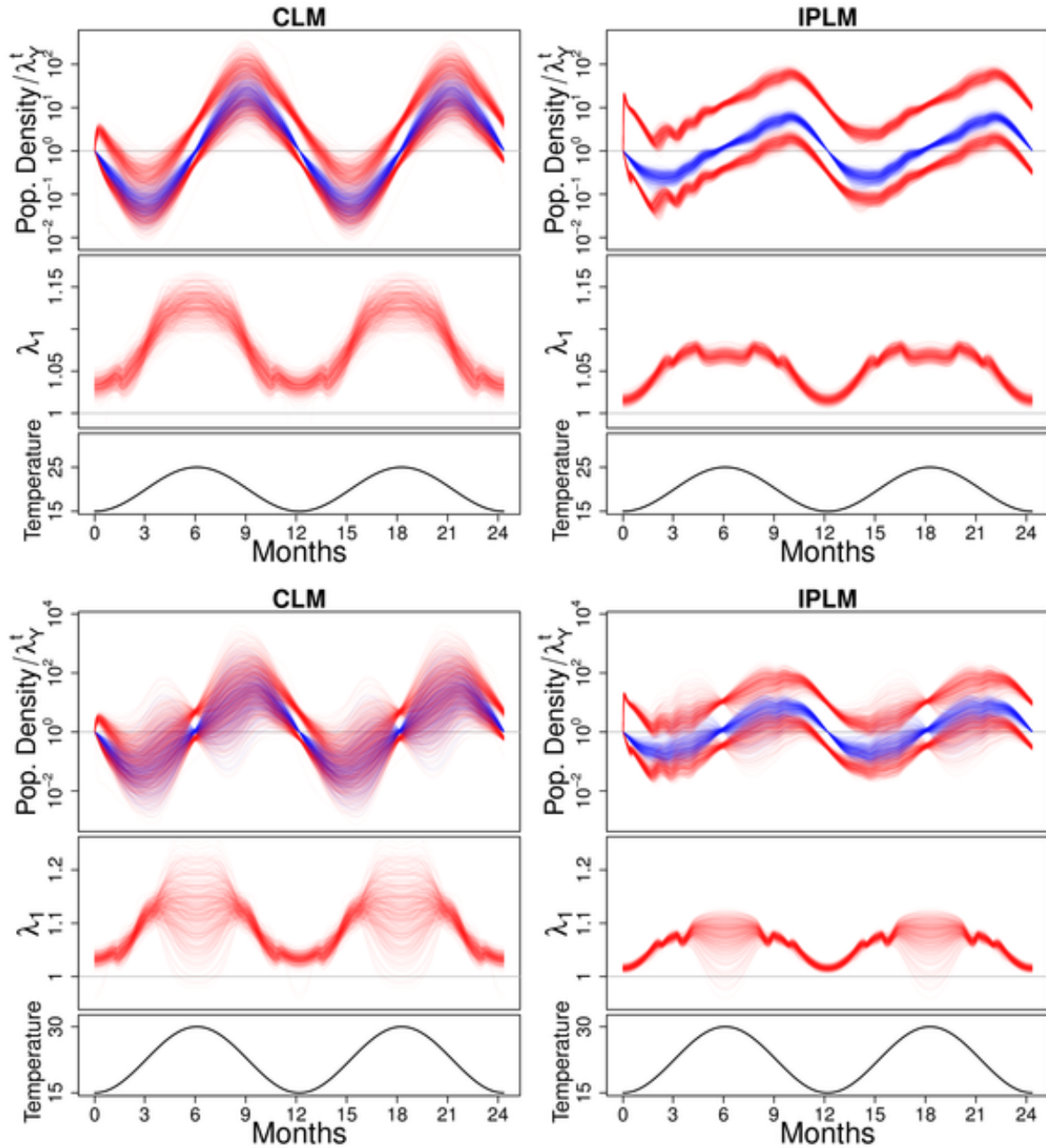


Figure 3.10. Growth rate (λ_1) and standardised density projections for *Culicoides* CLM (left) and IPLM (right) models forced with annual temperature fluctuations with ranges 15°C-25°C (top) and 15°C-30°C (bottom). The amplitude of λ_1 oscillations was greater for CLM than for IPLM. Uncertainty in λ_1 increased when mid-summer temperatures were increased from 25°C to 30°C. Maximum amplification and maximum attenuation trajectories (upper and lower red lines) were more complex for IPLM than CLM and exhibited clear spring-time oscillations characteristic of transient dynamics in the first year. Relative density projections were initialised at the 15°C stable stage distribution and standardised using annual growth λ_Y (blue lines). These trajectories only showed mild evidence of transient dynamics and few between-year differences indicating that these initial conditions generated negligible perturbation. Results from 1000 MCMC samples are shown.

3.4 Discussion

Here we have applied the integral projection Lefkovitch matrix (IPLM) framework to model the dynamics of *Culicoides* biting midges in order to provide an example of adapting IPLMs to a poikilothermic stage-structured population. We have shown that stage-specific vital rates, and related metrics, can be parameterised via non-linear responses to fixed or time-varying covariates to yield more realistic sojourn-mortality distributions. Moreover, by their structure, IPLMs can provide valid transition probabilities when stage distributions are non-stable, reduce errors in transient and/or non-linear dynamics and can therefore be expected to improve predictive performance when exogenous factors differentially affect vital rates.

In our *Culicoides* analyses, neglecting within-stage heterogeneity generated a small but systematic bias favouring over-estimation of growth rates (λ_1) and the probability of $\lambda_1 > 1$. Thus, neglecting within-stage heterogeneity can apparently affect predictions of potential ecological niche. It is increasingly recognised that perturbations and transient dynamics can be as important as asymptotic dynamics (Bierzychudek, 1999; Cushing *et al.*, 2002; Crone *et al.*, 2013). Neglecting within-stage heterogeneity led to underestimation of the duration and amplitude of transient oscillations, the potential range of relative densities (att_{\max} , amp_{\max}) and associated inertias. Whereas the 4×4 CLM yields just one pair of complex eigenvalues, the larger IPLM yields many more complex eigenvalues giving a richer characterisation of the transient oscillations that follow perturbation. The *Culicoides* analyses suggest that, for the chosen temperature profile – which possesses a smooth day-to-day change in temperature and mild extreme temperatures – , the importance of transient dynamics relative to asymptotic dynamics is small. Despite vital rates responding differentially to temperature, the perturbations generated by the chosen temperature profile only generate low-amplitude transient oscillations. However, the amplitude of transient oscillations is expected to increase as winter-summer, or day-to-day, temperature differences increase since cold winters exert strong differential mortality. Moreover, this unexpected result appears to arise as an artefact of using a linear (density independent) dynamic model. In chapter 4 we show that adding density dependence leads to an important role of transient dynamics that do not dampen with time throughout the season. These effects can have important consequences in wildlife management and other branches of ecology and evolution where perturbations limit the forecasting horizon of current methods. Temperature transfer experiments (Régnière *et al.*, 2012) are required to test this hypothesis.

As far as we know, our *Culicoides* analysis is the first time an IPM-based approach has been adapted to analyse

temperature effects on the within-stage development, transient dynamics and phenology of a poikilothermic arthropod. Although tracking within-stage development with temperature-dependent IPMs is analogous to tracking degree-day accumulation (Plant and Wilson, 1986), IPMs and degree-day accumulation models have hitherto evolved in relative isolation. The use of IPLMs in the *Culicoides* study bridges a historic gap between these schools of ecological modelling, overcomes many of the limitations of the pioneering work of Dennis *et al.* (1986), and avoids popular linearity assumptions – which are unrealistic over large temperature ranges. The IPLM framework readily accommodates stage-specific non-linear responses to temperature. We modelled these responses at unmeasured temperatures using spline interpolation subject to biologically justified unimodal constraints. Alternatively, mechanistic link functions based on metabolic theory (Régnière *et al.*, 2012) could have been used. This would have provided smoother development–temperature response curves, however, our non-parameteric approach provided greater parsimony in the *Culicoides* study. The model’s relative simplicity, and ability to exploit diverse data types, suggest that IPLM provides a valuable tool for modelling not only arthropods but many stage-structured populations.

An IPLM model is both an augmented matrix model and a discrete IPM approximation. Discretisation unavoidably introduces resolution parameters r with no biological meaning. The simulation results of chapter 2 suggested that, provided r is large enough, model fit can be relatively insensitive to r . By contrast, the *Culicoides* study shows that the degree of sensitivity of likelihoods to r is data dependent. Indeed, at low values, r functions as a shape parameter, suggesting that more flexible kernels should reduce sensitivity to r . Modelling development rate heterogeneity with beta distributions allowed us to demonstrate how the likelihoods of CLMs are greatly increased with just a few additional parameters. But we do not expect this distribution to be optimal in all situations, thus identifying and testing alternatives merits further research. Possible alternatives include probability distributions with more parameters and non-parameteric or semi-parameteric methods. It may even be possible to invert the estimation problem by fitting stage-duration distributions to data via classical survival analysis and solve to obtain a corresponding IPM-kernel. However, we do not know an analytic solution to this inverse problem, nor do we know of numerical solutions that could be more efficient than the projection approach adopted here.

Here we have focussed on temperature effects since temperature is the most fundamental covariate to affect poikilothermic dynamics. However, other covariates can be important and further research is required to generalise these methods for other taxa. For example, adult *Amblyoma* ticks exhibit a behavioural diapause in

their host-seeking behaviour during dry periods, an adaptative behaviour that minimises egg and larva mortality (Pegram *et al.*, 1988; Labruna *et al.*, 2003). Similarly, the eggs of *Aedes albopictus* and *Aedes aegyptis* are known to be highly resistant to dessication and can survive dry conditions for over one year (Bentley and Day, 1989; Bonizzoni *et al.*, 2013). Clearly, for approximating such effects, extra covariates are required in the definition of the IPLM kernel. But since IPM kernel definitions can be highly flexible, the approach should easily be adapted to these scenarios.

Other alternative data sources might be incorporated into model fitting. For example, time series from field studies could be used to reduce parameter uncertainty beyond the range of experimental conditions. A related problem to addressing this improvement of IPLMs is how to upscale from laboratory studies (with single generations) to field scenarios (with overlapping generations), and how to fit IPLMs to time series which typically just provide partial observations of incomplete life cycles. A Bayesian approach where the analysis of laboratory data affords informative priors for modelling field data would provide ecologists a very powerful tool to improve the forecast horizon of matrix models. In chapter 4, a preliminary study towards the development of generic methods for fitting IPLM models to such time series is presented.

3.5 Conclusion

In this chapter we have extended IPLM methods to include non-linear temperature-dependence in the developmental times and mortality rates of a stage-structured population. Fitting the model to *Culicoides* (biting midge) unmarked laboratory cohort data enabled an assessment of the relative roles of transient and asymptotic dynamics in constant and seasonal climates by comparing outputs of the IPLM model to the classic Lefkovich matrix (i.e. the matrix model that neglects within-stage heterogeneity). We have shown the traditional negligence of individual developmental heterogeneity affects asymptotic dynamic metrics in various ways and, moreover, greatly underestimates the importance (both amplitude and duration) of transient dynamics.

The IPLM framework permits the tracking of within-stage heterogeneity that arises from covariates without recourse to overparameterisation or unrealistic variance assumptions. Moreover, as shown here, much of the machinery of matrix model analysis is available including indices of asymptotic and transient dynamics. It is also perfectly feasible that sensitivities and elasticities of these indices to perturbation in parameters can be derived from the matrices using well known methods (Caswell, 2006), although we do not explore this possibility here.

Further research is required to explore the incorporation of additional covariates such as rainfall or soil moisture, that can be important for other taxa. Results presented in this chapter are only first step towards understanding the effects of within-stage variation on life histories on natural stage-structured populations. However, field populations are subject to sources of variation not considered here such as environmental stochasticity and density dependence. The incorporation of these factors and the methodological challenges to estimate associated parameters from field data are explored in the second part of the next chapter.

3.6 Annexe. Analysing linear assumption of degree-day approaches.

Degree-days (DDs) is a unit that measures the energy available for driving development processes per unit time (days in general) (De Reaumur, 1735; Russelle *et al.*, 1984; Higley *et al.*, 1986). The cumulative degree-days (CDDs) – or *thermal time* – synthesises the integrated effect of temperature on the many individual physiological processes involved during the time taken for the development of poikilothermic species (Plant and Wilson, 1986; Bonhomme, 2000). In the IPLM approach, progress by discrete increments Δ tracks development as a non-linear function of temperature and enables one to reconstruct cumulative maturation time distributions (Fig. 3.3), which is analogous to tracking non-linear responses to CDDs.

Degree-day based development studies have been a valuable tool in the quantitative analysis of plant, pest and crop phenology in agricultural sciences (Taylor, 1981; Higley *et al.*, 1986; Gu and Novak, 2006; Merrill *et al.*, 2010; Evans *et al.*, 2014; Lewis *et al.*, 2015) – for example CDD has been used to estimate the timing of relevant biological events that determine the timing and impacts of crop pest outbreaks. Traditional DD-based models assume a linear relationship between temperature and development rate, and impose the definition of a base temperature T_b under which individuals are assumed not to develop. This approach is highly popular due to its simplicity, but several authors have suggested this assumption only provides accurate approximation over small temperature ranges and can be erroneous when modelling the developmental responses of organisms under large temperature variations (McMaster and Wilhelm, 1997; Bonhomme, 2000). In general, the errors of this approach will be magnified whenever non-linearity in developmental rate curves becomes non-negligible, as can be the case in temperature regimes with a wide diurnal range. Some approaches that attempt to overcome this limitation (e.g. Osawa *et al.* (1983); Logan *et al.* (1976); Dennis *et al.* (1986); Régnière and Powell (2013)) were described in section 3.1.

Here we investigate the validity of the linearity assumption of DD models by using CDD estimates obtained from the *Culicoides* IPLM model. To do this, we first plot the cumulative distribution function (CDF) of the CDD required to complete each stage at each experimental temperature using the fitted IPLM model – this was done for a range of values of T_b in $0^\circ\text{C} - 14^\circ\text{C}$. Under linearity, these CDFs at each temperature should be identical, thus we select the T_b that minimises the distance between these CDFs to maximise the evidence for linearity per stage. Finally, with T_b fixed, we plot uncertainty in the expected CDD required to develop from egg to the completion of a first gonotrophic cycle. This is done over all temperatures, to evaluate if the thermal requirements for this

development are independent of temperature (as predicted by the linearity hypothesis) or not. Details of these steps are described below.

Methods. Daily degree-days, $DD(t)$, is typically obtained by employing some variant of the canonical expression $DD(t) = \frac{(T_t^{\max} + T_t^{\min})}{2} - T_b$, where T_b is the threshold temperature for development, and T_t^{\min} and T_t^{\max} (both assumed to be greater than T_b , otherwise $DD(t) = 0$) are the daily minimum and maximum temperature recorded at day t . If development takes t days, the corresponding cumulative degree-day required is

$$CDD(t) = \sum_{\tau=1}^t DD(\tau). \quad (3.3)$$

At a fixed temperature T , equation (3.3) reduces to

$$CDD(t) = tH(T - T_b), \quad (3.4)$$

with H the Heaviside step function.

Naturally, uncertainty of *Culicoides* IPLM estimates is lowest at (the fixed) experimental temperatures T_e (see table B.1). At these temperatures, we re-scaled (from 1000 MCMC samples) the estimated CDFs of maturation times to CDFs of thermal times by using equation (3.4) for a set of plausible T_b in the range $\{0^\circ C, 14^\circ C\}$ (step size = 0.1). This range was taken based on literature and the data we had (i.e. clearly $T_b < 15^\circ C$). We define a *distance* metric, d_l , that for a given T_b , sums the distance between the estimated CDFs of thermal time across empirical temperatures T_e for sample l ,

$$d_l(T_b) = \sum_{i \in T_e} \sum_{j \in T_e, j \neq i} \frac{1}{2} \int_0^\infty |F_i^{(l)}(CDD) - F_j^{(l)}(CDD)| dCDD, \quad (3.5)$$

where $F^{(l)}$ is the CDF of thermal time obtained using estimate set associated to sample l . Note, the total distance D (for N MCMC samples) is just $D(T_b) = \sum_{l=1}^N d_l(T_b)$. We determined \hat{T}_b , the “optimal” T_b as the one that minimises distance D , which in other words, is the T_b that maximises the evidence of linearity found in data.

Once \hat{T}_b had been identified for each stage, the expected CDD (ECDD) required to complete the four stages was calculated, from the CDF for thermal time (as used above) associated with T_b and each experimental temperature. Note, for adults, only time to completion of the first gonotrophic cycle (hereafter named GC1) was considered. Plots of these expected thermal times (per temperature) were then assessed for evidence of linearity

i.e. independence from temperature T . The ECDD for completing all four stages was also obtained.

Results. Minimising the total distance D (between CDFs of thermal time for stage completion at each experimental temperature) enabled \hat{T}_b to be determined for every stage (Fig. 3.11). For egg, larval and pupal stages, our estimates of \hat{T}_b were within the ranges of T_b reported in published works on *Culicoides variipennis* from which most of our data comes from (table 3.1). The expected thermal time (ECDD) for egg, larva, pupa, and GC1 completion showed that DD requirements are not independent of temperature (Fig. 3.12). Similarly, the ECDD required to develop from egg to GC1 completion is also not independent of temperature (Fig. 3.13).

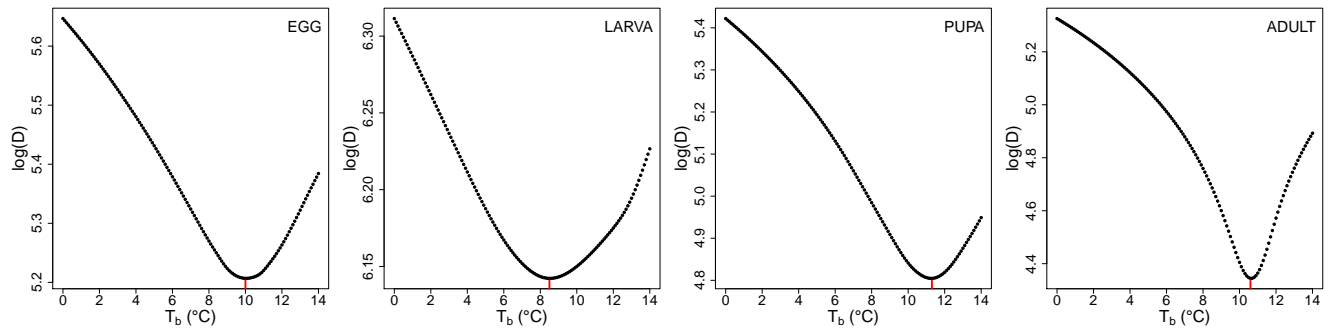


Figure 3.11. For every stage, a threshold temperature for development, T_b , was explored among values in the set $\{0^\circ\text{C}, 0.1^\circ\text{C}, \dots, 14^\circ\text{C}\}$ by minimising an auxiliary metric, *total distance* D , between the CDFs of CDD requirements for stage completion of different fixed experimental temperatures. Here, D is shown on the logarithmic scale. Optimised values (red bars) which maximise the evidence for linearity were $T_b^E = 10.0^\circ\text{C}$, $T_b^L = 8.5^\circ\text{C}$, $T_b^P = 11.3^\circ\text{C}$, $T_b^A = 10.6^\circ\text{C}$. For adults, maturation times for the first gonotrophic cycle only were considered.

Stage	T_b (°C)		
	Mullens <i>et al.</i> (1983)	Vaughan <i>et al.</i> (1987)	IPLM
E	9.8	-	10.0
L	10.7	7.5	8.5
P	10.6	11.7 [†]	11.3
1 st GC	-	-	10.6

Table 3.1. A comparison of base temperature (T_b) estimates published in the literature and estimated in the current work. [†] Data at 35°C not included for estimation of T_b by (Vaughan and Turner, 1987).

In Fig. 3.12, we can see that thermal requirement for larva development is much higher than for the other stages, and egg and pupa exhibit lower thermal requirements. For egg, larva and pupa, the ranges where linearity is valid do not exceed (roughly) a length of seven $^\circ\text{C}$ units. For adults, the evidence of non-linearity is weaker, although care must be taken given the high uncertainty at higher temperatures. Eggs and larvae show a clear optimal developmental temperature (around 20°C and 23°C respectively), associated to lower developmental

requirements. For pupae, the requirement is lowest at lower temperatures.

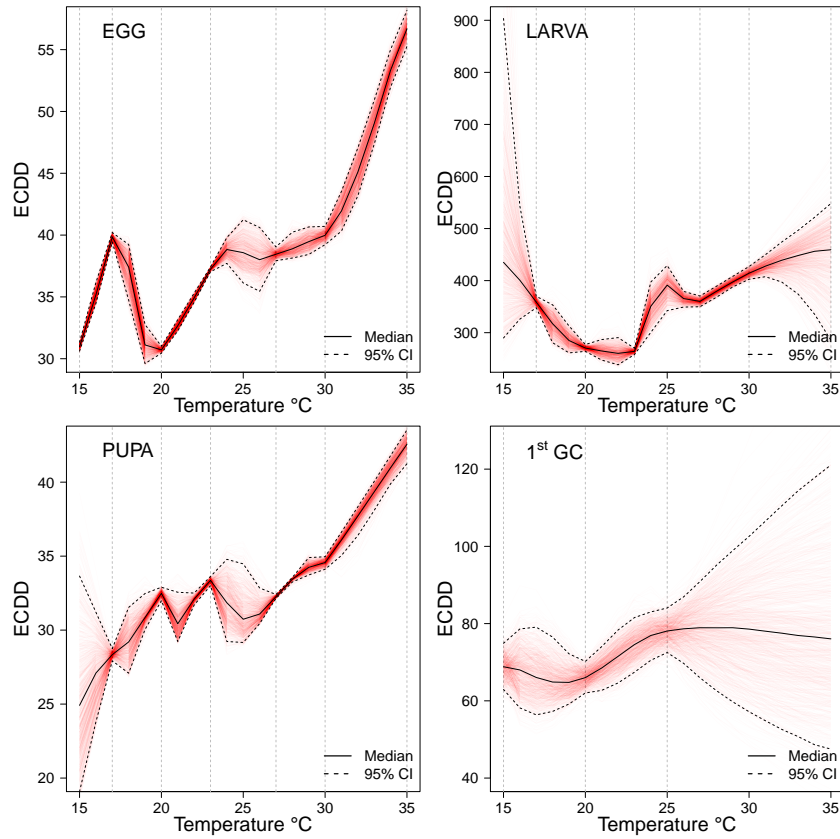


Figure 3.12. Expected cumulative degree-days (ECDD) requirements for stage completion from 1000 samples of MCMC are shown (red lines). For adult *Culicoides*, only time to completion of first gonotrophic cycle is considered. Median (continuous line) and 95% CI (dashed lines) are shown. The developmental threshold T_b used for every stage was identified in Fig. 3.11. Vertical lines are a reference to experimental temperatures.

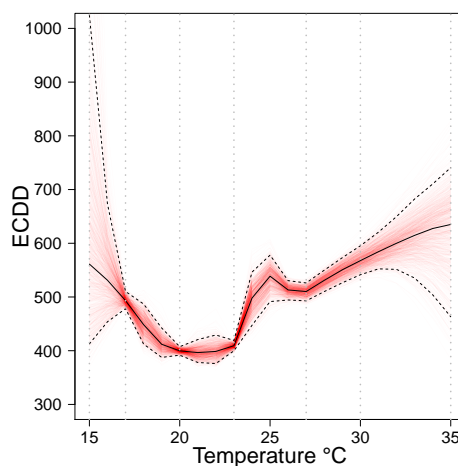


Figure 3.13. Expected cumulative degree-days (ECDD) requirement for *Culicoides* development from egg to completion of the first gonotrophic cycle (red lines). The developmental threshold T_b used for every stage was identified in Fig. 3.11. Median (continuous line) and 95% CI (dashed lines) of 1000 MCMC samples are shown. Vertical lines are a reference to experimental temperatures.

Discussion. Estimated maturation times from the *Culicoides* IPLM model enabled us to test linearity assumption underlying most degree-day development models. We found strong evidence of non-linearity between development and temperature, confirming that inferences based on linear DD approaches can be inaccurate in the face of typical temperature ranges encountered in field studies.

A striking similarity between the ECDD for larva and ECDD for development from egg to GC1 completion highlights the importance of the larva stage regarding the generation time of *Culicoides*. Little is known about the larval stages of most of *Culicoides* species of epidemiological interest and these results provide important information that could be used for motivating further studies on larva development or assessing vector control strategies focused on the use of larvicides.

Diapause (a metabolic status by which organisms reduce activity in order to survive environmental unfavorable periods) has been suggested to facilitate overwintering among *Culicoides* larvae in temperate regions. Some entomologists have speculated that photoperiod could be a main trigger for diapause onset, while temperature most likely trigger the end of diapause (Carpenter, personal communication). Most DD-based models, including the *Culicoides* IPLM, ignore life-history aspects that are not essentially driven by temperature, such as a photoperiod triggered onset of diapause. Thus, although of recognised utility, these DD-based approaches are limited when aiming predict or monitor thermal time for populations undergoing diapause. Although our approach could be extended to include diapause as another stage, we did not explore this possibility due to a paucity of data.

In addition to temperature, food availability and larval density are known to affect *Culicoides* larval development (Mellor *et al.*, 2000). Therefore, DD requirements for stage completion estimated from laboratory experiments that do not consider these factors might be biased when extended to applications in the field. In section 4.2 we describe how easily density-dependence and environmental stochasticity can be incorporated into the IPLM framework and make some preliminary studies regarding the identifiability of these parameters using simulated field data.

Conclusion. Analysis of the temperature-dependent *Culicoides* IPLM model demonstrates that the traditional assumption of a linear relationship between cumulative degree-day and development rate does not hold outside the bounds of a narrow range of mild temperatures. We have shown that IPLM methods provide a simple framework that readily accounts for non-linearity in the temperature effects on vital rates. The flexibility of the IPLM to incorporate non-linear responses of vital rates to covariates is an important feature that bolsters the utility of these methods for applied purposes such as conservation or pest / vector control.

Chapter 4

Extending the utility of the IPLM framework

In chapter 4, two potential research lines are explored by using the *Culicoides* IPLM of chapter 3. First, projections of *Culicoides* adult densities were used to analyse temporal variation in the R_0 of bluetongue following an induced reduction in adult survival. We show that generation time, determined by how temperature drives the *Culicoides* life-cycle, affects the efficiency of vector control to reduce or halt bluetongue transmission and we outline implications for integrated vector management.

Secondly, I explore methods for upscaling to natural scenarios by using IPLMs in a state-space model with a view to making inference from field data (e.g. adult time series). For this, density-dependence and environmental noise are included in the IPLM model and combined with a noisy observation process. In a preliminary study with simulated data, I explore a potential method for inference that uses a recent simulation-based approach (synthetic likelihood, Wood (2010)). The results indicate that, under certain conditions, these techniques are capable of identifying the strength of population regulation and environmental stochasticity from vector surveillance data. Methodological challenges and potential avenues for future research are outlined.

4.1 Estimating the effects of vector control on the R_0 of bluetongue

4.1.1 Introduction

In epidemiology, the basic reproduction ratio (or number) R_0 , the number of secondary infections arising from the introduction of one primary infectious individual in a susceptible population, is a common index used to analyse the potential of disease invasion (Murray, 2002; Heffernan *et al.*, 2005). A disease is expected to increase in frequency among a host population only if $R_0 > 1$, thus R_0 is a measure to assess the risk a disease poses to a population.

A few works have derived the R_0 for bluetongue virus (BTV) transmission from mechanistic models (Gubbins *et al.*, 2008; Hartemink *et al.*, 2009; Guis *et al.*, 2012; Turner *et al.*, 2013). These models have studied the effects of different (biotic and abiotic) factors and assumptions underlying this index, as well as the epidemiological consequences of manipulating such factors. For example, in Guis *et al.* (2012), the authors use R_0 to assess the effects of climate change on the risk of BTV emergence across Europe by including high-resolution climate projection (up to 2050) scenarios, while in Turner *et al.* (2013), the effects on R_0 of differences in vector abundance, competence, host preference and species (*C. pulicaris* and *C. obsoletus* groups) in South Africa have been analysed.

In Gubbins *et al.* (2008), the authors obtain an expression for R_0 by using next generation methods from a system of ordinary differential equations that model BTV transmission in a mixed population of cattle and sheep. Their models consists of susceptible (i.e. uninfected), infected and recovered hosts; and susceptible, latent (i.e. infected, but not infectious) and infectious vectors. The R_0 expression they obtain is

$$R_0 = \sqrt{\frac{b\beta a^2}{\mu} \frac{\gamma}{\mu + \gamma} \left(\frac{m_c \varphi^2}{r_c + d_c} + \frac{m_s (1 - \varphi)^2}{r_s + d_s} \right)}, \quad (4.1)$$

where b is the probability of transmission from vector to host, β the probability of transmission from host to vector, a the biting rate, μ the vector (adult) mortality rate, $1/\gamma$ the mean extrinsic incubation period, m the vector-to-host ratio (i.e. $m = A/H$, with A and H accounting for (adult) vector and host densities respectively), φ the proportion of bites on cattle, $1/r$ the duration of viraemia and $1/d$ the disease-induced mortality rate. Sub-indices c and s refer to cattle and sheep, the only hosts considered in the model. This expression assumes the duration of viraemia in hosts (cattle and sheep), the extrinsic incubation period and vector (adult) longevity follow exponential

distributions.

Some of the parameters involved in expression 4.1 can be manipulated via vector control practices. Vector control responses include use of insecticides or pathogens to (or direct removal of) larval breeding sites; application of insecticides on host animals or resting sites targeting adults; use of repellents or housing livestock to reduce contact; and use of host kairomones to lure and kill adult midges (Carpenter *et al.*, 2008). However, current knowledge of the ecology (e.g. breeding sites) and behaviour (e.g. host-seeking and resting behavior) of midges do is too limited to permit quantitative prediction of the success of vector control interventions that aim to reduce the spread of BTV.

Limited knowledge on breeding habits, combined with undesirable consequences in the use of larvicides (e.g. cross-resistance in larval population, difficulty achieving adequate coverage, potential impact on aquatic invertebrates and increasing environmental concerns (Clements and Rogers, 1968)) have increased interest in alternative methods that target *Culicoides* adults. Adult control methods either studied or applied in the field include insecticide-treated livestock (the main method used in Europe for BTV), insecticide-treated screens, environmental spraying (e.g. Linley and Jordan (1992) reported mortality increased up to 90% in field studies with *C. furens* exposed to different adulticides), removal trapping methods (e.g. traps baited with carbon dioxide and octenol) and biocontrol agents (e.g. fungicides) (Carpenter *et al.*, 2008; Ansari *et al.*, 2011). Because biting midges are relatively poor flyers (in general dispersion is lower than two miles from original breeding site (Lillie *et al.*, 1985)), control targeting adults can, in theory, quickly reduce local biting populations.

4.1.2 Methods

The aim here is to provide preliminary insights into the relative effectiveness of an adult-density reduction strategy to reduce the transmission risk of BTV. Specifically, we explore the effects of a sudden antropogenically induced reduction in adult survival – as could be the goal of *Culicoides* control with adulticides – on the associated R_0 . For this, we adapt and simplify expression 4.1 by assuming that $m_s = m_c = m$, and integrate adult density projections from the *Culicoides* IPLM model (chapter 3). The intervention reduces the daily probability of adult survival ν to $\nu_c = \alpha\nu$, with $0 < \alpha < 1$. We compute at a series of times t following the initialisation of vector control ($t = 0, 1, 2, \dots$), changes in the temperature-dependent basic reproductive ratio, $R_0^c(t)$, relative to its value immediately prior to starting vector control intervention, R_0 .

We assume the vector population grows according to the IPLM model and we neglect density dependence. As a consequence of the hypothetical control program, the vector (i.e. adult) mortality rate becomes $\mu_c > \mu$, while the vector-host ratio is $m_c = A_c/H$, with A_c being the adult population under control measures. We assume (realistically) that host density H is constant. Using these values in the expression for R_0 (equation 4.1), and taking the ratio $r_T(t) = R_0^c(t)/R_0$, we obtain

$$r_T(t) = \sqrt{\frac{\mu(\mu + \gamma_T)}{\mu_c(\mu_c + \gamma_T)} \frac{A_c(t)}{A_0}}, \quad (4.2)$$

where A_0 is assumed to be the adult population associated to ν (prior to vector control), T is temperature and the extrinsic incubation period γ_T is a monotonically increasing function of T , $\gamma_T = 0.0003 \times T \times (T - 10.4057)$, taken from Mullens *et al.* (2004). Assuming density-independent mortality among adults, the mortality rate μ of expression 4.1 relates to daily survival ν of the *Culicoides* IPLM model via the identity

$$\mu = \log\left(\frac{1}{\nu}\right). \quad (4.3)$$

Note, since the principal source of density-dependence in survival is often considered to affect immature stages, neglecting density dependence in this step is not expected to be an important source of bias for calculating μ .

We investigate the effects of reducing adult survival by a constant reduction factor α . For every $\alpha \in \{\frac{1}{50}, \frac{2}{50}, \frac{3}{50}, \dots, 1\}$, the dynamics of a *Culicoides* population are projected, at fixed temperatures, and the ratio $r_T(t)$ is evaluated at each (daily) time-step during 15 weeks. Average daily estimates of $r_T(t)$ are obtained for IPLM parameters obtained from 1000 lines of MCMC output (see section 3.3). Temperatures were fixed at $T \in \{15^\circ C, 20^\circ C, 25^\circ C\}$; these temperatures were chosen because 1) they cover a range where the uncertainty of the *Culicoides* IPLM model is relatively low (fig 3.5); and 2) R_0 estimates had previously been published at these temperatures. For every projection, the initial population was set to its stable stage distribution in the absence of control.

Various estimates of the R_0 of the BTV transmission are available in the literature (Gubbins *et al.*, 2008; Hartemink *et al.*, 2009; Guis *et al.*, 2012; Turner *et al.*, 2013). We consider two of them for aiding these analyses (table 4.1) while the remaining are addressed in discussion section. These works provide a rough range of R_0 from which we take the values $R_0 \in \{2, 3, 4\}$ to analyse our results.

	Temperature			Author
	15°C	20°C	25°C	
R_0	1.4-3.1	1.6-3.6	0.9-2.4	Gubbins <i>et al.</i> (2008)
	2.53		2.7	Turner <i>et al.</i> (2013)

Table 4.1. Available estimates of R_0 for BTV at 3 temperatures from published literature.

4.1.3 Results

The effects of the different values of α on $r_T(t)$ trajectories were plotted for each temperature considered (Fig 4.1). In all cases, control effects (i.e. changing ν to ν_c) caused a large initial reduction in $R_0^c(0)$. The non-linear relationship between α and r_T (expressions 4.2 and 4.5) is observed at $t = 0$ for the different proportional decreasing α -values (gray gradient bars) considered. Large reductions in $R_0^c(0)$ were possible at 15°C with relatively modest control efforts, while at higher temperatures this effect was attenuated. For example, $r_T(0) = 0.5$ was possible at 15°C with $\alpha \simeq 0.9$, whereas $\alpha \simeq 0.8$ was required at 25°C.

In general, oscillations are observed. These oscillations are associated to transient dynamics in adult populations following the sudden change of adult survival. If control is weak ($\alpha \simeq 1$), r_T trajectories indicate that after the initial reduction, R_0^c consistently increases at all considered temperatures. As α decreases (i.e. adult survival ν_c decreases), oscillations are characterised by an initial decline in r_T . The duration of this initial decline is temperature dependent and is linked to the expected generation times of the IPLM. After the initial decline, the trend in r_T is to either increase (for relatively high α values) or decrease (for α less than some threshold).

The values of α such that $r_T(0)$ is approximately $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ were identified in red, pink and blue respectively in the legends of figure 4.1. Trajectories obtained for corresponding values of α were indicated in solid lines following the same color code. If R_0 was initially $R_0 = 2$, $R_0 = 3$ or $R_0 = 4$ respectively, these values of α would correspond to the epidemiologically important threshold $R_0^c(0) = 1$. Horizontal dotted lines compare trajectories in $r_T(t)$ to $R_0^c(t = 0)$ in these cases.

Colored curves represent trajectories of $r_T(t)$ associated with the largest value of α that reduces an initial R_0 of 2, 3 or 4 to $R_0^c(0) < 1$. These trajectories indicate that the effects of initially achieving $R_0^c(0) < 1$ can be lost in many cases if the vector population is free to grow ($\lambda_1 > 1$) during the control program. For example, at 15°C and $R_0 = 2$, $\alpha = 0.92$ predicts BT decline (i.e. $R_0^c(t) < 1$) for between 5 to 6 weeks after which $R_0^c(t)$ increases to above the threshold $R_0^c(t) = 1$ (red dashed line). Note, that below a threshold of roughly $\alpha \approx 0.86$, trajectories display a continued (slightly oscillatory) decline of $r_T(t)$. At the same temperature, an initial R_0 of 3

or 4 demands a stronger reduction in ν_c ($\alpha < 0.84$ and $\alpha < 0.78$ respectively) in order to achieve $R_0^c(0) < 1$. The corresponding trajectories show a trend of continued (slightly oscillatory) decline of $R_0^c(t)$ indicating that $\lambda_1 < 1$.

Initial values of $R_0 = 2$, $R_0 = 3$ or $R_0 = 4$ at higher temperatures demand a stronger reduction of adult survival ν_c in order to achieve a) $R_0^c(0) < 1$, and b) $R_0^c(t) < 1$ for $t \geq 15$ weeks. In general, whether or not projections predict a continued long-term decline in $R_0^c(t)$ depends on temperature and the strength of control. At 15°C , long-term decline (associated with $\lambda_1 < 1$) is achieved with $\alpha \lesssim 0.86$, while at 20°C and 25°C , this is achieved with $\alpha \lesssim 0.74$ and $\alpha \lesssim 0.66$ respectively.

An R_0 mapping study in the Netherlands suggests that for temperatures between 15°C and 20°C , R_0 for BTV can take values greater than 10 or even than 20 (Hartemink *et al.*, 2009). According to our analyses, at 15°C , reducing initial $R_0 = 10$ or $R_0 = 20$ to values $R_0^c(0) < 1$ require $\alpha \lesssim 0.46$ and $\alpha \lesssim 0.2$ respectively, while at 20°C , $\alpha \lesssim 0.32$ and $\alpha \lesssim 0.032$ would be required. Some of these α values are unlikely to be reached with the application of adulticides, which means that in such scenarios, an integrated vector management approach should be considered.

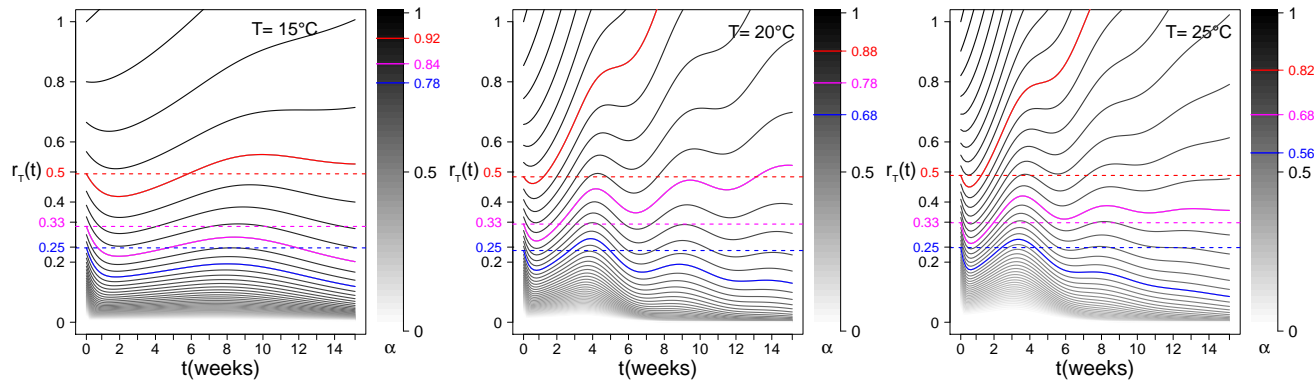


Figure 4.1. Plots of the mean ratio $r_T = R_0^c(t)/R_0$ (equation 4.2) that measures temporal changes in the basic reproduction ratio for bluetongue transmission under vector control that reduces adult survival probability ν_c by a factor α . $R_0^c(t)$ is the basic reproduction ratio t time-steps after initial control, and R_0 is the basic reproduction ratio immediately prior to vector control intervention. For each temperature, 50 equidistant values of α were considered ($\alpha \in \{1, \frac{49}{50}, \dots, \frac{2}{50}, \frac{1}{50}\}$), indicated by the gray color gradient bars. For each α considered, $r_T(t)$ trajectories over 15 weeks are shown. Colored curves represent trajectories of $r_T(t)$ associated with the largest value of α (following same color code) that reduces an initial R_0 of 2, 3 or 4 to $R_0^c(0) \lesssim 1$ (in red, pink and blue respectively) immediately upon starting vector control. Horizontal dotted lines represent threshold values in $r_T(t)$ below which $R_0^c(t=0) < 1$ for these cases. These trajectories indicate that the initial effects of initially reducing $R_0^c(t)$ to below 1 can be lost if the vector population is free to grow during the control program. Long-term decline (associated with $\lambda_1 < 1$) is achieved at 15°C with $\alpha \lesssim 0.86$, while at 20°C and 25°C , this is achieved with $\alpha \lesssim 0.74$ and $\alpha \lesssim 0.66$ respectively.

4.1.4 Discussion

The basic reproduction ratio, R_0 , provides a quantitative framework to address the question of level of risk posed by a disease to a population. While a full treatment of estimating R_0 from data was beyond the scope of this analysis, we assessed the question of how much the R_0 of BTV transmission changes over the course of vector control targeting *Culicoides* adult survival, via a novel index $r_T(t)$.

The expression for R_0 underlying our analysis was taken from Gubbins *et al.* (2008), where uncertainty and sensitivity analyses carried out by the authors identified temperature and the vector-to-host ratio, $m = A/H$, as being among the most important parameters in determining the magnitude of R_0 . The *Culicoides* IPLM model (chapter 3) provided a basis for a temperature-dependent framework for $r_T(t)$.

Estimates on adult *Culicoides* daily survival decrease monotonically in the range $15^\circ\text{C} - 25^\circ\text{C}$ (Fig. 3.5), from which we might expect a lower control effort would be required at higher temperatures to achieve $R_0^c(0) < 1$, independently of the initial R_0 considered. However, our results suggest the opposite, i.e. stronger efforts to reduce adult survival ν_c (via the factor α) are necessary as temperature increases in order to ensure $R_0^c(0) < 1$ whichever R_0 is considered (see for example $R_0 = 2, 3$ or 4). This may reflect the importance of the (non-linear) relationships between R_0 and the extrinsic incubation rate γ_T , whose variation among vector species is poorly known. Improving estimates of γ_T for specific species involved in the transmission of BTV can thus enhance our R_0 estimates. However, changes in λ_1 and generation time with respect to temperature are probably important too.

At any temperature considered, whenever α is sufficiently large that $r_T(0)$ is greater than some threshold in (roughly) $(0.8, 1)$, the subsequent trajectories of $r_T(t)$ consistently increase monotonically and the vector control can be successful for a period ranging from (roughly) 2 to 6 weeks only for initial R_0 very close to 1. For the remaining cases, the transient decline observed in $r_T(0)$ during the first weeks does not provide a useful indicator for the successful long-term reduction of R_0 , given that if α is not low enough, the long-term trend indicates that BTV will (if introduced) spread after the temporary reduction. However, some advantage could be taken from the predicted transient period, for instance to help implement vaccination strategies (Pioz *et al.*, 2012) during a period when $R_0^c(t) < 1$, or to reduce the proportion of the year where $R_0^c(t) > 1$. Thus, even if only transitory periods of $R_0^c(t) < 1$ can be achieved by vector control, the contribution can be important in an integrated disease management program where other control measures are also used to reduce BTV spread.

Temperature influences many processes involved in the transmission of BTV, such as the vector mortality

rate or the extrinsic incubation period (Gerry *et al.*, 2001; Mullens *et al.*, 2015). Moreover, temperature is a driving force for immature developmental rates that influence the number of generations produced and the adult population size that can result in a season (Mullens *et al.*, 2004). The IPLM framework captures this driving force in vector dynamics and gives an alternative to field estimates of some vector-related parameters (e.g. biting rate can be estimated via the inverse of gonotrophic cycle duration), thus providing a step towards obtaining numerical methods for assessing the risk of *Culicoides*-borne disease transmission, and – more importantly – for estimating associated uncertainties.

Note, exponential growth (with rate λ_1) underlies the (density-independent) IPLM *Culicoides* life-cycle model (chapter 3). Although this is not true in most of natural systems, there are many situations where populations can exhibit near-exponential behavior. This can arise when densities are well below carrying capacities, in scenarios such as invasion, during vector control, trophic release (i.e. removal of key predators), recovery from extreme climatic events or land use change that augments the carrying capacity (e.g. habitat creation). In any of such cases, the exponential growth assumption could be adequate for guiding control. Moreover, if α is sufficiently small such that $R_0^c(t) < 1$ for large t , then having a density-dependent model is probably superfluous. That said, control programs need to take care that juvenile survival does not increase as adult survival are reduced.

Limitations in the effectiveness of reducing adult populations in scenarios where high R_0 values are expected – as is the case in some projections of BTV transmission in Europe (Hartemink *et al.*, 2009; Guis *et al.*, 2012) – suggest that an integrated vector management approach appears as a more efficient option to reduce initial R_0 , since other measures could concur with the application of adulticides to debilitate the potential of BTV transmission. Currently, vaccination is the most effective BTV control measure, while insecticide-treated livestock has also been implemented. However, for emergent viruses, vaccines are often not ready in time – this was the case for the Schmallerberg virus epidemics of 2011-2013 in Europe. In these scenarios, having alternative strategies at hand can be highly useful. Larval control (e.g. reduction/elimination of breeding sites, or application of larvicides) appears as the most effective complementary measure to reduce adult abundance of the next generation (Carpenter *et al.*, 2013). Physical barriers (not available yet in the field) is another alternative that could be considered to reduce biting rates.

Expression 4.1 relies on a BTV transmission model where structure is defined by health status dynamics described via ordinary differential equations (ODEs). For adult vectors, stages are susceptible, exposed, infected

and removed (usually known as SEIR model (Keeling and Rohani, 2008)), and expression 4.1 assumes an exponential probability distribution function of the extrinsic incubation period γ_T . Limitations associated to this assumption could be overcome by approaching BTV transmission with regression-based stage-specific integral projection models (IPMs). In this way, the IPM approach would allow, for example, to capture differential effects of temperature on the *Culicoides* exposed state thus providing a more realistic distribution for studying temperature dependence of γ_T .

Unfortunately, important gaps in the knowledge of BTV and of key indices associated to BTV's vector (i.e. *Culicoides*) exist that hamper the understanding of transmission dynamics. For example, it is still unclear how BTV overwinters in temperate regions (Mayo *et al.*, 2014), or, particularly for Europe – struck by unprecedented epidemics of BTV over the last decade –, several parameters used to describe vector features/behavior of local species involved in the BTV spread are still not available.

Moreover, epidemiologists lack field methods that enable to obtain reliable (i.e. unbiased) estimates of the effects of various biotic and abiotic factors involved in R_0 (or other indices of disease transmission), as well as methods to predict or evaluate the efficiency of different vector control strategies other than vaccination (Græsbøll *et al.*, 2014). Obtaining estimates required to efficiently determine alternative vector control targets in the field makes the ultimate objective – interrupting BTV transmission – a challenging interdisciplinary task (Guis *et al.*, 2012; Mullens *et al.*, 2015).

In this study, we used R_0 to evaluate the efficacy of a particular control measure. A more complete R_0 -based analysis is necessary to determine how combinations of control measures might be more effective against BTV transmission. A major drawback of common approaches is that calculating R_0 relies on parameters whose estimates often have high uncertainty. Not accounting for this uncertainty greatly limits the potential use of R_0 -based analysis for assessing potential control interventions. We do not explore this avenue further here, but clearly, Bayesian methods can play an important role in quantifying the effects of uncertainty on R_0 and calculating the probability of $R_0 < 1$ at any time.

4.1.5 Conclusion and perspectives.

Temperature-dependent projections of adult densities from the *Culicoides* IPLM (chapter 3) enable us to model the temperature response of two of the parameters of greatest sensitivity involved in R_0 of bluetongue virus (BTV)

(Gubbins *et al.*, 2008), namely adult vector survival and vector-to-host ratio. This marks an improvement to current techniques that typically employ more simplified approximations such as constant parameters. We analysed the impact of vector control targeting adult *Culicoides* survival on the temporal evolution of R_0 at different temperatures.

We found that for $R_0 > 1$ and even relatively low (up to 4), reducing adult survival is not, in most of cases, sufficient to predict reduction of R_0 below 1. Also, we found that $R_0^c(t)$ is sensitive to λ_1 and generation time, determined by how temperature drives the whole midges life-cycle.

Here, for simplicity, we analysed just one BTV intervention strategy carried out once. However, it is well accepted that control programs that integrate repeated and/or multiple control strategies have a greater chance of success. Additional work should address both optimal intervention frequency and how other strategies, either affecting different parameters or the same parameters to different degrees, further affect trajectories of $R_0^c(t)$.

4.2 Towards a state-space IPLM framework

Multiple factors acting on different life-cycle stages influence the dynamics of stage-structured populations in non-linear ways that complicate the assessment, management or control of populations. Field data is critically important to determine which of these factors are the most relevant, what life stages they impact the most, and how these impacts can be translated into ecological inferences. The relationships among endogenous and exogenous driving forces and the life stages they influence are often difficult to piece together into statistical models without sacrificing model realism for the sake of tractability. Thus, developing frameworks that address the complexity and sources of variance that characterise natural systems / populations is a major challenge for ecologists.

In this section, we take some preliminary steps towards developing such predictive framework based upon IPLMs. We set up a state-space model (SSM) with an stochastic density-dependent IPLM process model with a view to taking advantage of field time-series data that provide a primary source of statistical power for analysing populations (i.e. fitting models to data, generating predictions and quantify uncertainties).

This section is organised as follows. First, we analyse outputs from an IPLM model extended to include density-dependence and environmental stochasticity. Next, we perform a preliminary analysis using a simulation-based method – the synthetic likelihood – within an MCMC algorithm for inference with an IPLM state-space model. Finally, preliminary results and some perspectives for further research are outlined with a view towards an IPLM-based state-space framework for the study of stage-structured population's dynamics *in natura*.

4.2.1 Extending the IPLM model

Let us recall the *Culicoides* IPLM life-cycle model of chapter 3. The female only egg-larva-pupa-adult (ELPA) model is

$$\begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_t = \begin{bmatrix} \mathbf{W}_E & \mathbf{0} & \mathbf{0} & \mathbf{F}_A \\ \mathbf{B}_E & \mathbf{W}_L & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_L & \mathbf{W}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_P & \mathbf{G}_A \end{bmatrix} \begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_{t-1}, \quad (4.4)$$

where $\mathbf{G}_A = \mathbf{W}_A + \mathbf{B}_A$ models multiple gonotrophic cycles, and stage-specific parameter sets are $\Theta_S = \{\mu_S, \kappa_S, \nu_S, r_S\}$, with $S \in \{\mathbf{E}, \mathbf{L}, \mathbf{P}, \mathbf{A}\}$. This model was previously used to obtain estimates for temperature-

dependent stage-specific developmental rates (mean and standard deviation, $\{\mu_s, \sigma_s\}$), and daily survival ν_s from *Culicoides* laboratory data (Fig. 3.5).

So that our models provide a more realistic representation of dynamics *in natura*, here the IPLM approach is extended to include density-dependent population regulation and environmental stochasticity. A parameter, ϵ_t , accounts for the effects of stochasticity on stage-specific mortality ν'_s , and is assumed to account for unmeasured fluctuations in environmental (biotic and abiotic) covariates. The effects are defined on the logit scale, such that

$$\text{logit}(\nu'_{s,t}) = \text{logit}(\nu_{s,t}) + \epsilon_{s,t}, \quad (4.5)$$

where ϵ_t is drawn from a normal Gaussian distribution, $\epsilon_t \sim \text{Normal}(0, \sigma_\epsilon)$, and ν_s is survival of stage S in the absence of density-dependence (as in chapter 3).

Density-dependence can reflect one or more regulatory mechanisms such as cannibalism, parasitism or competition for resources. In *Culicoides*, the strongest evidence of density-dependence corresponds to the development and survival of larvae (Akey *et al.*, 1978; Linley, 1985; Mullens *et al.*, 2015). We focus thus on this stage and introduce a new parameter, ω , that accounts for density-dependence in larval survival. For simplicity, we assume ω is constant. Both ϵ_t and ω are assumed independent processes, thus survival ν'_L in the extended IPLM is

$$\text{logit}(\nu'_{L,t}) = \text{logit}(\nu_{L,t}) + \epsilon_{L,t} - \omega \mathbf{L}_t, \quad (4.6)$$

where \mathbf{L} is the total density over the r_L larval substages.

Exploratory simulation analysis with the extended IPLM. To explore the behavior of the density-dependent IPLM, we projected *Culicoides* dynamics in response to a seasonal temperature profile,

$$T_t = v + w \cos\left(t \frac{2\pi}{365}\right), \quad (4.7)$$

where v and w were set such that $\min(T_t) = 15^\circ\text{C}$ and $\max(T_t) = 30^\circ\text{C}$, with $t = 0$ the coldest day of the year. Temperature-dependent life cycle parameters were taken from interpolated posterior estimates ($\{\mu_s, \kappa_s, \nu_s\}$) obtained in chapter 3, with resolution parameters fixed in all cases to their maximum *a posteriori* (MAP) estimates,

$$r_E^{(\text{MAP})} = 6, r_L^{(\text{MAP})} = 9, r_P^{(\text{MAP})} = 31 \text{ and } r_A^{(\text{MAP})} = 8.$$

Firstly, for simplicity the median (from 1000 MCMC samples) of $\{\mu_S, \kappa_S, \nu_S\}$ was used (continuous black lines in Fig. 4.5), and fecundity was set to a constant ($F = 100$) so to be able to focus on effects of both noise and density dependence. Projections covered five years, with an initial population $N_0 = 10^7$ that was allocated among all substages, according to substage proportions sampled from a Dirichlet distribution $\text{Dir}(\alpha_1, \dots, \alpha_{r_T})$, where $\alpha = 1 \ \forall \alpha$. With this setup, we first explored the dynamics of the deterministic density-dependent IPLM, and then we included environmental noise.

In a second analysis, new projections were generated; this time developmental and fecundity parameters were obtained using (1000) MCMC samples to establish their daily value accordingly to the temperature profile given by expression 4.7. This enabled us to address the impact of uncertainty in demographic parameter uncertainties (Fig. 4.5) on the projections of both the deterministic and noisy density-dependent IPLMs.

Deterministic density-dependent IPLM. Different orders of magnitude of ω were considered ($\omega \in \{10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}\}$), giving oscillating adult populations that did not descend below $A = 10^5$. Results (Fig. 4.2) show that ω affects the carrying capacity K , thus the initial population either increases or decreases (depending on ω) to reach a regime in which the population oscillates with a repeating pattern that stabilises after (approximately) 2 to 4 years (depending on initial population size). Decreasing ω by one order of magnitude roughly translates into an order of magnitude increase in K . There is no evidence of a bifurcation in the chosen range of ω and (once stabilised) the oscillations for different values of ω follow a very similar seasonal pattern. Notably, we observe a seasonal effect on generation time that is very consistent (e.g. in the final year) for all ω .

Note, that unlike the density-independent IPLM of chapter 3 (see Fig. 3.10), where the same temperature profile was used, the results here (Fig. 4.2) indicate a much stronger influence of transient dynamics that extends beyond the first year and are not damped in succeeding years. This long-term seasonal cyclicity manifests as large broad peaks in late autumn and early winter, and smaller narrow peaks in hotter months. This pattern arises from the differential responses of stages to temperature. For example, egg and adult survival decrease at higher temperatures, while larval and pupal survival show relatively less change through the temperature regime (Fig. 4.5). Moreover, a complex interplay between seasonality shifting stage-distributions, generation times, density-dependence and population growth rates is observed. However, despite this complex set of interactions, the dynamics settle into a regular predictable pattern.

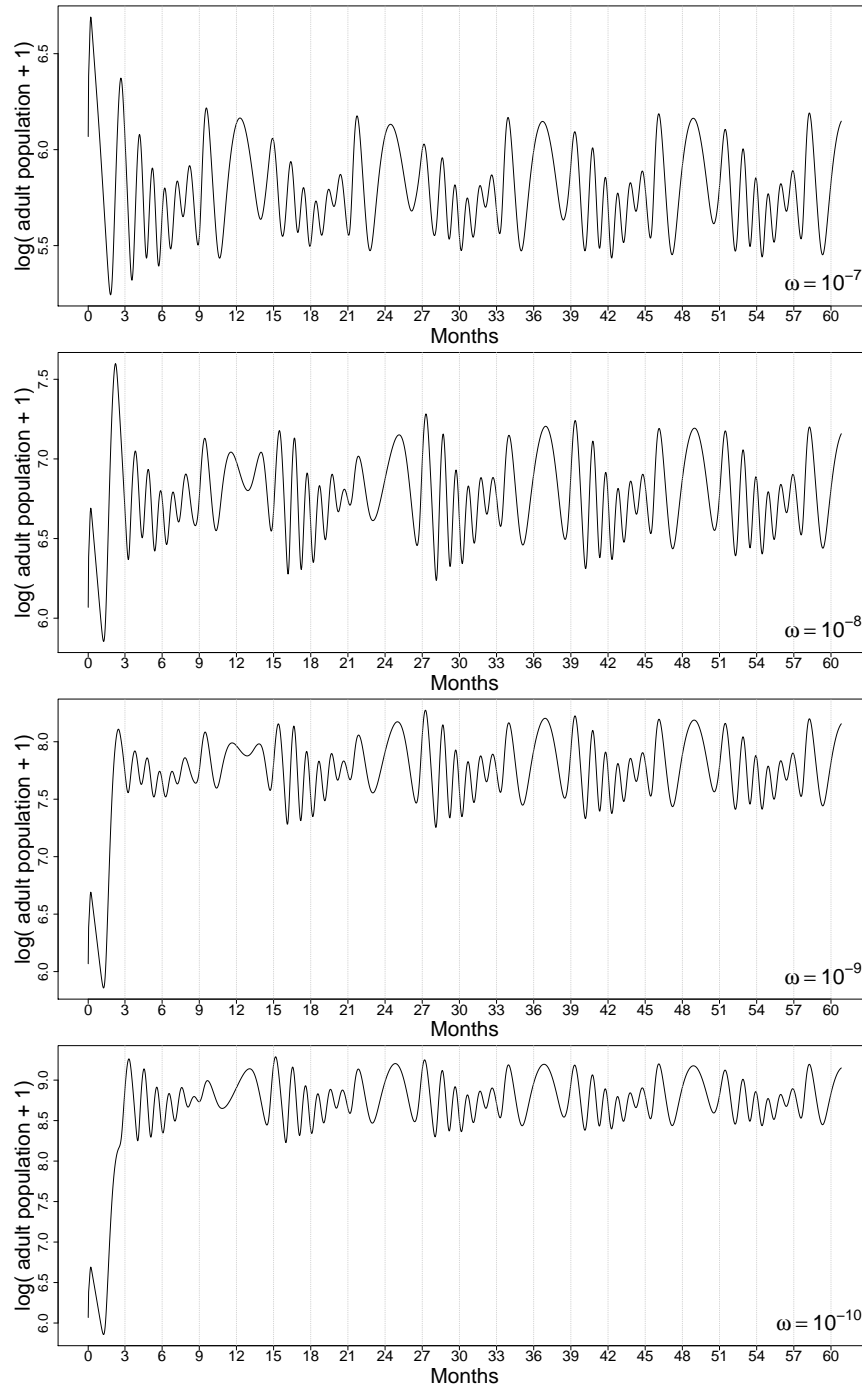


Figure 4.2. Simulated adult *Culicoides* density response to a seasonal temperature profile, for different strengths of density-dependence in larval survival (ω). Daily temperature varies following a (cosine) profile of period one year, and affects *Culicoides* life-cycle parameters according to estimates from chapter 3. Here, the median of life-cycle parameters (bold line of Fig. 4.5) was used, and fecundity was set to $F = 100$.

Note that the temperature range we are using represents mild winters (i.e. $\lambda_1 > 1$). We choose not performing

simulation with lower temperature regimes given the high uncertainty of egg survival for temperatures lower than 15°C .

We also explored the sensitivity of the system to initial conditions by simulating 1) different initial substage proportions for a constant total population $N_0 = 10^7$, 2) different orders of magnitude of N_0 , ranging from 10^2 to 10^{10} , and 3) different degrees of density dependence with $\{\omega \in 10^{-15}, \dots, 10^5\}$. Monte Carlo simulation indicated (figures not shown) that the system does not exhibit chaotic or near chaotic behaviour (by chaotic we mean a system that features sensitivity to small changes in system states and parameters). A consistent result was that in all cases, by the final year (year 5), populations showed signs of convergence (similarly to figure 4.2). These results are preliminary and more detailed analyses should be carried in order to explore more thoroughly the parameter space of N_0 and ω .

Given the non-linearity of parameters (w.r.t. temperature) and the density-dependence, it could be expected that the system would be a candidate to exhibit some chaotic behavior (May *et al.*, 1976; Cushing *et al.*, 2002). Nonetheless, our exploratory analysis suggests the opposite. This is perhaps due to the seasonal temperature profile regulating the system, damping out any asynchronised transients, although such aspects of the model have not been addressed further in this work. Alternatively, the possibilities of bifurcations with increased fecundity (as in May *et al.* (1976)) have not been explored. Interestingly, exploring the system at fixed temperatures (figures not shown) confirmed that the observed cyclicity is driven by the seasonality without which stage-distributions stabilise and oscillations cease. Thus, increasing the strength of this seasonality would be expected to further accentuate the oscillations displayed by the deterministic dynamics.

Stochastic environment. In a second exploratory analysis, we included environmental noise to the density-dependent *Culicoides* IPLM. Scenarios with and without environmental noise are considered. When environmental noise was included, its standard deviation was set to $\sigma_{\epsilon} = 0.25$. In both cases, density-dependence was fixed with $\omega = 10^{-9}$. Again, first the median (from 1000 MCMC samples) of $\{\mu_S, \kappa_S, \nu_S\}$ for each stage S was used (continuous black lines in Fig. 4.5), and fecundity was set to a constant ($F = 100$). Monte Carlo (MC) methods (10^4 simulations) provided estimates of the distribution of daily adult abundance. In each simulation, $N_0 = 10^7$ and individuals were initially distributed among all substages draw from a Dirichlet distribution with parameters as defined above.

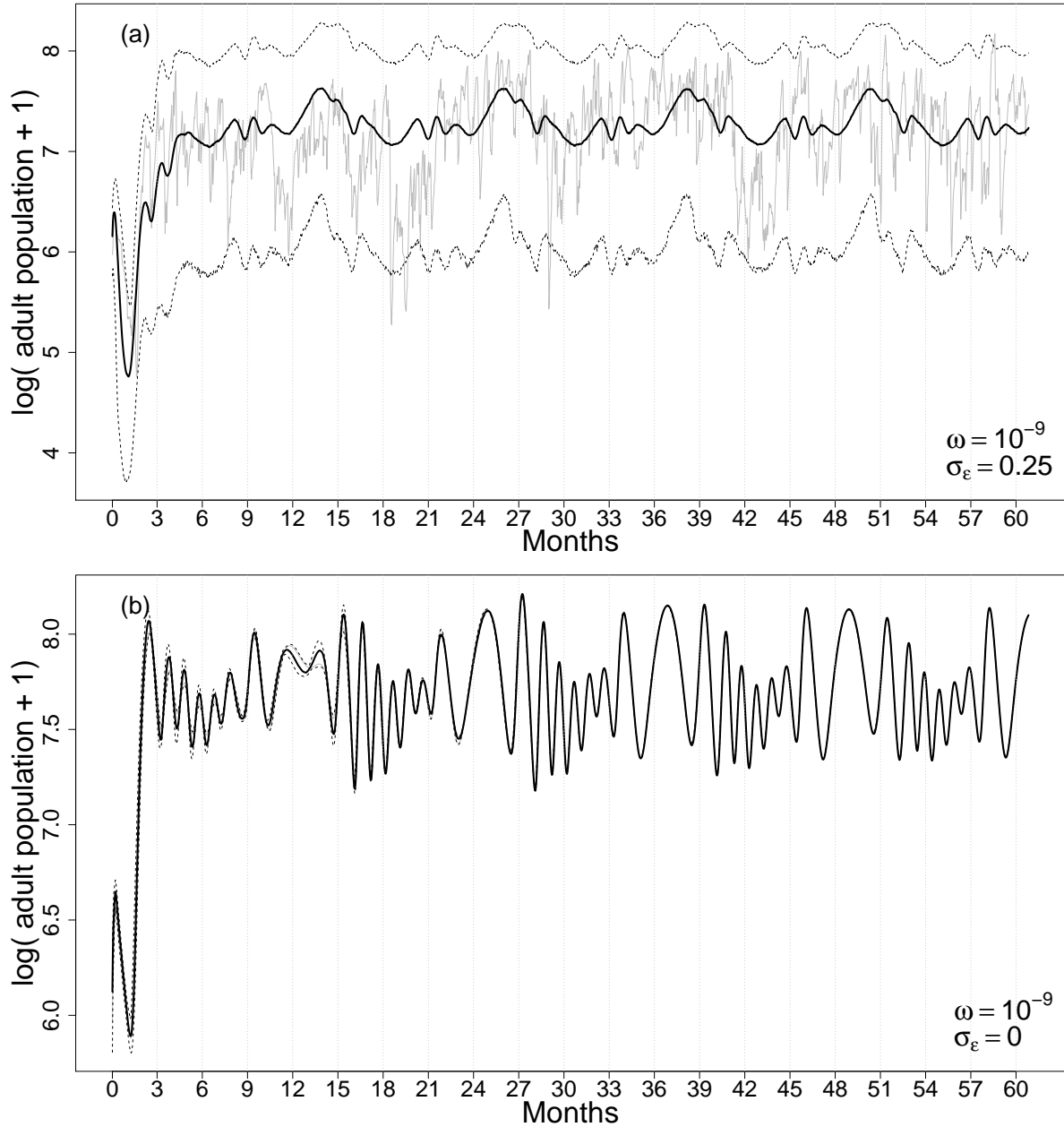


Figure 4.3. Simulated adult *Culicoides* time series of an IPLM model including density-dependence larval survival ($\omega = 10^{-9}$). Daily temperature varies following a (cosine) profile of period one year, and affects *Culicoides* life-cycle parameters according to estimates from chapter 3. Here, the median of life-cycle parameters (bold line of Fig. 4.5) was used, and fecundity was set to $F = 100$. In (a), environmental noise ϵ_t is included in the model ($\epsilon_t \sim \text{Normal}(0, \sigma_\epsilon)$, with $\sigma_\epsilon = 0.25$). In (b), the system is deterministic (thus no environmental noise). The continuous line shows the median of a Monte Carlo estimation (10^4 simulations) of log adult abundance, while dashed lines refer to 1% and 99% quantiles. In case (b), slight differences in trajectories arise from different initial Dirichlet random allocations of $N_0 = 10^7$ individuals among all substages. The trajectory of a single simulation is shown in gray in both plots.

Figure 4.3 summarises the MC results. The median (bold line) and $\{1\%, 99\%\}$ percentiles (dashed lines) of

the distribution of daily adult densities, as well as the output of a single simulation (gray line) are shown for both stochastic (top) and deterministic (bottom) scenarios. Comparing the two medians (bold line) indicates that environmental noise produces a general reduction of the adult population (a direct consequence of environmental noise increasing the probability of lower larval survival) and blurs the oscillations and transients generated by density dependence. For example, the median of the deterministic output displays eight peaks per year, this is reduced to six in the median of the stochastic system's output. By contrast, the number of peaks in a single simulation is greatly augmented by the stochastic survival term. Although the amplitude of oscillations in the median is diminished, noise increases notably the day-to-day variance in adult density, as indicated by the dramatic expansion of the 1% and 99% percentiles of trajectories (dashed lines) or the jagged profile of a single trajectory (gray line). Analogous patterns in the whole behaviour were obtained when varying N_0 across different orders of magnitude explored ($N_0 \in \{10^2, \dots, 10^{10}\}$, results not shown).

We repeated the above analysis, i.e. simulating adult abundance from the density-dependent IPLM, with and without environmental noise by including, this time, the seasonal uncertainty associated to estimates of *Culicoides* life-cycle parameters and fecundity (1000 MCMC samples) (Fig. 4.5). Results (Fig. 4.4) show that the added uncertainty decreases the range of oscillation in the median of the trajectories and that, in general, noise augmenting the probability of lower larval survival greatly increases the probability of population crashes that could lead to local extinction. That this feature was not detected in Fig. 4.3 reinforces the importance of accurately assessing parameter uncertainty when making assessments of extinction risk for conservation purposes. Moreover, the contrast between the two plots of Fig. 4.4 highlights the importance of correctly identifying the level of environmental stochasticity in a system. In fact, turning this argument on its head, Fig. 4.4 suggests time series data can provide a rich data source for estimating the degree of environmental stochasticity faced by field populations. In other words, parameters related to this stochasticity should be identifiable from data if correct care with parameterisation is taken.

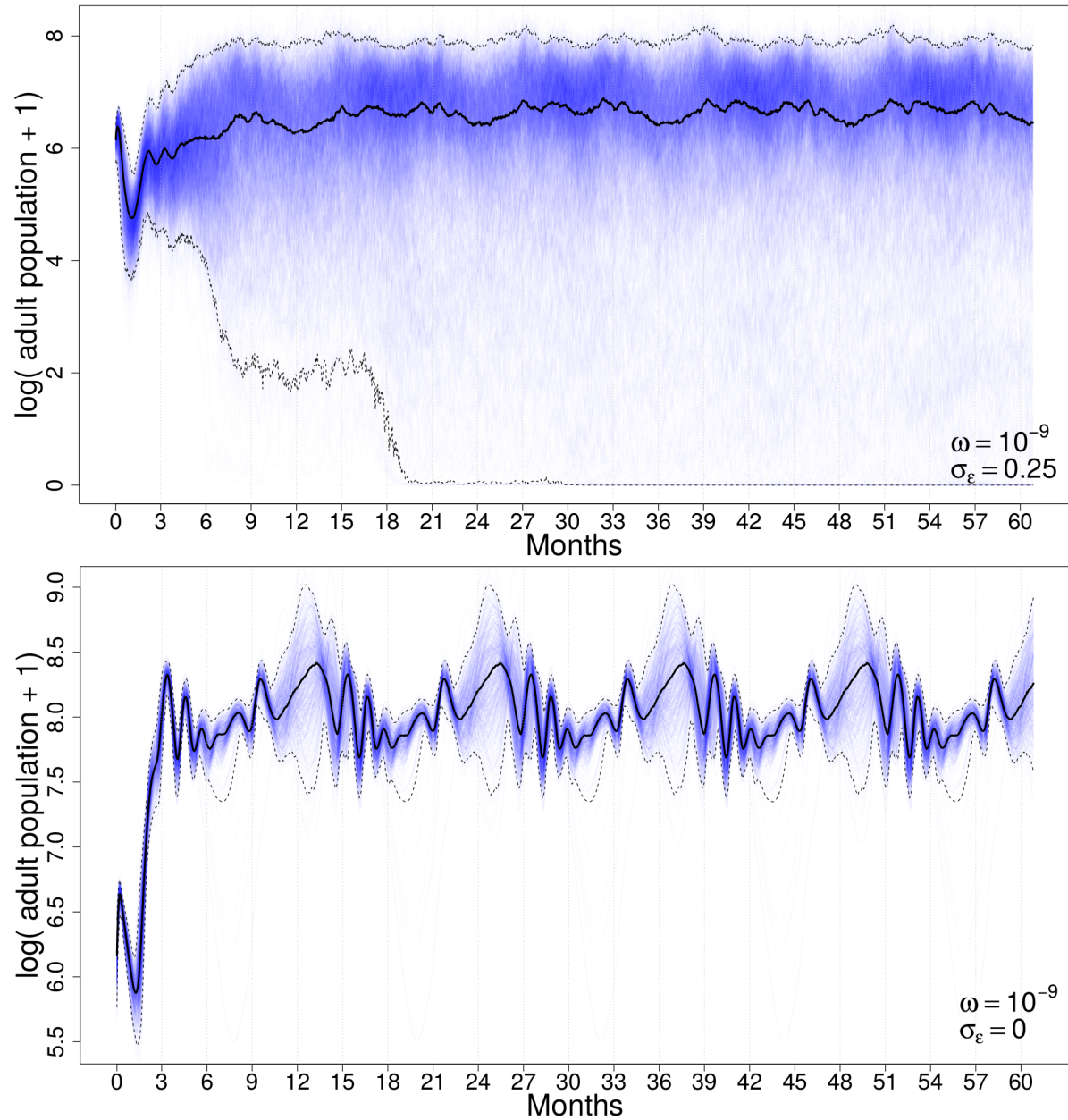


Figure 4.4. Simulated time series of adult *Culicoides* abundance generated from an IPLM model including in density-dependent larval survival ($\omega = 10^{-9}$) and environmental stochasticity in survival (top only). Daily temperatures follow a cosine profile of period one year through the five years shown, and affects *Culicoides* life cycle parameters according to interpolated estimates (1000 MCMC samples) from chapter 3 (see Fig. 4.5). On top, environmental noise ϵ_t is included in the model ($\epsilon_t \sim \text{Normal}(0, \sigma_{\epsilon})$, with $\sigma_{\epsilon} = 0.25$). At the bottom, the system is deterministic (no environmental noise). Continuous line shows the median of the Monte Carlo estimates (10^3 simulations) of log adult abundance, while dashed lines refer to 1% and 99% quantiles.

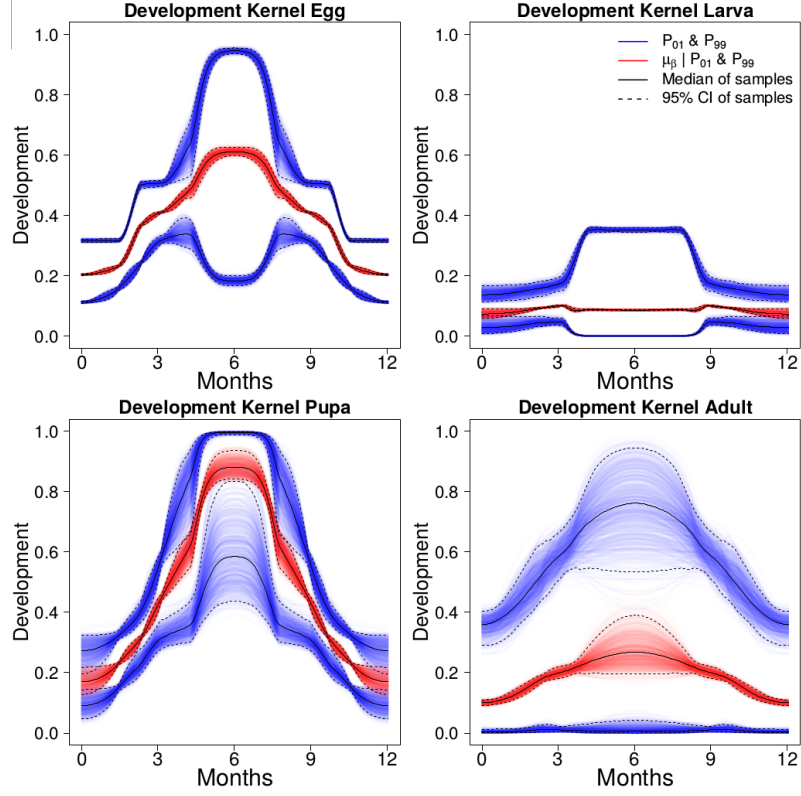


Figure 4.5. Annual non-linear developmental responses to temperature at the maximum *a posteriori* (MAP) resolutions ($r_E^{MAP} = 6$, $r_L^{MAP} = 9$, $r_P^{MAP} = 31$, $r_G^{MAP} = 8$) for egg, larvae, pupae and adult midges. Results from 1000 MCMC samples are plotted with unimodal spline interpolation. Red and blue lines show median and 1% tail percentiles ($\mathcal{P}_{01}, \mathcal{P}_{99}$) of development kernel $f(\Delta|\theta)$. Median (black line) and 95% credibility intervals (CI) for each parameter are shown (dashed lines). Seasonality in fecundity is not shown but is derived, as in this case, from results giving Fig. 3.5.

These preliminary analyses provide useful insight when designing methodological strategies for estimating parameters of stochastic density-dependent IPLMs for observed time series. Our aim is to have methods to estimate density-dependence and stochasticity for the basic IPLM using field data. This requires that we confront the process model with data for inference and prediction, which brings us naturally to the *state-space model* framework.

4.2.2 Parameter estimation with IPLM state-space models

In this section, we perform a preliminary analysis using synthetic likelihood (see Appendix C.1) methods (Wood, 2010) within an MCMC algorithm for inference with an IPLM state-space model (SSM). The process model of our SSM is given by the noisy density-dependent IPLM described above (expressions 4.4, 4.5 and 4.6). For the

observation model, we suppose that the number of adult individuals y_t observed at every time step t is a Poisson random variable with mean (and variance) ϕ , and distribution function

$$f_Y(y_t = n) = \frac{e^{-\phi} \phi^n}{n!}, \quad n = 0, 1, 2, \dots \quad (4.8)$$

Weekly observations are assumed (i.e. the observation expectancy is zero for six days of the week and the resulting “zero observations” are ignored). Noisy observations are modelled with an auxiliary parameter ϕ'_t such that

$$\begin{cases} \text{logit}(\phi'_t) = \text{logit}(\phi) + \epsilon_t^{\text{obs}} \\ \epsilon_t^{\text{obs}} \sim \text{N}(0, \sigma_{\text{obs}}). \end{cases} \quad (4.9)$$

Thus, the IPLM state-space model can be expressed as

$$\begin{cases} \mathbf{x}_{t+1} = M(\mathbf{x}_t, \Theta_t) \mathbf{x}_t \\ y_t \sim \text{Poisson}(\phi'_t \mathbf{A}_t), \end{cases} \quad (4.10)$$

where $M(x_t, \Theta_t)$ and x_t are the matrix and vector from expression (4.4), and Θ_t indicates the time-dependent set of parameters defining the process model and \mathbf{A}_t is the total adult density density at time t .

Methods.

In order to explore the performance of parameter estimation with synthetic likelihood (SL), we simulated weekly adult density data across three years. *True* model parameters were set to $N_0 = 10^8$, $\omega = 10^{-9}$, $\sigma_\epsilon = 0.5$ and $\text{logit}(\phi) = -12 + \epsilon_t^{\text{obs}}$, with $\sigma_{\text{obs}} = 1$; while stage-specific temperature-dependent parameters $\{\mu, \kappa, \nu\}$ and fecundity f were sampled taken from a single MCMC obtained in chapter 3. For simulations, the process model was updated each day and observations were generated each seven days.

Two parameter estimation scenarios were performed using MCMC with synthetic likelihood (see Appendix C.1.1): **(a)** the expected proportion of observed adults was assumed known (i.e. $\text{logit}(\phi) = -12$), and the parameter set $\{N_0, \omega, \sigma_\epsilon\}$ was estimated; and **(b)** the parameter set $\{N_0, \omega, \sigma_\epsilon, \phi\}$ was estimated. During MCMC, both N_0 and ω were sampled on the log scale; ϕ on the logit scale; and a change-of-variables correction was used to remove bias in the priors associated to these transformations. For both cases, priors (and hyperpriors) of the model were

$$\begin{cases} \log(N_0) & \sim N(7, 1.5) \\ \log(\omega) & \sim N(-9, 3) \\ \sigma_\epsilon & \sim \text{Cauchy}_0(0, 10) \\ \sigma_{\text{obs}} & = 1 \end{cases} \quad (4.11)$$

where Cauchy_0 is the zero-truncated Cauchy distribution with mean 0 and scale 10, while for case **(b)**, the additional prior $\text{logit}(\phi) \sim N(-12, 0.5)$ was used. In both scenarios, the basic IPLM parameters that were estimated in chapter 3 were assumed known and were fixed at their true values used for data simulation.

Target posteriors distributions for cases **(a)** and **(b)** were sampled using a block Metropolis-Hastings sampler in NIMBLE. This was done using the nimble function `RW_llFunction_block_sampler` in order to incorporate the SL (see manual NIMBLE Development Team (2016)). The numbers of simulations n_s used within the SL for MCMC runs were $n_s = \{25, 50, 100\}$. For $n_s \in \{25, 50\}$, 45 different datasets for the “true” data were explored, while 13 were explored for $n_s = 100$. In all cases, a total of 5000 burn-in samples plus 5000 post burn-in samples was generated for each MCMC. Convergence diagnostics were performed using CODA (Plummer *et al.*, 2006).

Results. The trajectories and autocorrelation from three typical MCMC runs with $n_s = 25$, $n_s = 50$ and $n_s = 100$ for case **(a)** are shown in Fig. 4.6. Chains showed a relatively rapid convergence to the true parameter value (red lines) and better mixing (and reduced autocorrelation) for either $n_s = 50$ or $n_s = 100$ than for $n_s = 25$. This similarity between the results when $n_s = 50$ or $n_s = 100$ is also observed in Fig. 4.7, which compares the true adult time series to the median and 95% CI of replicated data generated from the model during MCMC. The median of the replicated time series shows a very good fit to data in all cases, and uncertainty was reduced drastically when the number of replicates was raised from $n_s = 25$ to $n_s = 50$, while very similar performance was obtained with $n_s = 100$. We compared posteriors from MCMCs generated with different “true” datasets for $n_s = 50$ and $n_s = 100$ (Fig. 4.8). These results confirm that similar identifiability is attained with the less computational demanding $n_s = 50$ case. Moreover, posteriors are highly different to the weakly informative priors (expression 4.11), suggesting that the “true” simulated data \mathbf{y}_t were highly informative regarding estimation of N_0 , ω and σ_ϵ .

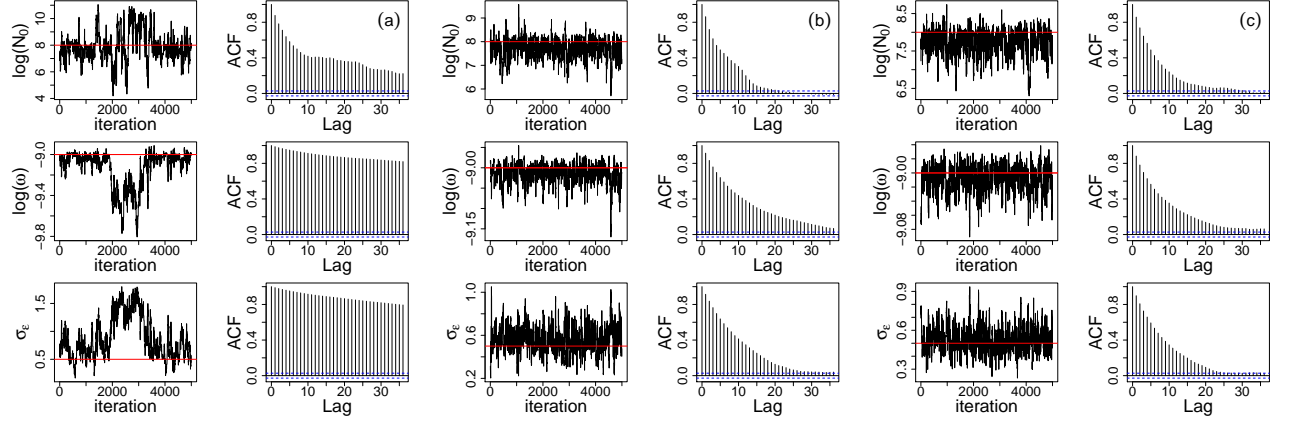


Figure 4.6. Trajectories and autocorrelation of parameter estimates from a typical MCMC run for case (a), i.e. $\{N_0, \omega, \sigma_\epsilon\}$ estimated; ϕ known. The number of replicates used for the vector of summary statistics were $n_s = 25$ (left), $n_s = 50$ (center) and $n_s = 1000$ (right). In all cases, the same “true data” was used to fit the model. Trajectories of replicated data from these three runs are shown in Fig. 4.7.

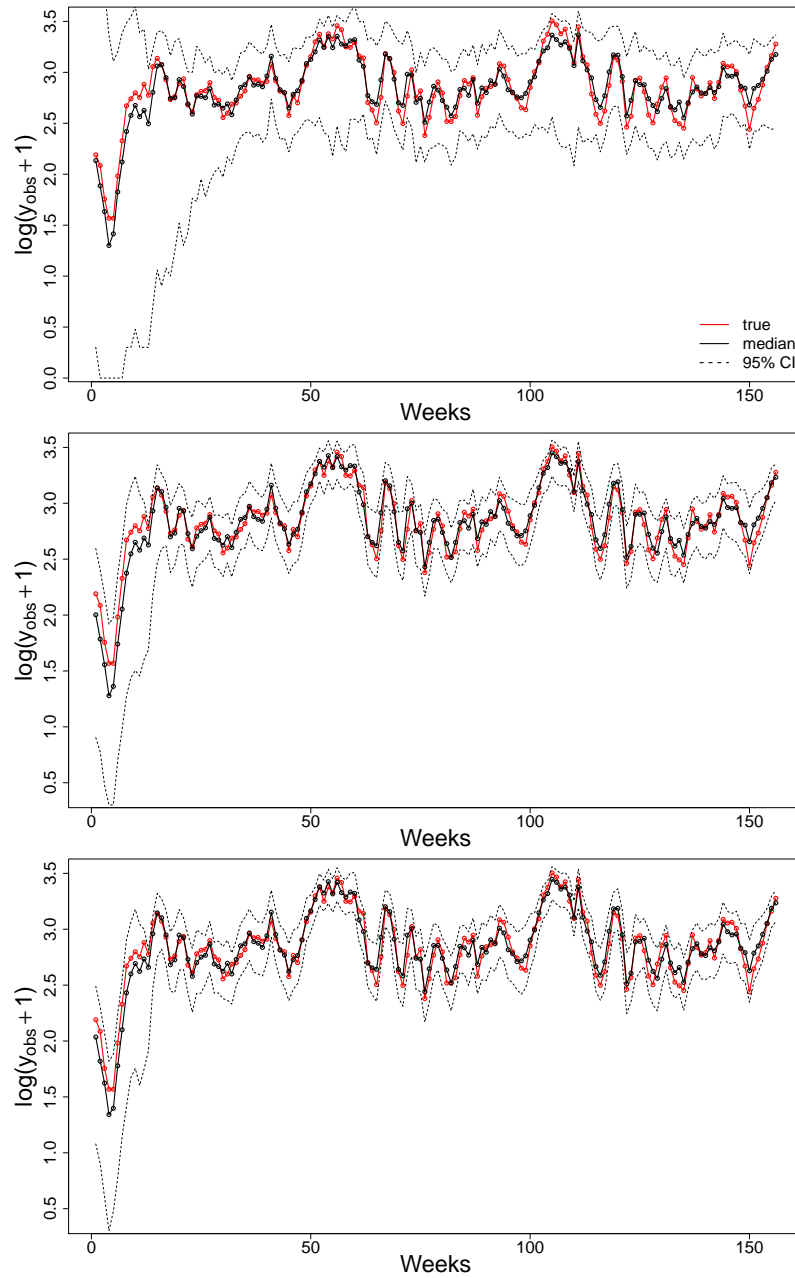


Figure 4.7. Comparing MCMC realisations with different number of simulations in the synthetic likelihood ($n_s = 25$ (a), $n_s = 50$ (b) and $n_s = 100$ (c)) for the same “true” weekly adult abundance observation data, \mathbf{y}_{obs} . The “true” observation data (red line), the median (black) and 95% CI (dashed) of adult trajectories \mathbf{y}_T from 5000 post burn-in MCMCs are shown. Plots (a), (b) and (c) correspond to the same MCMC runs as in Fig. 4.6.

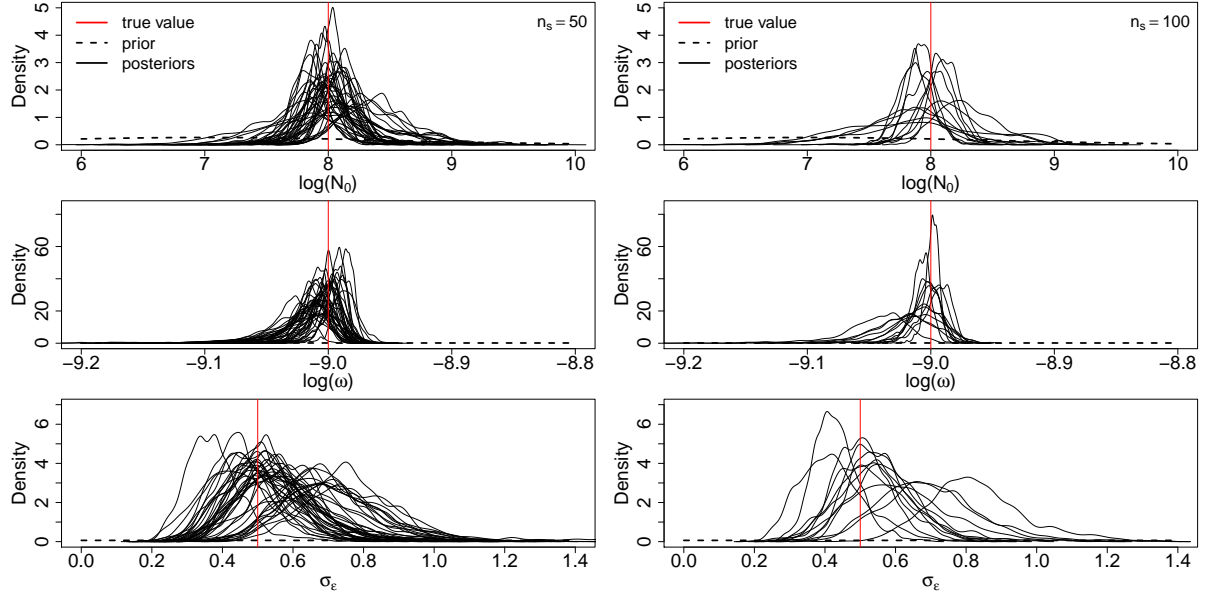


Figure 4.8. Comparing posterior densities (solid black) of parameters on their (sampling) scale when transformation was used to associated prior distributions (dashed), and true values (red), for $n_s = 50$ (left) and $n_s = 100$ (right). Densities were obtained from 5000 post burn-in MCMC samples, and 45 different “true” datasets were used for $n_s = 50$, while 13 “true” datasets were used for $n_s = 100$.

For case **(b)**, i.e. when ϕ was included in the set of parameters to be estimated, chains showed poor mixing regardless of the n_s used to calculate the SL (Fig. 4.9). Very strong correlations were observed between ϕ and ω (on their transformed scales), suggesting that the true parameters might not be identifiable given data \mathbf{y} and that developing efficient proposals for MCMC could be challenging.

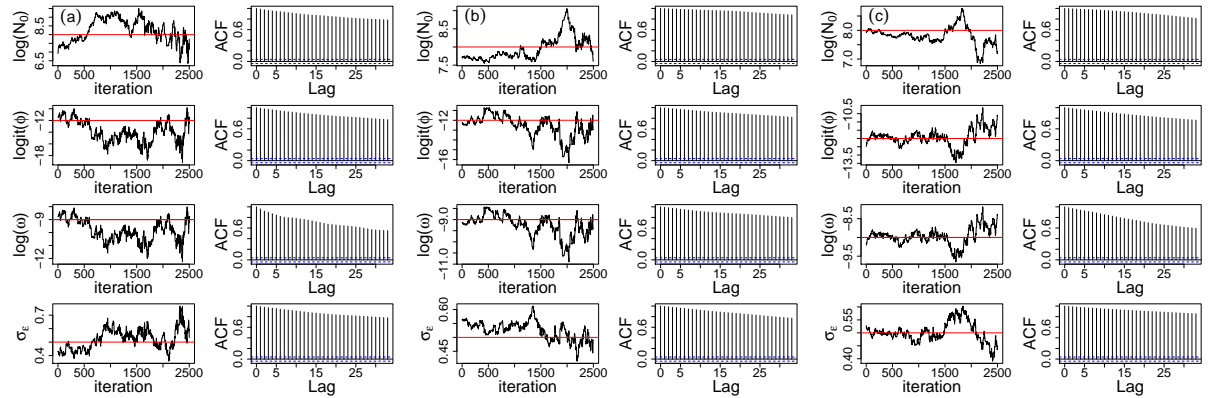


Figure 4.9. Trajectories and autocorrelation of parameter estimates from a typical MCMC run for case **(b)**, i.e. $\{N_0, \omega, \sigma_\epsilon, \phi\}$ estimated. Number of replicates used for the vector of summary statistics were $n_s = 25$ (a), $n_s = 50$ (b) and $n_s = 1000$ (c). In all cases, the same “true data” was used to fit the model.

Discussion. Here we have shown how IPLMs can be extended to better characterise the dynamics of stage-structured populations in natural (i.e. field) conditions. Including the IPLM into the state-space framework to study and analyse time-series of ongoing population dynamics increases the complexity of the modelling challenge. While the intractability in resulting likelihoods hampers the performance of likelihood-based model inference, we have shown that a recent simulation-based approach – the synthetic likelihood (SL) – offers a promising alternative.

With simulated multi-annual time-series of weekly adult flight-trap data, we have explored the identifiability of key cryptic parameters *in natura* (environmental stochasticity in mortality, density-dependence, initial population density and expected efficiency of flight-traps). We have shown that when flight-trap efficiency and σ_{obs} are known, the other parameters can be identified to a high level of precision with reasonable computational demand (in terms of the number of replicates the SL uses). If the efficiency of the observation process is erroneously assumed known then bias in the estimates of ω (and other correlated parameters) can be expected. When trying to estimate flight-trap efficiency ϕ , very strong correlation with the density-dependence parameter ω hampers the ability of the sampler to adapt, suggesting that additional data sources are required to reduce uncertainty in ϕ and to calibrate the model for epidemiological purposes

More extensive simulation and estimation scenarios, and sampling strategies, should be explored. For example, in these preliminary studies, several parameters were fixed at their true values. In the analysis of field data we do not have this luxury, and so the number of parameters to be estimated must be expanded in order to be aware of potential problems or limitations when working with field data. This risks placing extra-demand on the set of summary statistics, which may need to be expanded in order for parameters to be identifiable.

Identifying the most informative combinations of summary statistics (and distance metrics) involved in SL calculation via more systematic methods (Scranton *et al.*, 2014), could improve estimates with lesser computational efforts. Also, although not included in this study, random variables to correlate population dynamics – which were explored via the simulation study of chapter 2 – could be used to include different genotypes, group-level dynamics determined by conditioning phenotypic responses of sub-populations (to possibly unmeasured) local factors, or other random effects affecting a particular population in question.

A further avenue that should be explored is to identify how parameter uncertainty increases and identifiability decreases as noise in the observation model increases and the level of information in the observations decreases. It may prove challenging to discriminate noise in the observation process and in the process model. Similarly, how

the frequency of making observations affects estimations should also be explored in order to guide surveillance programs about optimal sampling strategies.

Here, a relatively mild seasonality was used. This was due to large uncertainties outside the chosen temperature range. To reduce these uncertainties one would ideally perform temperature transfer experiments (Régnière *et al.*, 2012) and adapt techniques in chapter 3 to obtain parameters across large temperature ranges. Alternatively, however, it could also be possible to reduce uncertainties in $\{\mu, \kappa, \nu\}$ at low temperatures using techniques developed in this chapter.

The results of simulation suggest the dynamics generated from a combination of strong seasonality and density dependence could be a relatively rich source of information that the IPLM framework is well adapted to exploit. This possibility is clearly an exciting avenue for further research.

4.2.3 Conclusion and perspectives

The ultimate aim guiding this prospective analysis was to explore methodological strategies that enable Bayesian estimates for the stochastic density-dependent IPLM when including field (time series) data. For this, we have extended the IPLM model to include key aspects of observing process of natural systems, e.g. regulatory mechanisms and sources of stochasticity in both the population dynamics and data sampling.

Promising results with simulated data from our IPLM state-space model coupled with synthetic likelihood based MCMC set the basis for further work that provide the methodological tools for combining survey (i.e. census) data and individual-based demographic data for more detailed inference about population dynamics. To conclude, our preliminary study suggests that synthetic likelihood based MCMC inference of the IPLM state-space model shows great promise as a framework for parameter estimation and model inference of realistic stage-structured population models. Further research is required to extend this preliminary study and to graduate from analysing simulated to real datasets.

Chapter 5

General conclusion

This thesis has introduced the integral projection Lefkovitch matrix (IPLM) model, a new matrix-based approach that is expected to improve the current predictive performance of models for biological stage-structured populations (SSPs). The new framework incorporates vital rate heterogeneity at the level of within-stage developmental; is flexible enough to include complexities encountered by natural populations such as non-linear responses to time-varying covariates, density-dependence or environmental stochasticity; and enables the use of different types of data for estimation, inference and simulation. Moreover, the ability to fit IPLMs with individual-level fixed-traits, or group-level random effects, provides a potentially powerful tool for confronting theoretical ecological, or even eco-evolutionary, models to data.

By comparing outputs from the fitted *Culicoides* IPLM to those from its “classic” Lefkovitch matrix model (CLM) counterpart, we have demonstrated the effects of assumptions regarding variance in stage durations (and the processes that generate that variation) in common metrics of asymptotic and transient dynamics. The oversimplified assumptions of CLMs translate into erroneous estimation of the amplitude and duration of transient dynamics, or the range of temperatures predicting population growth. The importance of this effect is exacerbated in fluctuating environment such as those studied here because perturbations are inherent features of these systems. The effects of these assumptions can have implications regarding potential ecological niche predictions from mechanistic niche models and, in a more broad and applied sense, can have serious consequences in wildlife management, conservation and pest / vector control.

Many phenological models rely on an assumed linearity between cumulated degree-days and development.

By modeling the effects of heat energy accumulation via a kernel that tracks within-stage development in a Bayesian framework, our stage-specific IPMs couple theoretical ideas (Plant and Wilson, 1986) and previous biologically limited approaches (Dennis *et al.*, 1986) with established statistical techniques to provide more informative estimates of how developmental rates vary in response to covariates whilst providing more informative assessments of uncertainty than is encountered traditionally. For example, we used the results of our Bayesian model fitting to test the linearity assumption of classic degree-day models, showing that linearity only provides a robust developmental model over extremely narrow temperature intervals. A further limitations of most degree-day based models is that they usually do not account for diapause (a biological mechanism of reducing activity in the face of adversity to maximise survival), which is observed in some *Culicoides* species and other arthropods. Extending IPLMs to account for key covariates other than temperature (e.g. photoperiod or soil moisture) driving diapause could provide a more realistic model for the study of phenology, although diapause data is scarce for most arthropods of epidemiological interest (Isaev, 1974, 1976; Faraji and Gaugler, 2015).

The emergence and resurgence of vector-borne diseases represents a global public health issue. Improving models to enable inference from vector surveillance data is a necessary step towards advancing our knowledge of vector ecology and the epidemiology of the pathogens they vector. In the case of our motivating example, despite the epidemiological risk posed by biting midges, their life-cycle parameters are poorly documented and under-exploited. Modeling has prioritised empirical analyses of flight-trap data and, as far as we know, to date no stage-structured models have been developed to predict the responses of *Culicoides* populations (and their associated diseases) to global change. We have demonstrated a potential application of the *Culicoides* IPLM model by analysing the effects of a hypothetical adulticide-based vector control scenario on the temporal variation of the basic reproduction ratio (R_0) of BT disease. Although quite simple, our analysis shows the potential utility of including temperature-dependent generation times. For example, the approach can reduce error associated with epidemiological predictions, highlight the transient nature of R_0 dynamics in response to control intervention and, with appropriate data, can be easily extended to help guide effective vector management schemes incorporating other integrated vector management strategies (e.g. larvicide or vaccination).

A popular use of matrix models is to perform sensitivity analysis, which permits the exploration of effects of parameter perturbation (Caswell, 2006). In the IPLM framework this would enable the evaluation of the relative importance of various parameters across the life cycle. This can potentially help to determine key parameters for

vector control. Since these developments follow standard methods, we did not explore this potential use of IPLM models. Moreover, we have frequently used Bayesian methods to indicate the possible distribution of indices, such as λ_1 , given uncertainty in IPLM parameters. However, future work should permit that sensitivities and elasticities can be readily obtained from IPLM models by users.

Another interesting potential study not addressed in this work is comparing the predictive performance of our models to those of purely statistical approaches. This is particularly important in the light of the recent rise of machine learning and the ensuing controversy regarding whether machine learning tools can outperform true models on predictive tests (Perretti *et al.*, 2013; Hartig and Dormann, 2013). Given the strong determinism of environmental factors in the generation times and vital rates of poikilothermic organisms, our working philosophy has been one of improving the resolution of existing “mechanistic” models, and incorporating them into a state-space framework, in order to maximise the value of our knowledge regarding prediction. Presumably, if machine learning tools outperform IPLM predictions then there are holes in our ecological understanding that would need to be addressed. Comparing the predictions of our mechanistically informed models to those of mechanistically naive models could provide valuable information for guiding future research directions.

A perennial goal of ecology is the establishment of methods that realistically scale-up from laboratory to field conditions. Yet, it is well known that complex population models are challenging to fit to real data for many reasons. We have extended the basic IPLM framework to include key aspects relating to observed stage-structured dynamics in field conditions, e.g. regulatory mechanisms and sources of stochasticity in both the population dynamics and data sampling. We have taken preliminary steps to explore the use of cutting edge Bayesian techniques for fitting IPLM models to partially observed (i.e. adult only) time series typical of the data obtained from vector surveillance studies. Promising results with simulated data from our IPLM state-space model coupled with synthetic likelihood based MCMC need to be further extended to better identify the balance between biological detail, computation time, and predictive power that current data and computer power support.

Our long-term aim is to develop robust statistical methods that link IPLM models to field data where populations (with overlapping generations) are partially observed. This would allow combining survey (i.e. census) data and individual-based demographic data for more detailed inference about population dynamics. Yet – and taking *Culicoides* as a paradigmatic example – there are considerable technical challenges for various reasons. For example, the dynamics of juvenile stages are rarely the subject of surveillance studies in natural conditions,

primarily due to their small size and related difficulties sampling their populations. Most of the data pertinent to modeling the juvenile stages of *Culicoides* comes from laboratory studies on species that are not necessarily the species of key epidemiological importance – there is therefore uncertainty regarding how pertinent parameters gleaned from laboratory studies are for parameterising models of dynamics *in natura*. Thus modellers need to be able to use surveillance data to adjust parameters that have been obtained from laboratory studies in order to achieve greater realism – but adding this flexibility comes at the cost of parsimony. Moreover, most surveillance studies of *Culicoides* rely on trapping flight-active adults, which is known to provide a biased representation of the true abundance of active adults. For example, traps are often located close to livestock so the distribution at other locations can be left unsampled. Furthermore, it is known that large between-species and between-sex differences exist in how individual insects respond to various trap types or different hosts (Viennet *et al.*, 2013). Nonetheless, these data sources are often the best indicators available, and developing methods that enable extracting as much biological information as the data can support is a well known goal in ecological modelling.

Ecologists have always been fascinated by the sources of cyclicity in natural populations. Here we have shown how seasonal shifts in generation time, stage distributions and density dependence naturally lead to multiple population cycles within a given year. It is not difficult to conjecture that models which remain overly simplistic regarding the links between within-stage development and environmental fluctuation have poor hope of characterising such cyclicity. Moreover, the lack of cyclicity under constant temperatures suggests there remain other bifurcations associated with seasonality that have yet to be identified and mapped. In ideal scenarios, the information in these patterns should be able to help model selection. However, we have also seen how environmental noise can blur the characteristic deterministic cyclicity of density-dependent IPLMs, which can be expected to make model selection a more challenging problem. Moreover, it is still not known if, and under what conditions / parameterisations, an IPLM model might demonstrate bifurcations leading to chaotic behaviour, and it is likely that discriminating between chaos and the effects of fluctuating climate in real data could be difficult. In conclusion, many challenges remain to be addressed regarding the analysis of IPLMs.

The key contribution of this work is that the methods explored here can be adapted to study the life-cycles of many organisms. This is particularly powerful for species with instar or egg stages, for which typical developmental states such as size or weight provide poor or unavailable predictors for the underlying dynamics. The generality of IPLMs – including the ability to fit IPLMs for many taxa not previously studied with IPMs –

suggests that many applications are possible, including mechanistic niche modelling, demographic compensation analysis, eco-evolutionary and ecological forecasting for conservation, agricultural and epidemiological purposes.

Appendix A

Appendix to Chapter 2

A.1 Discretising within-stage IPLM

The main text describes an integral projection model (IPM) for characterising the dynamics of within-stage development δ (equation 2.3). In this model, an individual's developmental status δ progresses via increments $\Delta = \delta' - \delta$ with PDF $f(\Delta|\theta)$, CDF $F(\Delta|\theta)$ and parameters θ . For parsimony, we assume both $f(\Delta|\theta)$ and survival ν are independent of δ . The IPM-kernel is therefore $K_{\Theta}(\Delta) = \nu f(\Delta|\theta)$, where the PDF $f(\Delta|\theta)$ accounts for developmental rate heterogeneity. Throughout this work, we assume $f(\Delta|\theta)$ is the PDF of a beta distribution with mean (i.e. expected developmental increment) μ and variance $\kappa\mu(1 - \mu)$ (parameters defined in the main text). The full parameter set for a given stage is thus $\Theta = \{\nu, \theta\} = \{\nu, \mu, \kappa\}$.

A matrix approximation of this IPM for within-stage development is obtained by discretising the within-stage developmental status δ into r equally sized substages. We assume individuals always start a new stage in the first substage and develop by increments of up to r substages each time-step. Within-stage transition probabilities for discrete developmental increments of l substages are given by

$$p_l = \int_{\frac{l}{r+1}}^{\frac{l+1}{r+1}} f(\Delta|\theta) d\Delta, \quad (\text{A.1})$$

where $p_r = 1 - \sum_{l=0}^{r-1} p_l$ gives the probability to complete the entire stage in just one time-step. These probabilities are used to define the projection matrix \mathbf{W}_S for each stage S (equation 2.4 in main text). The full parameter set for a

given stage in this discretised IPM-approximation is $\{\nu, \mu, \kappa, r\}$, where r can be treated either as a computational parameter to fix or a parameter to estimate. In the latter case, sensitivity to r implies r functions as a shape parameter, in which case alternative more flexible distributions for $f(\Delta|\theta)$ might be explored.

A.2 Quality-dependent development with Gaussian copulas

Copulas are tools for generating multivariate distributions from an arbitrary set of marginal distributions (Kruskal, 1958; Nelsen, 2006). In the simulation-estimation study of chapter 2, Gaussian copulas were used to model correlation between individual quality q and development increments Δ . Here, we outline the details required to condition development kernels on individual quality. We assume q is fixed through an individual's lifespan and conditions the distribution of increments at each time-step.

Marginal distributions of both Δ and q were described by beta distributions with densities

$$f_{\Delta}(\Delta|\alpha_1, \alpha_2) = \frac{\Delta^{\alpha_1-1}(1-\Delta)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)}, \quad (\text{A.2})$$

$$f_q(q|\xi_1, \xi_2) = \frac{q^{\xi_1-1}(1-q)^{\xi_2-1}}{B(\xi_1, \xi_2)}, \quad (\text{A.3})$$

and with corresponding cumulative distribution functions F_q and F_{Δ} . A standard uniform distribution was obtained for q by setting $\xi_1 = \xi_2 = 1$.

Correlation between Δ and q is established via random variables x_{Δ} and x_q , which have a bi-variate Gaussian density with correlation coefficient ρ and standard normal marginal densities $\phi(\cdot)$ with distribution functions $\Phi(\cdot)$. These variables are linked via the following probability integral transformations:

$$F_q(q) = u_q = \Phi(x_q), \quad (\text{A.4})$$

$$F_{\Delta}(\Delta) = u_{\Delta} = \Phi(x_{\Delta}), \quad (\text{A.5})$$

where u_q and u_{Δ} follow standard uniform distributions. The joint density of Δ and q is therefore

$$f_{\Delta,q}(\Delta, q | \alpha_1, \alpha_2, \xi_1, \xi_2, \rho) = f_{x_\Delta, x_q}(x_\Delta, x_q | \alpha_1, \alpha_2, \xi_1, \xi_2, \rho) \left| \begin{array}{cc} \frac{\partial \Delta}{\partial x_\Delta} & \frac{\partial \Delta}{\partial x_q} \\ \frac{\partial q}{\partial x_\Delta} & \frac{\partial q}{\partial x_q} \end{array} \right|^{-1}, \quad (\text{A.6})$$

where $f_{x_\Delta, x_q}(\cdot, \cdot | \alpha_1, \alpha_2, \xi_1, \xi_2, \rho)$ is the bi-variate normal density with correlation parameter ρ and the Jacobian determinant provides a change of variables correction for the transformations.

The above specification gives the following conditional density of Δ given q :

$$f_{\Delta|q}(\Delta | \alpha_1, \alpha_2, q, \rho) = \frac{f_{x_\Delta|x_q}(x_\Delta | \alpha_1, \alpha_2, x_q, \rho) f_\Delta(\Delta)}{\phi(x_\Delta)}, \quad (\text{A.7})$$

where $f_{x_\Delta|x_q}(x_\Delta | \alpha_1, \alpha_2, x_q, \rho)$ is normal with mean $\mu_\Delta = \rho x_q$ and variance $\sigma_\Delta^2 = 1 - \rho^2$. The corresponding conditional distribution function is given by the identity

$$F_{\Delta|q}(\Delta | \alpha_1, \alpha_2, q, \rho) = \int_0^\Delta f_{\Delta|q}(y | \alpha_1, \alpha_2, q, \rho) dy = \int_{-\infty}^{x_\Delta} f_{x_\Delta|x_q}(z | \alpha_1, \alpha_2, x_q, \rho) dz = F_{x_\Delta|x_q}(x_\Delta | \alpha_1, \alpha_2, x_q, \rho). \quad (\text{A.8})$$

A.3 Generating sojourn-mortality probabilities with IPLMs

In the IPLM approach, the fate of every individual in a given stage can be described by two pieces of information: 1) the time-to-event, $y_A \in \{1, \dots, t_c\}$ (where t_c is the time beyond which data are right censored); and 2) the event-type, $y_B \in \{\text{stage completion, mortality, censored}\}$. Probabilities associated with combinations of these two pieces of information can be obtained as follows. Given stage-specific parameters $\{\mu, \kappa, \nu, r\}$, construct the IPM-approximation (equation (2.5), main text). Project the unit density pulse vector $(\mathbf{n}^T, c_s)_0 = (1, 0, \dots, 0)^T$ and, for each time-step $t \in \{0, \dots, t_c\}$, record the probabilities to complete a stage, $p_d(t)$, or die, $p_m(t)$. The probability of right censor beyond t_c is $p_c = 1 - \sum_{t=1}^{t_c} (p_d(t) + p_m(t))$.

Mean and variance of sojourn-time distribution The probabilities p_d , p_m and p_c define the right censored sojourn-mortality time distribution for individuals in a given stage.

These probabilities provide the basis for calculating the likelihood of IPLM parameters given the observed data.

The mean and variance of the sojourn time distribution can be calculated as

$$\tilde{\mu} = \lim_{t_c \rightarrow \infty} \frac{\sum_{t=1}^{t_c} p_d(t)t}{\sum_{t=1}^{t_c} p_d(t)} \quad (\text{A.9})$$

and

$$\tilde{\sigma}^2 = \lim_{t_c \rightarrow \infty} \frac{\sum_{t=1}^{t_c} (t - \tilde{\mu})^2 p_d(t)}{\sum_{t=1}^{t_c} p_d(t)}. \quad (\text{A.10})$$

In practice, these quantities are approximated by setting t_c large enough that p_c become negligibly small.

A.4 Markov chain Monte Carlo strategy

Parameters μ , κ and ν are bounded on $(0, 1)$. We adopted the strategy of transforming these parameters to the logit scale to enable sampling on unbounded domains.

Thus, in general we sampled the logit transformed parameters

$$\mu' = \log \left(\frac{\mu}{1 - \mu} \right), \quad (\text{A.11})$$

$$\kappa' = \log \left(\frac{\kappa}{1 - \kappa} \right), \quad (\text{A.12})$$

$$\nu' = \log \left(\frac{\nu}{1 - \nu} \right). \quad (\text{A.13})$$

A simple change of variables correction reveals the prior densities of these transformed parameters to be

$$f(\mu') = \frac{1}{(1 + e^{\mu'})(1 + e^{-\mu'})}, \quad (\text{A.14})$$

$$f(\kappa') = \frac{1}{(1 + e^{\kappa'})(1 + e^{-\kappa'})}, \quad (\text{A.15})$$

$$f(\nu') = \frac{1}{(1 + e^{\nu'})(1 + e^{-\nu'})}. \quad (\text{A.16})$$

In the simulation-estimation study, a burn-in period consisting of a series of runs of 10^3 iterations was iterated until $\overline{LP}_{\text{run}} < 2 + \overline{LP}_{\text{run}-1}$, where $\overline{LP}_{\text{run}}$ is the mean log posterior density of the model over a given short

run. Expected sample size (ESS) (Plummer *et al.*, 2006) was calculated for each parameter from the final pre-run. Thinning was then set to $2 \times \min(\text{ESS})$, to remove much of the auto-correlation from subsequent samples. Thereafter, 10^4 thinned post-adaption samples were generated per model and convergence diagnostics were performed using CODA. These post-MCMC sampling steps generated a unique MCMC output file per stage and further CODA diagnostics were performed on those outputs. NIMBLE and R scripts used in these analyses are available on github https://github.com/scastano/IPLM_code.

Appendix B

Appendix to Chapter 3

B.1 *Culicoides* biting midges

B.1.1 *Culicoides* as disease vectors

Culicoides (Diptera: Ceratopogonidae) is a genus of biting midges of small size (approximately 1-3 mm). Most of the 1300 species known worldwide blood-feed on vertebrate mammals (including humans), birds, reptiles and other insects, and are distributed from the tropics to the tundra and from sea level to altitudes over 4200 m (Borkent and Wirth, 1997; Mellor *et al.*, 2000; Meiswinkel *et al.*, 2004b,a; Carpenter *et al.*, 2013). Yet, due to their diversity, small size and fragility which limit studies of their ecology in the field and complexify laboratory colonization; and due to their relatively limited impact on human health; and despite their importance in veterinary health; *Culicoides* remain the least studied of the major Dipteran vector groups (Carpenter *et al.*, 2013).

Adult *Culicoides* transmit several pathogens to humans and animals. They transmit parasites of veterinary importance, such as hemoparasites (*Haemoproteus* sp. in birds, *Hepaticystis kochi* in monkeys) and filariasis (*Onchocerca* sp. in horses and cattle). Nonetheless, their prime importance is due to the viruses they transmit.

Currently, the only known virus transmitted by *Culicoides* to humans is the Oropouche virus (OROV), which causes Oropouche fever, characterised by headaches in most of cases, and, less likely, by generalized arthralgia, anorexia and in rare cases meningitis (Linley *et al.*, 1983; Mellor *et al.*, 2000). OROV epidemics have occurred in Brazil, Peru and Panama since the beginning of the 1960s, with associated incidence rates remaining undetermined for the vast majority of outbreaks (Pinheiro *et al.*, 1981b; Watts *et al.*, 1997). Several studies suggest *Culicoides*

paraensis to be the main vector of OROV between humans during urban epidemics (Anderson *et al.*, 1961; Pinheiro *et al.*, 1981a; Roberts *et al.*, 1981; Hoch *et al.*, 1990), with estimates of people infected in Brazil alone of up to half a million (Pinheiro *et al.*, 1998). The Oropuche virus is currently restricted to the Amazonian region of South America, Central America and some Caribbean islands (Anderson *et al.*, 1961; Tesh, 1994; Baisley *et al.*, 1998; Vasconcelos *et al.*, 2001; da Silva Azevedo *et al.*, 2007).

Several species of *Culicoides* are involved in the transmission of viruses of veterinary importance. These pathogens include bluetongue virus (BTV); epizootic hemorrhagic; disease of deer virus; African horse sickness virus (AHSV); equine encephalosis virus; Akabane virus and Schmallenberg virus (Howerth *et al.*, 2001; Maclachlan and Guthrie, 2010; Maclachlan, 2011; Carpenter *et al.*, 2013; Mellor and Hamblin, 2004; Lievaart-Peterson *et al.*, 2012; Hoffmann *et al.*, 2012; Lehmann *et al.*, 2012; Beer *et al.*, 2013). Among these, AHSV and BTV are notifiable to the OIE (Office International des Epizooties, World Animal Health Organisation) because of their potential of rapid spread and economic impact (Office International des Epizooties, online). *Culicoides* have been responsible of massive outbreaks of AHSV in Asia in 1959-61 (leading to the death of over 300,000 equids), and in 1987-91 in Spain and Portugal (Rodriguez *et al.*, 1992), proving for the first time that the virus could overwinter in Europe (Mellor and Hamblin, 2004).

Yet, the paradigmatic illustration of the devastating effect of a *Culicoides*-borne virus on naive livestock populations is the unprecedented series of outbreaks of BT in Europe which started in 1998 and is still ongoing in some of the affected countries. These epidemics, involving multiple strains of BTV, spread across most of the Mediterranean basin and the Balkan areas, reaching historically uninfected areas 800 km further north in Europe than previously reported (Mellor and Wittmann, 2002; Purse *et al.*, 2005), causing the most severe outbreak of this disease ever recorded (Gubbins *et al.*, 2008). The virus reached new regions where a key vector (and invasive species), *Culicoides imicola*, was present, but also regions where *C. imicola* was absent, leading to the discovery that some species of indigenous European *Culicoides* were also able to transmit the virus. This alarming scenario imposed strict monitoring and drastic movement restrictions heavily impacting international trade issues with major economic consequences.

B.1.2 Life-cycle

The *Culicoides* genus contains species of holometabolic midges, i.e., they undergo complete metamorphosis. A typical *Culicoides* life cycle (B.1) consists of a series of immature aquatic/semi-aquatic stages (egg, four larval instars and pupa) followed by a mature aerial stage (adult). Adult *Culicoides* live between 10 and 20 days, although exceptionally life-span can extend to more than 90 days. In most species, females are hematophagous, a blood meal being required for eggs maturation. Eggs breed in moist conditions in a wide variety of habitats, particularly damp, muddy areas containing organic matter such as marshes, bogs, beaches, swamps, tree holes, irrigation pipe leaks, streams, saturated soil, animal dung, and rotting fruit and other vegetation (Mellor *et al.*, 2000). The mean egg number laid per cycle varies among species (from about 30 to over 400 eggs). Hatching takes place 3 to 5 days after eggs are laid. Larvae go through four stages of development, which can take from five days to many months, to develop into pupae. Temperate region species overwinter in the last larval state via a developmental diapause, developing to pupae and emerging as adults once environmental conditions are favorable (Kettle, 1984). The pupal stage is in general short (2 to 3 days) and gives rise to emergent adults. Developmental times within immature stages depend on environmental (particularly on temperature and humidity) and demographic factors. Under favorable environmental conditions, development from egg to adult takes about 15 days, but during overwintering periods it can take up to 7 months. Emergent females are called nulliparous (i.e. that have not laid eggs). The flight range of *Culicoides* (for seeking a mate, taking blood-meal or searching an oviposition site) usually is short (a few hundred meters from breeding sites) although some species disperse to a few kilometers (Lillie *et al.*, 1985) and individuals have been known to disperse some hundred of kilometers (Sellers *et al.*, 1977), a phenomenon that has received recent attention from modellers (Eagles *et al.*, 2014; Burgin *et al.*, 2012) and population geneticists (Jacquet *et al.*, 2016). After fecundation, females take a blood meal (although a few species are autogenous) to get the energy needed for producing and laying eggs. Once the first eggs are laid, females are said to be parous. The time interval between two consecutive blood-meals corresponds to one gonotrophic cycle, and takes usually between 3-5 days although may be reduced further at climatically optimal periods (Holmes and Birley, 1987). Female longevity determines the number of blood meals obtained within a lifetime.

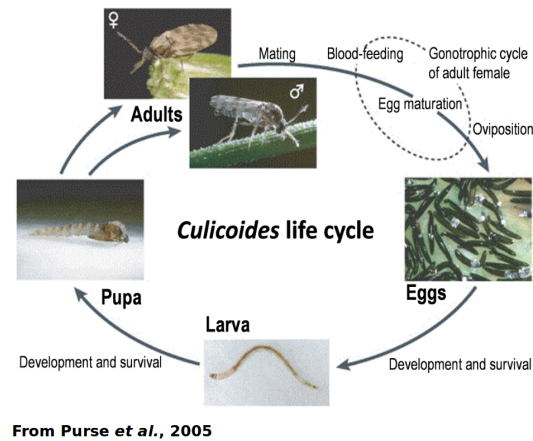


Figure B.1. *Culicoides* life cycle, adapted from Purse *et al.* (2005).

A female can become infected by taking a blood meal on a viraemic host. The female will then be able to transmit the virus to other hosts during the following blood meals. Infected females stay infectious for their entire life-span. No vertical transmission of BTV from a female to its eggs has ever been described in *Culicoides*, although genetic evidence from nuliparous *C. obsoletus/scoticus* and *C. punctatus* suggests Schmallenberg virus can be transmitted vertically (Larska *et al.*, 2013). Thus, only parous females are involved in BTV transmission. Transmission can only occur if the life-span of a female is greater than the duration of the gonotrophic cycle since the female needs to take at least two blood meals for transmission to occur.

B.2 *Culicoides* life cycle data

Laboratory data from two *Culicoides* species were used to parameterise development, survival and fecundity for a complete life cycle (see B.1). Note, these species share similar developmental responses across the 15°C–35°C range (Purse *et al.*, 2015). Our own insectarium (ASTRE, Montpellier) provided individual-level data on gonotrophic cycle durations, number of gonotrophic cycles and number of eggs laid for *C. nubeculosus* females at 15°C, 20°C and 25°C (Balenghien *et al.*, 2016). The laboratory also provided egg maturation and survival data at 15°C. Two similar *C. variipennis* studies (Mullens and Rutz, 1983; Vaughan and Turner, 1987) provided maturation time data for immature stages (egg, larva and pupa) – these data were available either as individual-level maturation times or summary statistics (sample means and standard deviations). The study of Vaughan and Turner (1987) provided developmental data at 20°, 23°, 27°, 30° and 35°C whilst that of Mullens and Rutz (1983) provided

developmental data at 17°, 20°, 23°, 27°, 30° and 35°C. In the latter case, published data were complimented with original notes from the author’s lab-book which gave initial sample sizes and individual-level data for pupal and composite larval-pupal stage studies. Published larvae-only data from both studies were not used due to ambiguity regarding sample sizes. The likelihoods functions used to analyse data type are given in section B.3.

Temp	DEVELOPMENT, SURVIVAL & FECUNDITY DATA							
	<i>C. nubeculosus</i>			<i>C. variipennis</i>				
	Balenghien <i>et al.</i>			Mullens <i>et al.</i>			Vaughan <i>et al.</i>	
	gonotrophic cycle	egg	fecundity	egg	larva & pupa	pupa	egg	pupa
15	1k	1c	3					
17				2u [†]	1c	2u		
20	1k		3	2u [†]	1c	2u	2p [†]	2p [†]
23				2u [†]	1c	2u	2p [†]	2p [†]
25	1k		3					
27				2u [†]	1c	2u	2p [†]	2p [†]
30				2u [†]	1c	2u	2p [†]	2p [†]
35							2p [†]	2p [†]

Table B.1. Available laboratory data for estimating *Culicoides* life cycle parameters. Data sets marked “1” provide sojourn time frequency distributions with “k” indicating that mortality date frequency distributions are *known* and “c” indicating *clumping* of mortality and right censored data. Data sets marked “2” provide the mean & standard deviation of observed sojourn times with “p” indicating the *proportion* surviving at each temperature is known and “u” indicating survival related information was *unreported*. The data set marked “3” provides the number of eggs laid per female in the gonotrophic cycle study. The notation † indicates missing sample size data requiring imputation steps described in section B.4.

B.3 Likelihood functions

For every column of table B.1, data at different temperature are assumed to be independent. Thus, for a given stage and data set, equation 2.11 in the main text is

$$f(\mu_1, \dots, \mu_K, \kappa_k, \dots, \kappa_K, \nu_1 \dots \nu_K, r | \mathbf{y}) \propto \prod_{k=1}^K f(y_k | \mu_k, \kappa_k, \nu_k, r) f(\mu_k, \kappa_k, \nu_k, r), \quad (\text{B.1})$$

where $\mathbf{y} = \{y_1, \dots, y_K\}$ and y_k is data collected at the k^{th} temperature. The Bayesian framework easily permits to extend expression (B.1) for the cases where models were fitted using multiple data sources – they are assumed independent each other.

Details of the various expressions used for the likelihood function $f(y_k|\mu_k, \kappa_k, \nu_k, r)$, which depend on the statistical information provided in data $\mathbf{y} = \{y_1, \dots, y_K\}$ (see table B.1), are presented below. The following expressions for each data type are valid at every empirical temperature and implicitly dependent of r , thus r and subscripts k are dropped for brevity.

Modelling sojourn-mortality time frequency data. Where data provide counts for the number of individuals that can be associated with $p_d(t)$, $p_m(t)$ or p_c at each time step, then a multinomial likelihood function can be defined. When such data is available at the individual-level – i.e. for each observation the total number of counts in one – then a categorical likelihood function can be used. This was the case for adult females in the *Culicoides* study – note, data sets marked "1k" in table S1 provide the sojourn-mortality time frequency distribution for each gonotrophic cycle of individual females. In that study, the sojourn-time frequency distribution was used to model the time required to complete each gonotrophic cycle, where the completion of a gonotrophic cycle was indicated by the completion of egg-laying.

The fate of each individual i in a given stage (or gonotrophic cycle) is described by time-to-event $y_{iA} \in \{1, \dots, t_c\}$ and event-type $y_{iB} \in \{\text{development, mortality, censored}\}$ data. Each observation can therefore be modelled by assuming

$$(y_{iA}, y_{iB}) \sim \text{Categorical}(p_d(1), p_m(1), \dots, p_d(t_c), p_m(t_c), p_c). \quad (\text{B.2})$$

Sojourn time frequency data with clumped mortality – right-censor information. Sometimes it is difficult or impossible to know precisely when mortality has occurred or even how many individuals have died prior to right censor time t_c . In such scenarios, the inability to distinguish dead from censored individuals requires clumping of the probabilities $p_m(t)$ and p_c to match the clumping of the associated data. Once data and probabilities are correctly clumped, then the likelihood is derived following a similarly procedure to the previous case.

In the *Culicoides* study, data sets marked "1c" in table S1 provide sojourn time frequency distribution data but do not distinguish the number of dead individuals from immature individuals still alive after t_c . Let y_t be the number of individuals completing maturation in t days, let y_{mc} be the number of dead or right censored individuals

and y^{total} be the sample size. Likelihoods for such data are simply obtained by assuming

$$y_1, \dots, y_{t_c}, y_{\text{mc}} \sim \text{Multinomial}\left(y^{\text{total}}, p_d(1), \dots, p_d(t_c), 1 - \sum_{t=1}^{t_c} p_d(t)\right), \quad (\text{B.3})$$

where probabilities $p_d(t)$ are obtained as described previously.

Likelihood for mean and standard deviation of sojourn times. Sometimes the only data available are summary statistics, typically means and standard deviations, obtained from publications. In such cases, we need the likelihood of the available summary statistics given the parameters.

For data sets marked "2u" and "2p" in table S1, data y consists of sojourn-time mean, μ^{obs} , and standard deviation, σ^{obs} . Ideally the sample size N is available too. Assuming normality, the following likelihood can be written for this data:

$$f(\mu^{\text{obs}}, \sigma^{\text{obs}} | \mu, \kappa, \nu) = \frac{S^{\text{Post}}}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{\sigma^{\text{obs}2} + (\mu^{\text{obs}} - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right\}, \quad (\text{B.4})$$

where $\{\mu, \kappa, \nu\}$ are the model parameters, and $\tilde{\mu}$ and $\tilde{\sigma}$ are moments of the joint sojourn-mortality time distribution estimated from probabilities $p_d(t)$ (Appendix A.3), calculated with t_c high enough that p_c falls below a precision threshold and could be assumed negligible. Note, in equation (B.4) above, μ^{obs} and σ^{obs} are not independent of mortality ν . However, greater precision in mortality ν is possible when survival data are available too. Sometimes, detailed survival data are not available, but the number, or proportion, of individuals surviving until t_c is reported. Data sets marked "2p[†]" (in table S1) reported the proportion of individuals surviving each experimental temperature, π , plus the total sample size at the start of each experiment, $S^{\text{Pre}}_{\text{total}}$. The proportions π were used in the imputation of S^{Post} (B.4). The posterior likelihood of daily survival ν , given $\{\mu, \kappa\}$, was given by the beta-binomial model

$$f(\nu | \mu, \kappa, S^{\text{Pre}}, S^{\text{Post}}) \propto \hat{\pi}^{S^{\text{Post}}} (1 - \hat{\pi})^{S^{\text{Pre}} - S^{\text{Post}}}, \quad (\text{B.5})$$

where $\hat{\pi} = 1 - \sum_{t=1}^{t_c} p_m(t)$.

B.3.1 Bayesian model for expected fecundity

Fecundity data obtained from the gonotrophic cycle experiment (marked "3" in table S1) provided the number of eggs laid at the end of each gonotrophic cycle. The posterior likelihood for the expected fecundity given this

data was obtained using the following approach. Expected fecundity at each temperature was estimated from oviposition data using a Poisson model with Jeffrey's prior, resulting in posterior distributions

$$F_A(T_k) \sim \text{Gamma}\left(\text{shape} = \frac{1}{2} + \sum_{i=1}^{N_{T_k}} n_{i_{T_k}}, \text{rate} = N_{T_k}\right) \quad (\text{B.6})$$

where, at temperature T_k , N_{T_k} is the number of observed ovipositions and $n_{i_{T_k}}$ is the fecundity of the i^{th} oviposition. As for survival and development, a unimodal constraint with respect to temperature was used to facilitate unimodal spline interpolation (see section B.7).

B.4 Imputation of missing data

Three of the data sets used for modeling the *Culicoides* life cycle presented missing sample size data (see table S1). Bayesian imputation steps to account for associated uncertainties are described here.

Sample sizes at experimental temperatures T_k were typically reported in one of the two forms : the sample size at the start of maturation experiment k , S_k^{Pre} , and the sample size at the end of a maturation experiment k , S_k^{Post} . It was assumed that any differences between S_k^{Pre} and S_k^{Post} could be accounted for by mortality and right censoring.

In Vaughan *et al.*'s egg and pupae studies, the total sample size $S_{\text{total}}^{\text{Pre}}$ was published, but how those numbers were divided among the five experimental temperatures was missing data. Moreover, neither S_k^{Post} or $S_{\text{total}}^{\text{Post}}$ were published, although the proportion that survived each experiment, π_k , was available. The duration of the experiment was not published, curtailing the possibility to account for potential right censor. It was assumed *a priori* that the expected sample sizes in each of the n_T experimental temperatures were equivalent. Dropping the k notation for brevity, this gave the following prior for each experimental group,

$$S_1^{\text{Pre}}, \dots, S_{n_T}^{\text{Pre}} \sim \text{Multinomial}\left(S_{\text{total}}^{\text{Pre}}, \frac{1}{n_T}, \dots, \frac{1}{n_T}\right). \quad (\text{B.7})$$

The proportions π_k that survived at each experimental temperature T_k permitted imputed values for S_k^{Post} to be determined as

$$S_k^{\text{Post}} = \pi_k S_k^{\text{Pre}}. \quad (\text{B.8})$$

In the Mullens and Rutz egg data, a total sample size at the end of the maturation experiment was reported, but

how those sample sizes were distributed among the different temperature groups was not reported. However, those authors did note in their paper that those values were roughly equivalent. Thus, we adopted the following prior

$$S_1^{\text{Post}}, \dots, S_{n_T}^{\text{Post}} \sim \text{Multinomial}\left(\mathbf{S}_{\text{total}}^{\text{Post}}, \frac{1}{n_T}, \dots, \frac{1}{n_T}\right). \quad (\text{B.9})$$

B.5 Posterior distributions of *Culicoides* IPLM model parameters

For stages $S = \{\text{egg}, \text{adult}\}$, posterior densities of parameters at each experimental temperature were of the form

$$f(\mu_S, \kappa_S, \nu_S, r_S, \hat{\mathbf{y}}_S | \mathbf{y}_S) \propto f(\mathbf{y}_S | \mu_S, \kappa_S, \nu_S, r_S, \hat{\mathbf{y}}_S) f_{\mathcal{U}}(\mu_S, \kappa_S) f_{\mathcal{U}}(\nu_S) f(r_S) f(\hat{\mathbf{y}}_S), \quad (\text{B.10})$$

where \mathbf{y}_S represents the full set of data, from different various sources, for stage S at a given empirical temperature, \mathcal{U} indicates the unimodality constraints, and $\hat{\mathbf{y}}_S$ represents imputed missing data. Recall, unimodal constraint \mathcal{U} includes: 1) unimodal response of parameters ν , \mathcal{P}_{01} and \mathcal{P}_{99} to temperature, and 2) unimodality in $f(\Delta | \alpha_1, \alpha_2)$, which is ensured when both α_1 and α_2 are greater than one. Posterior densities for larvae and pupae parameters per experimental temperature were

$$\begin{aligned} f(\mu_L, \mu_P, \kappa_L, \kappa_P, \nu_L, \nu_P, r_L, r_P, \hat{\mathbf{y}}_P | \mathbf{y}_P, \mathbf{y}_{LP}) &\propto f_{\mathcal{U}}(\mu_L, \kappa_L) f_{\mathcal{U}}(\nu_L) f(r_L) f_{\mathcal{U}}(\mu_P, \kappa_P) f_{\mathcal{U}}(\nu_P) f(r_P) f(\hat{\mathbf{y}}_P) \\ &\quad \times f(\mathbf{y}_P | \mu_P, \kappa_P, \nu_P, r_P, \hat{\mathbf{y}}_P) \\ &\quad \times f(\mathbf{y}_{LP} | \mu_L, \kappa_L, \nu_L, \mu_P, \kappa_P, \nu_P, r_L, r_P). \end{aligned} \quad (\text{B.11})$$

Likelihoods for data \mathbf{y}_E , \mathbf{y}_P and \mathbf{y}_A were calculated from sojourn-mortality distributions obtained by projecting the unit pulse vector $(\mathbf{n}^T, c_S)_0 = (1, 0, \dots, 0)^T$ with equation 2.5) (main text). The likelihood for \mathbf{y}_{LP} was calculated similarly using the following two-stage projection matrix for the IPM-approximation:

$$\mathcal{M}_{LP} = \begin{bmatrix} \mathbf{W}_L & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_L & \mathbf{W}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_P^1 & \mathbf{1} \end{bmatrix}. \quad (\text{B.12})$$

B.6 MCMC strategy

For the *Culicoides* IPLM model, stage-specific parameters μ , κ and ν were transformed to the logit scale, as in Chapter 2 (Appendix A.4). In this case, target posterior distributions were sampled using a parallel tempering algorithm (Swendsen and Wang, 1986; Liu, 2008; Łacki and Miasojedow, 2015) adapted from NIMBLE’s library of functions for adaptive MCMC. Tempering was not applied to the transformed priors (equations (A.14–A.16) since the prior on the original scale is already flat, neither was tempering applied during the imputation steps. Depending on the analysis, the tempering was performed with temperature ladders of 10 to 15 different temperatures (n_{Temps}). Temperature ladders were initialised as $\mathbf{T} = e^{\log(0 \times 10)}, e^{\log(1 \times 10)}, \dots, e^{\log(n_{\text{Temps}} \times 10)}$ and were adjusted to target a 0.234 acceptance rate using techniques described in (Łacki and Miasojedow, 2015) during an adaptive burn-in phase.

A burn-in period, consisting of a series of runs of 10^4 iterations was iterated until $\overline{LP}_{\text{run}} < 2 + \overline{LP}_{\text{run}-1}$, where $\overline{LP}_{\text{run}}$ is the mean log posterior density of the model over a given short run. Expected sample size (ESS) (Plummer *et al.*, 2006) was calculated for each parameter from the final pre-run. Thinning was set to $2 \times \min(\text{ESS})$, to remove much of the auto-correlation from subsequent samples. Thereafter, 10^4 thinned post-adaption MCMC samples were generated per model and convergence diagnostics were performed using CODA.

To avoid mixing difficulties in the *Culicoides* study, the above sampling strategy was applied with resolution parameters fixed at values (or combinations of values for larvae-pupae) given by $r_E, r_P, r_A \in \{1, \dots, 50\}$ and $r_L \in \{1, \dots, 15\}$. Integration over each $r_S \in \{r_E, r_A\}$ was achieved in a post-MCMC step by sampling among lines of model output with weights

$$w_l(r_S) = \frac{f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}}|\mathbf{y}, r_S)}{\sum_{r'_S=1}^{50} f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}}|\mathbf{y}, r'_S)}, \quad (\text{B.13})$$

where $l \in \{1, \dots, 10^4\}$ indicates the MCMC output line. For the larvae-pupae analysis, equation (B.13) was adjusted to include integration over both r_L and r_P such that

$$w_l(r_L, r_P) = \frac{f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}}|\mathbf{y}, r_L, r_P)}{\sum_{r'_L=1}^{15} \sum_{r'_P=1}^{50} f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}}|\mathbf{y}, r'_L, r'_P)}. \quad (\text{B.14})$$

These post-MCMC sampling steps generated a unique MCMC output file per stage and further CODA diagnostics

were performed on those outputs. NIMBLE and R scripts used in these analyses are available on github https://github.com/scastano/IPLM_code.

B.7 Unimodal cubic Hermite spline interpolation

Posterior estimates of $\text{logit}(\mathcal{P}_{01S})$, $\text{logit}(\mathcal{P}_{99S})$, $\text{logit}(\nu_S)$ and $\log(F_A)$ at unsampled temperatures were obtained via interpolation with unimodal (i.e. up to two piece-wise monotonic) cubic Hermite splines. Recall, given n points (x_k, y_k) , where $k \in \{1, \dots, n\}$ and $x_k < x_{k+1}$ for all k , a cubic Hermite spline between two successive points is defined

$$f_{\text{interpolated}}(t) = y_k h_{00}(t) + (x_{k+1} - x_k) m_k h_{10}(t) + y_{k+1} h_{01}(t) + (x_{k+1} - x_k) m_{k+1} h_{11}(t), \quad (\text{B.15})$$

where m_k is the gradient at point k , $t = \frac{x - x_k}{x_{k+1} - x_k}$ and h_{ii} are the cubic Hermite spline basis functions $h_{00}(t) = (1 + 2t)(1 - t)^2$, $h_{10}(t) = t(1 - t)^2$, $h_{01}(t) = t^2(3 - 2t)$ and $h_{11}(t) = t^2(t - 1)$. Equation (B.15) is available in R as `splinefunH` in the `stats` package.

The Fritsch-Carlson method (Fritsch and Carlson, 1980) provides a deterministic algorithm for setting gradients m_k such that the fitted spline is piece-wise monotonic. Pseudo-code for a classic implementation of the Fritsch-Carlson method is given in Algorithm B.7.1.

We use a stochastic variation of the Fritsch-Carlson algorithm that permits uncertainty in gradients m_k to be explored within the piece-wise monotonic constraint $\alpha_k^2 + \beta_k^2 < 9$. We assume throughout that the set of points $(x_1, y_1), \dots, (x_n, y_n)$ contains at most one local maximum – a condition used as a constraint during MCMC. Moreover, we initialise each m_k with a draw from the following unconstrained conditional (on \mathbf{y}) priors,

$$\arctan(m_k) \sim \begin{cases} \text{Uniform}(\arctan(\Delta_k), \arctan(\Delta_{k-1})) & \text{if } y_{k-1} < y_k \text{ and } y_k > y_{k+1}, \text{ else} \\ \text{Uniform}(0, \arctan(3\Delta_{k-1})) & \text{if } y_{k-1} < y_k, \text{ else} \\ \text{Uniform}(\arctan(3\Delta_{k-1}), 0) & \text{if } k \text{ s.t. } y_{k-1} > y_k \end{cases} \quad (\text{B.16})$$

where $\Delta_0 = \Delta_1$ and impose the constraint that $\alpha_k^2 + \beta_k^2 < 9$, which truncates these priors, to avoid “overshoot”. We use a stepping-in algorithm to re-sample $\arctan(m_k)$ wherever the unimodality constraint is violated. Pseudo-code for this sampling strategy is given in Algorithm B.7.2.

Algorithm B.7.1: FRITSCH-CARLSON METHOD(\mathbf{x}, \mathbf{y})

comment: Set slopes of secant lines

$$\Delta_k = \frac{x - x_k}{x_{k+1} - x_k} \quad \forall k \in \{1, n-1\}$$

comment: Initialise tangents

$$m_1 = \Delta_1 \text{ and } m_n = \Delta_{n-1}$$

for $k \in \{2, n-1\}$

$$\text{do } \begin{cases} m_k = \frac{\Delta_{k-1} + \Delta_k}{2} \\ \text{if } \text{sign}(\Delta_{k-1}) \neq \text{sign}(\Delta_k) \\ \text{then } m_k = 0 \\ \text{if } \Delta_k = 0 \\ \text{then } m_k = m_{k+1} = 0 \end{cases}$$

comment: Derive parameters α and β

$$\alpha_k = m_k / \Delta_k \quad \forall k \in \{1, n-1\}$$

$$\beta_k = m_{k+1} / \Delta_k \quad \forall k \in \{1, n-1\}$$

comment: Reset gradients where monotonic constraint is violated

for any k s.t. $\alpha_k^2 + \beta_k^2 > 9$

$$\text{do } \begin{cases} m_k = \frac{3\alpha_k \Delta_k}{\sqrt{\alpha_k^2 + \beta_k^2}} \\ m_{k+1} = \frac{3\beta_k \Delta_k}{\sqrt{\alpha_k^2 + \beta_k^2}} \end{cases}$$

Finally, for each stage S (suffix dropped for brevity), development kernel parameters α_1 and α_2 at any temperatures T absent in the set of experimental temperatures for S were derived from interpolated values $\mathcal{P}_{01}(T)$ and $\mathcal{P}_{99}(T)$ by numerical minimisation. Thus, α_{1T} and α_{2T} were calculated as

$$\alpha_{1T}, \alpha_{2T} = \arg \min_{\alpha_1, \alpha_2} \sum_{p \in \{1, 99\}} \left(\hat{\mathcal{P}}_p(\alpha_1, \alpha_2) - \mathcal{P}_p(T) \right)^2, \quad (\text{B.17})$$

where $\hat{\mathcal{P}}_p(\alpha_1, \alpha_2)$ was calculated with R function `qbeta` and minimisation was performed using the Nelder-Mead algorithm of R function `optim`.

Algorithm B.7.2: STOCHASTIC UNIMODAL CUBIC SPLINE(x, y)

comment: Set slopes of secants and identify mode

$$\Delta_k = \frac{x - x_k}{x_{k+1} - x_k} \quad \forall k \in \{1, n-1\} \text{ and } k_{max} = \arg \max_k(y_k)$$

comment: Initialise bounds and sample gradients

for $k \leftarrow 1$ **to** n

do $\left\{ \begin{array}{l} \text{if } k = k_{max} \text{ and } k \notin \{1, n\} \text{ then } L_k = \Delta_k \text{ and } U_k = \Delta_{k-1} \\ \quad \text{else if } y_k < y_{k+1} \text{ then } L_k = 0 \text{ and } U_k = 3\Delta_{k-1} \\ \quad \text{else if } y_k > y_{k+1} \text{ then } L_k = 3\Delta_{k-1} \text{ and } U_k = 0 \\ \quad \arctan(m_k) \sim \text{Uniform}(\arctan(L_k), \arctan(U_k)) \end{array} \right.$

$$\alpha_k = m_k / \Delta_k \text{ and } \beta_k = m_{k+1} / \Delta_k \quad \forall k \in 1, \dots, n-1$$

while $\max\{\alpha_k^2 + \beta_k^2 : k \in 1, \dots, n-1\} > 9$

do $\left\{ \begin{array}{l} \kappa = \arg \max_k(\alpha_k^2 + \beta_k^2) \\ \text{if } \beta_\kappa > \alpha_\kappa \text{ then } \kappa = \kappa + 1 \\ \text{if } \kappa = k_{max} \text{ and } m_\kappa < 0 \text{ then } L_\kappa = m_\kappa \\ \quad \text{else if } \kappa = k_{max} \text{ and } m_\kappa > 0 \text{ then } U_\kappa = m_\kappa \\ \quad \text{else if } \kappa < k_{max} \text{ then } U_\kappa = m_\kappa \\ \quad \text{else if } \kappa > k_{max} \text{ then } L_\kappa = m_\kappa \\ \quad \arctan(m_\kappa) \sim \text{Uniform}(\arctan(L_\kappa), \arctan(U_\kappa)) \\ \quad \alpha_k = m_k / \Delta_k \text{ and } \beta_k = m_{k+1} / \Delta_k \quad \forall k \in 1, \dots, n-1 \end{array} \right.$

Appendix C

Appendix to Chapter 4

C.1 A brief description of the synthetic likelihood method

The synthetic likelihood (SL) ¹ is a modern simulation-based approach for making statistical inference from noisy and highly non-linear dynamic systems (Wood, 2010). It can be used for statistical inference with state-space models to provide a “synthetic” likelihood for use in optimisation or MCMC algorithms. In a Bayesian context, the idea is to provide an approximation of the posterior $p(\theta|y)$. Note, that the state x is marginalised out of this posterior via Monte Carlo simulation.

Like Approximate Bayesian Computation (ABC), the SL is based on a set of summary statistics that characterises the variation in observed and simulated data sets. Unlike ABC, the SL method makes the assumption that the distribution of summary statistics s can be approximated by a multivariate normal distribution

$$s \sim N(\mu_\theta, \Sigma_\theta) \quad (C.1)$$

with μ_θ the mean vector and Σ_θ the associated covariance matrix. For the true data y_{obs} , both μ_θ and Σ_θ are unknown, since they are generally intractable functions of the vector of unknown model parameters, θ . However, for any given θ they can be estimated by simulating replicated data from the model, in which case a sort of “synthetic likelihood” can be evaluated (Fig. C.1).

Based on s , the synthetic likelihood of any given parameter vector θ can be evaluated as follows. First

¹adapted from (Wood, 2010)

we parameterised the model via Θ to simulate N_r replicate data sets, $\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_{N_r}^*$. We then convert these to replicate statistics $\mathbf{s}_1^*, \mathbf{s}_2^*, \dots, \mathbf{s}_{N_r}^*$, exactly as \mathbf{y} was converted to \mathbf{s} . Then we evaluate $\hat{\boldsymbol{\mu}}_\theta = \sum_i \mathbf{s}_i^* / N_r$, $S = (\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \dots)$ and, hence, $\hat{\Sigma}_\theta = SS^T / (N_r - 1)$ (Ling, 1990). Dropping irrelevant constants, the log synthetic likelihood is

$$l_s(\boldsymbol{\theta}) = \frac{1}{2}(\mathbf{s}_\theta - \boldsymbol{\mu}_\theta)^T \hat{\Sigma}_\theta^{-1} (\mathbf{s}_\theta - \boldsymbol{\mu}_\theta) - \frac{1}{2} \log |\hat{\Sigma}_\theta^{-1}|. \quad (\text{C.2})$$

Like any log likelihood, $l_s(\boldsymbol{\theta})$ measures the consistency of the parameter values $\boldsymbol{\theta}$ with the observed data.

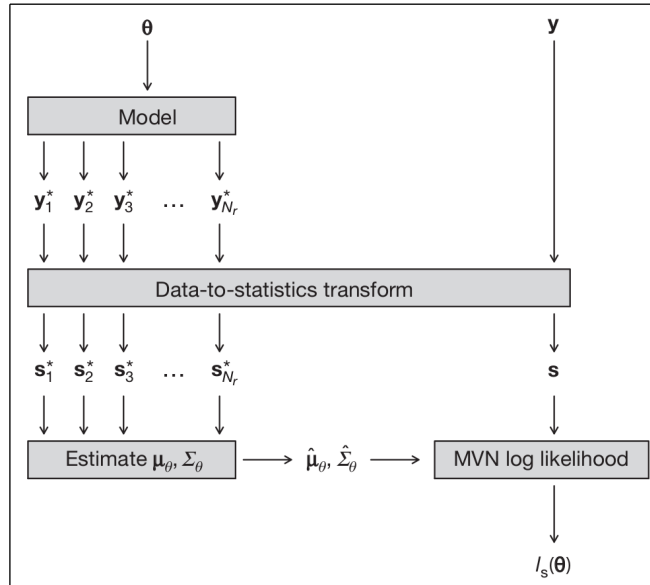


Figure C.1. A schematic representation of calculating a synthetic likelihood. From the top, we wish to evaluate the fit of the model with parameter vector θ to the raw data vector \mathbf{y} . Replicate data vectors $\mathbf{y}_1^*, \dots, \mathbf{y}_{N_r}^*$ are simulated from the model, given θ . Each replicate, and the observed or “true” data \mathbf{y} , is converted into a vector of statistics, \mathbf{s}_i^* or \mathbf{s} , in the same way. The \mathbf{s}_i^* are used to estimate the mean vector $\hat{\boldsymbol{\mu}}_\theta$, and covariance matrix, $\hat{\Sigma}_\theta$, of \mathbf{s} , according to the model with parameters θ . We use $\hat{\boldsymbol{\mu}}_\theta$, $\hat{\Sigma}_\theta$ and \mathbf{s} respectively as the mean vector, the covariance matrix and the argument of the log multivariate normal (MVN) probability density function, to evaluate the log synthetic likelihood, l_s . Adapted from (Wood, 2010).

This method is general enough to deal with hidden state variables, complicated observation processes, missing data and multiple data series. Calculated as described, l_s is invariant to reparameterisation and is robust to the inclusion of uninformative statistics, so very careful selection of statistics is not strictly necessary provided the set of statistics characterises key features of the dynamics. There is complete freedom to transform statistics to improve the Normal approximation in equation C.2. Furthermore, l_s behaves like a conventional log likelihood in

the limit as $N_r \rightarrow \infty$, giving access to much of the machinery of likelihood-based inference. For example, in the current thesis (chapter 4) l_s is used with a standard adaptative Metropolis-Hastings sampler.

C.1.1 Choice of summary statistics

The aim when generating summary statistics from some raw data series \mathbf{y} is to quantify local dynamic structure and the distribution of observations (Tavaré *et al.*, 1997; Wood, 2010). There is no rule to provide the choice of summary statistics; the most common approach is an *ad hoc* exploration of different combinations of summary statistics in order to improve fitting.

In our study (chapter 4, section 4.2) we define the adult abundance data set to which we fit our model as the *true* data set and each model simulation as a *simulated* data set. Let $N_{1:t}$ represent the series of adult density observations. In this preliminary study, the following summary statistics were considered: (1) mean of $N_{1:t}$; (2) standard deviation of $N_{1:t}$; (3) the median of $N_{1:t}$; (4) number of days where adult abundance was above the 0.9 percentile of the true data; and (5) the coefficients of the autocovariance function to lag l , where lag l was chosen to correspond to the first non-positive coefficient for the true data set in question.

Appendix D

Article: *Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes*

The following article has been submitted on January 16th to the journal *Methods in Ecology and Evolution*, and is composed of most of the developments presented in chapters 2 and 3 of this thesis.

Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes

Authors: María Soledad Castaño^{1,2,3}, Perry de Valpine⁴, Hélène Guis^{5,6,7}, David Richard John Pleydell^{1,2}

Email: soledadcastano@gmail.com, pdevalpine@berkeley.edu, helene.guis@cirad.fr, david.pleydell@inra.fr

Affiliations:

- (1) CIRAD, UMR117 ASTRE, 97170, Petit-Bourg, Guadeloupe, France;
- (2) INRA, UMR1309 ASTRE, 34398, Montpellier, France;
- (3) EA 4540 LAMIA, Université des Antilles et de la Guyane, BP 592, Campus de Fouillole, 97157, Pointe-à-Pitre, Guadeloupe;
- (4) Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA;
- (5) CIRAD, UMR117 ASTRE, Antananarivo, Madagascar;
- (6) Unité d'Epidémiologie, Institut Pasteur de Madagascar, Antananarivo, Madagascar;
- (7) Fofifa DRZV, Antananarivo, Madagascar.

Short running title: Integral projection Lefkovitch matrices

Key words: stage-structured models, stage duration distributions, integral projection models, individual heterogeneity, within-stage development, correlated stage-durations, temperature-dependence, degree-day accumulation, transient dynamics, hidden process models

Article Type: Standard

Number of words in the main text: 4738

Number of words in figure legends: 434

Number of words in references: 1540

Number of words in main, figures & references: 6712

Number of words in the abstract: 341

Number of figures, tables, and text boxes (Main Text): 6 figures, 0 tables and 0 text boxes

Number of figures and tables (Supporting Information): 7 figures and 1 table

Corresponding author:

David Pleydell, UMR CIRAD/INRA ASTRE, Domaine Duclos, Prise-d'Eau, Petit-Bourg, 97170, Guadeloupe.
 Tel: +590 5 90 25 54 44
 Fax: +590 5 90 94 03 96
 EMail: david.pleydell@inra.fr

Statement of authorship: SC, DP and HG conceived the study. SC, DP and PV wrote NIMBLE code. SC and DP performed the analyses and drafted the paper. All authors contributed substantially to revisions and gave final approval for publication.

Abstract

1. The predictive performance of many stage-structured population models is limited by unrealistic assumptions regarding individual variation. Errors become particularly prominent when fluctuating conditions affect vital rates differentially across stages. Integral projection models (IPMs) enable more realistic individual heterogeneity assumptions. However, IPMs perform badly when appropriate developmental trait (e.g. size, weight) data are prohibitively difficult to obtain.

2. To overcome these limitations, we incorporate IPMs for within-stage development into classic Lefkovich matrix (CLM) models, and fit the model using maturation-time data. A Bayesian inference framework is developed and tested in two case studies. First, the estimation procedure was tested using data simulated from a model where unmeasured individual-level fixed-traits generate correlated maturation-times. Second, a temperature dependent model was fit to *Culicoides* (biting midge) experimental maturation-time data, permitting, for the first time, an analysis of transient dynamics for these insects under fixed and seasonal temperatures.

3. The simulation study demonstrated that accurate maturation-time distributions can be estimated using data from modestly sized marked cohort studies – even when individual-level fixed-traits correlate maturation-times. The *Culicoides* study indicated that: the posterior likelihood of a CLM model was negligibly small compared to the new model; the non-linear responses of vital rates to temperature differed markedly among stages; the inclusion of within-stage development greatly augmented the amplitude and duration of transient dynamics and altered maximum and minimum inertia.

4. By tracking within-stage development, the new matrix model greatly reduces stage-duration errors, improves robustness to perturbation, and enables realistic incorporation of both time-varying covariates and/or unmeasured local or genetic factors. Moreover, by using maturation-time (and not size) data, our methods can greatly improve the precision of stage-structured IPMs whenever size is a poor, or unavailable, predictor of stage duration. This scenario is ubiquitous in ecology: egg and exoskeleton dimensions often remain relatively constant, and more appropriate developmental metrics can be too expensive or difficult to collect routinely. The new model enables improved ecological forecasting, mechanistic niche modelling, demographic compensation analysis and eco-evolutionary analysis for stage-structured taxa. Diverse applications are expected for conservation, agricultural, epidemiological or theoretical purposes.

1 Introduction

A central premise of population biology is that the timing of life-cycle events drive a population's dynamics (Caswell, 2006). When life-cycles progress via a series of developmental stages, describing how each stage's vital rates vary with genetic and environmental factors provides a basis for studying a population's dynamics (Manly, 1990). Analysis and simulation with stage-structured models (SSMs) has featured numerous biological complexities including density-dependence, stochasticity, time-varying parameters and dispersal (Van Tienderen, 1995; Cushing *et al.*, 2002). However, the assumptions of many SSMs, regarding variation in the time required to mature through a given stage (i.e. the maturation-, development- or sojourn-time), are oversimplified, unrealistic and lack generality. This can result in poor stage-duration distribution (SDD) approximations and inaccurate predictions of population growth rates and related quantities (Bolnick *et al.*, 2011; Vindenes and Langangen, 2015). Overcoming these shortcomings would significantly increase the forecast horizon (Petchey *et al.*, 2015) of SSMs.

Most SSMs are developed as Markov processes where between-stage transition probabilities are independent of the time-duration spent in a given stage. This generates memoryless (exponential or geometric) SDDs with artifactual most-likely stage-durations. For discrete time models, a time-step Δt might be chosen to minimise errors generated by the geometric SSDs (Cushing *et al.*, 2002), but the values of Δt required for different stages or environmental conditions may vary. Alternatively, additional sub-stages provide greater SDD flexibility (Longstaff, 1984; Birt *et al.*, 2009), but in these approaches sojourn-time variance is tied to the number of sub-stages making the incorporation of time-varying covariates elusive.

Delay equations provide an alternative (Nisbet, 1997). In their basic form SDD variance is zero, although delays can be distributed (Berezansky *et al.*, 2010) or modeled with covariates (Yamanaka *et al.*, 2012) to improve realism. However, these models lack generality because delays are determined at one point in time and are impervious to subsequent changes in covariates.

Markov or delay assumptions generate negligible artifacts when stage distributions are stable. However, when exogenous factors (environmental fluctuations or biotic interactions) affect stages differentially, these assumptions yield inaccurate characterisation of transient and/or non-linear dynamics (Blythe *et al.*, 1984; Bierzychudek, 1999).

Stage-duration distribution models (SDDMs) provide an empirical alternative. Based on survival analysis – the statistical characterisation of time-to-event data – SDDMs explicitly model SDDs, typically by fitting two-parameter probability distributions to cohort data at each stage (Manly, 1990; Hoeting *et al.*, 2003). This approach

has been used for estimating mortality and correlated SDDs in stable environments (De Valpine and Knappe, 2015) and tracking degree-day accumulation in synchronised insect cohorts (Murtaugh *et al.*, 2012). But, as for delay equations, these models do not track within-stage development and challenges arise when incorporating time-varying development rates. This limitation is overcome when individual-based models (IBMs) are parameterised with SDDM output (Régnière *et al.*, 2012). However, computational costs can prevent IBMs from scaling well and more general solutions are required. Despite improved realism, wider application of SDDMs has been elusive due to the need to incorporate fluctuating conditions, density-dependence, or other feedback mechanisms.

Integral projection models (IPMs) offer a promising alternative (Easterling *et al.*, 2000). An IPM tracking within-stage development could maintain projection validity even when stage distributions are unstable (De Valpine, 2009). Typically, IPM-kernels are parameterised using regression analyses of time-lagged traits such as size or weight (Rees *et al.*, 2014). The approach has become popular in plant (Ellner and Rees, 2007; Merow *et al.*, 2014) and animal (Coulson *et al.*, 2011; Ozgul *et al.*, 2012) studies where key traits are easily measured. But IPMs are rarely used when direct measurement of state variables – such as accumulated contamination, parasitic load, physical damage, or degree-days – is impracticable. Moreover, size can be a poor predictor of stage duration, for example when eggshells, exoskeletons or hosts effectively hide within-stage development. In these scenarios, more performant development metrics are often too difficult or expensive to collect routinely. Where IPMs have been used to model egg stages, assumptions leading to geometric maturation-time distributions, and thus bias, have been used (Ozgul *et al.*, 2012; Smallegange *et al.*, 2014).

Here, we extend standard matrix models by incorporating IPM approximations that track individuals through a series of developmental sub-stages to yield more realistic SDDs. Estimation is achieved by treating within-stage development as an unobserved state variable. These new "integral projection Lefkovitch matrix" (IPLM) models facilitate parameterisation with time-varying covariates, provide valid transition probabilities for non-stable stage distributions, and reduce errors in transient or non-linear dynamics analyses. They can therefore improve SSM forecast horizons.

In section 2, we outline the IPLM framework and describe how to parameterise these models using maturation-time data. In section 3, we test our estimation methods in a simulation study where variance in individual quality generates correlated sojourn-times. In section 4, we show how laboratory data on biting midge development at different temperatures can be used to fit an IPLM model for predicting population dynamics in the field that demonstrates more realistic dynamics than when developmental heterogeneity is neglected. Finally, in section 5 we discuss the implications of these developments for ecology and evolution.

2 Integral projection Lefkovitch matrices

Lefkovitch matrix models. The following recursive formula is a popular tool for studying demographic dynamics

$$\mathbf{N}_t = \mathbf{M}_t \mathbf{N}_{t-1}, \quad \text{eqn 1}$$

where \mathbf{N}_t denotes a vector of densities for a series of k age (Leslie, 1945) or stage (Lefkovitch, 1965) classes at time t and \mathbf{M}_t is a projection matrix. Lefkovitch (i.e. stage-structured) matrices can be constructed in many ways to match the great diversity of life-cycle strategies found in nature. Here, we focus on matrix models of the form:

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_k \end{bmatrix}_t = \begin{bmatrix} W_1 & F_2 & F_3 & \cdots & F_k \\ B_1 & W_2 & 0 & \cdots & 0 \\ 0 & B_2 & W_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & B_{n-1} & W_k \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_k \end{bmatrix}_{t-1}. \quad \text{eqn 2}$$

We call the matrix in eqn 2 a ‘classic Lefkovitch matrix’ (CLM). A tempting misinterpretation of eqn 2 is that, in time-step t , individuals in some stage $S \in \{1, \dots, k\}$ remain with probability W_S , advance one stage with probability B_S , contribute to the next generation with fecundity F_S and survive with probability $\nu_S = W_S + B_S$. However, this neglects within-stage developmental heterogeneity, assumes geometric sojourn-time distributions and only yields valid transition probabilities when stage distributions are stable (De Valpine *et al.*, 2014). Thus, such matrices can generate highly erroneous results unless vital rates are relatively constant and impervious to exogenous sources of variation.

Within-stage development. These limitations can be overcome by replacing scalar elements W_S , B_S and F_S of matrix \mathbf{M}_t with sub-matrices \mathbf{W}_S , \mathbf{B}_S and \mathbf{F}_S characterising within-stage development, between-stage development and fecundity respectively. Thus, every scalar n_S of eqn 2 is replaced by \mathbf{n}_S , a vector of r_S discrete sub-stages. Note, r_S can vary between stages.

We define sub-matrices \mathbf{W}_S , \mathbf{B}_S and \mathbf{F}_S via stage-specific IPMs. Unlike previous models (Longstaff, 1984; Birt *et al.*, 2009), this improves SDD approximations independently of r_S . An IPM for within-stage development can be written

$$n(\delta', t) = \int_0^1 K_\Theta(\delta, \delta') n(\delta, t-1) d\delta, \quad \text{eqn 3}$$

where $n(\delta, t)$ is the density of individuals with developmental status δ at time t , Θ is a parameter set, and the IPM-kernel $K_{\Theta}(\delta, \delta')$ quantifies the proportion of individuals with development δ that survive and develop to δ' in one time-step. Transition to next stage occurs once $\delta \geq 1$, whereby development in the new stage is initialised with $\delta = 0$.

For practical purposes, we simplify eqn 3 by assuming that the increments by which individuals develop are drawn independently from a distribution at each time step. Therefore, we re-write $K_{\Theta}(\delta, \delta')$ as $K_{\Theta}(\Delta)$, with $\Delta = \delta' - \delta$. In the examples below, we use for K_{Θ} a beta distribution with parameters $\{\mu, \kappa\}$ accounting for developmental rate heterogeneity, combined with survival probability ν . The beta distribution is a natural choice since δ ranges in $[0, 1)$ for each stage. This model provides the same level of parsimony as SDD models: each defines the distribution of sojourn-times and mortality with three parameters. In our model, the stage-specific kernel propagates individuals through a developmental process to derive probabilities of stage completion or death in any given time interval – these probabilities provide the basis for estimating parameters from data. Consequently, some computation time is required for fitting, but the benefit is a model formulated in discrete time-steps that can accommodate time-dependent parameters.

Like other IPMs, we approximate the continuous state variable δ by a series of discrete states. Given a series of discrete states between 0 and 1, we calculate the probability p_l of completing l discrete increments in a time-step by integrating K_{Θ} over an interval (see Appendix S1, Supporting Information). These transition probabilities provide, for a stage S , elements for the following $r_S \times r_S$ lower-triangular matrix:

$$\mathbf{W}_S = \nu_S \begin{bmatrix} p_{0,S} & & & & \\ p_{1,S} & p_{0,S} & & & \\ p_{2,S} & p_{1,S} & p_{0,S} & & \\ \vdots & \vdots & & \ddots & \\ p_{r-1,S} & p_{r-2,S} & & & p_{0,S} \end{bmatrix}. \quad \text{eqn 4}$$

Note, the probabilities $\{p_{0,S}, \dots, p_{r,S}\}$ depend on the stage-specific parameters $\{\mu_S, \kappa_S, \nu_S\}$. Matrix \mathbf{B}_S provides the proportion of individuals making the transition to the next stage, where development is initialised in the first sub-stage. Thus, if \mathbf{B}_S^1 denotes the first row of matrix \mathbf{B}_S , element j of \mathbf{B}_S^1 is $\sum_{l=r+1-j}^r p_{l,S}$. Each matrix \mathbf{F}_S is constructed assuming all individuals completing stage S contribute F_S to the next generation. Thus, the first row of \mathbf{F}_S is $\mathbf{F}_S^1 = F_S \mathbf{B}_S^1$. All other elements of \mathbf{B}_S and \mathbf{F}_S are zero and $F_S = 0$ for non-reproductive stages.

129 Alternative definitions for \mathbf{B}_S and \mathbf{F}_S are possible, but are not explored here for simplicity.

130 The matrix approximation of the IPM (eqn 3) for stage S is therefore

$$\begin{bmatrix} \mathbf{n}_S \\ c_S \end{bmatrix}_t = \begin{bmatrix} \mathbf{W}_S & \mathbf{0} \\ \mathbf{B}_S^1 & 1 \end{bmatrix}_t \begin{bmatrix} \mathbf{n}_S \\ c_S \end{bmatrix}_{t-1}, \quad \text{eqn 5}$$

131 where \mathbf{n}_S gives the population density in the r_S sub-stages, and c_S is the cumulative density of individuals that
132 have completed stage S . We call $\Theta_S = \{\mu_S, \kappa_S, \nu_S\}$ the parameter set of the discretised IPM-kernel.

133 The augmented equivalent of eqn 2 is

$$\mathcal{N}_t = \begin{bmatrix} \mathbf{W}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \dots & \mathbf{F}_k \\ \mathbf{B}_1 & \mathbf{W}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \mathbf{W}_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{B}_{m-1} & \mathbf{W}_k \end{bmatrix}_t \mathcal{N}_{t-1}, \quad \text{eqn 6}$$

134 where $\mathcal{N}^T = (\mathbf{n}_1^T, \dots, \mathbf{n}_k^T)$ and the model parameter set is $\Theta = \{\Theta_1, \dots, \Theta_k\}$. We call any matrix built on these
135 principals an integral projection Lefkovitch matrix (IPLM). When all $r_S = 1$, an IPLM reduces to a CLM. In
136 practice, we seek r_S small enough to maintain computational efficiency yet large enough to characterise sojourn-
137 time variance for stage S . Since the dimension of Θ_S is independent of r_S , parsimony is unaffected as matrix
138 dimension increases. The k IPM-kernels of an IPLM can incorporate fluctuating environmental conditions and
139 other sources of heterogeneity. Thus, these developments can greatly augment the range of scenarios studied with
140 the powerful tools of matrix model analysis.

141 **IPLM based survival analysis.** We consider fitting IPLMs using either marked or unmarked cohort development
142 data. Marked cohort data provide the time or time-interval of each stage transition per individual. Unmarked cohort
143 data include the number of individuals maturing from a stage in a time interval given that their development was
144 synchronised at $t = 0$. Typically, the number dying in one or more time intervals is also reported. We leave the
145 harder problem of fitting an IPLM to partially observed time-series data from overlapping generations – which
146 often arises in studies of natural insect populations – to future work.

147 When within-stage development is unmeasurable, the regression approach for fitting size-based IPM-kernels
148 is unfeasible. Instead, we take a SDDM inspired approach where the likelihood of observed stage-duration data

149 y_S depends on the probabilities (given Θ_S) of surviving and completing stage S in each time-step. We calculate
 150 these probabilities by iterating the discretised IPM. Specifically, we: **1)** initialise $\mathbf{n}_S(t=0) = (1, 0, \dots, 0)^T$ and
 151 $c_S = 0$; **2)** project $\mathbf{n}_S(t)$ forward (eqn 5); **3)** for each t , record the matured proportion c_S , and the loss of density
 152 over the vector $(\mathbf{n}_S^T, c_S)_t$, to construct a sojourn-mortality distribution (Fig. 1); and **4)** use these probabilities to
 153 evaluate the likelihood of data y_S given Θ_S . For marked data, individual-level covariates or random effects (e.g.
 154 individual qualities) can be included via individual-specific kernel calculations.

155 Heterogeneity in developmental rates – represented by the increments Δ – is modelled by the beta distribution
 156 (see Appendix S1) with PDF

$$f(\Delta|\theta) = \frac{\Delta^{\alpha_1-1}(1-\Delta)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)}, \quad \text{eqn 7}$$

157 where $\theta = \{\alpha_1, \alpha_2\}$ are parameters and $B(\cdot, \cdot)$ is the beta function. Bi-modality is avoided by constraining α_1
 158 and α_2 to be greater than one. Since α_1 and α_2 do not yield biological interpretation, we use the alternative
 159 parameterisation $\theta = \{\mu, \kappa\}$, where $\mu = E[\Delta] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$ is the expected developmental increment and $\kappa \in (0, 1)$
 160 is a scale parameter such that $\text{Var}(\Delta) = \kappa\mu(1 - \mu)$. The probability p_l of completing l discrete increments in a
 161 time-step is thus

$$p_l = F\left(\frac{l+1}{r+1}|\theta\right) - F\left(\frac{l}{r+1}|\theta\right), \quad \text{eqn 8}$$

162 where $F(\Delta|\theta)$ is the CDF associated with $f(\Delta|\theta)$.

163 We adopt a Bayesian approach to estimate stage-specific parameters $\{\mu, \kappa, \nu, r\}$. Throughout, we use standard
 164 uniform priors for $\{\mu, \kappa, \nu\}$ and the prior $\text{Uniform}(0, R_{\text{Max}})$ for r , where R_{Max} is a maximum resolution chosen to
 165 be large enough to optimise model fit but small enough to maintain computational efficiency. Markov chain Monte
 166 Carlo (MCMC) (Gelman *et al.*, 2003) enables sampling posteriors of the form

$$f(\mu, \kappa, \nu, r|\mathbf{y}) \propto f(\mu)f(\kappa)f(\nu)f(r)f(\mathbf{y}|\mu, \kappa, \nu, r), \quad \text{eqn 9}$$

167 where \mathbf{y} represents a set of independent data sets (Appendix S9).

168 We demonstrate these techniques with two case studies. The first, a simulation study, tests our parameter

estimation procedure with an IPLM model in which (time-fixed) random variables, quantifying individual quality, generate correlated sojourn-times. The second, constructs a temperature-dependent IPLM for a poikilothermic population by estimating development kernels from laboratory cohort data at different temperatures and using non-parametric regression to estimate non-linear effects of temperature on the development kernel. The resulting model is used to predict the effects of development heterogeneity on asymptotic and transient dynamics under seasonally varying temperatures. NIMBLE (NIMBLE Development Team, 2016; De Valpine *et al.*, 2016) and R (R Core Team, 2013) code for both studies is available (see Supporting Information).

3 Case study I: simulation study with correlated stage-durations

To test the identifiability of IPLM parameters, a simulation-estimation experiment was conducted. Motivated by recent directions in eco-evolution, the basic IPLM model was modified to incorporate correlated stage-durations arising from heterogeneous individual qualities. Quality parameters are used in eco-evolution to parsimoniously quantify net effects of genetic or local factors on vital rates (Wilson and Nussey, 2010; Vindenes and Langangen, 2015) and generate correlated stage-durations (De Valpine, 2009). Despite much theoretical work, the estimation of individual qualities, associated sojourn-mortality distributions and their evolutionary consequences in real populations remains challenging. Here, we outline how Gaussian copulas (Kruskal, 1958) enable individual qualities to condition IPLM kernels, and demonstrate that, even with modest sample sizes, a quality-dependent IPLM fitted to simulated marked-cohort data for just two sequential stages can accurately reproduce sojourn-mortality distributions.

Quality conditioned development. Copulas are tools for modelling correlations in arbitrary sets of random variables (Hougaard, 2012). Here, individual quality (q) and development increments (Δ) are correlated via Gaussian copulas. We assume q is fixed through an individual's lifespan and conditions the distribution of increments at each time-step.

Let $f_{\Delta}(\Delta|\alpha_{\Delta})$ and $F_{\Delta}(\Delta|\alpha_{\Delta})$ denote the marginal (beta) PDF and CDF of developmental increments, with parameters α_{Δ} . Let $f_q(q|\alpha_q)$ denote the marginal PDF of individual qualities, with CDF $F_q(q|\alpha_q)$ and parameters α_q . We assume q follows a standard uniform distribution, noting that any other distribution could be derived via a probability integral transform. A Gaussian copula with correlation ρ was used to establish the joint distribution $f(\Delta, q, |\alpha_{\Delta}, \alpha_q, \rho)$ while preserving the specified marginal distributions (Appendix S2). This implicitly defines the conditional distribution of development increments given quality, $f_{\Delta|q}(\Delta|\alpha_{\Delta}, q, \rho)$, and its corresponding

197 CDF, $F_{\Delta|q}(\Delta|\alpha_{\Delta}, q, \rho)$. The later provides the matrix elements p_l given q :

$$p_l|q = F_{\Delta|q}(\frac{l+1}{r+1}|\alpha_{\Delta}, q, \rho) - F_{\Delta|q}(\frac{l}{r+1}|\alpha_{\Delta}, q, \rho). \quad \text{eqn 10}$$

198 Since stage-durations of individuals are correlated via q , unique IPM-kernels are required for each individual at
199 each stage.

200 **Simulation.** A two-stage development study, following $N = 50$ individuals for $t_c = 21$ days, was simulated
201 500 times. In each simulation, $\{\mu_1, \kappa_1, \nu_1, r_1, \mu_2, \kappa_2, \nu_2, r_2\}$, $\mathbf{q} = \{q_1, \dots, q_N\}$ and ρ were drawn initially
202 from $\text{Unif}(0, 1)$, r_1 and r_2 from $\text{Unif}(0, 100)$, and rejection sampling (Appendix S3) was used to ensure: **1)**
203 all parameters α of the density $f_{\Delta}(\Delta|\alpha_1, \alpha_2)$ were greater than one, and **2)** the number of individuals completing
204 both stages was greater or equal to $N_{\min} = 35$.

205 The fate of each individual i in each stage was described by two pieces of information: 1) the time-to-event,
206 $y_{iA} \in \{1, \dots, t_{\max}\}$; and 2) the event-type, $y_{iB} \in \{\text{stage completion, mortality, censor}\}$. Probabilities associated
207 with combinations of y_{iA} and y_{iB} were obtained as follows. The probabilities $p_l|q_i$ permitted conditional (on
208 quality) construction of the IPM-approximation, eqn 5. The unit pulse vector $(\mathbf{n}^T, c_s)_0 = (1, 0, \dots, 0)^T$ was
209 projected to give, for each time-step $t \in \{0, \dots, t_{\max}\}$, the probabilities to complete a stage, $p_d(t|q_i)$, or to die,
210 $p_m(t|q_i)$. The probability of right-censor beyond t_{\max} is $p_{ci} = 1 - \sum_{t=1}^{t_{\max}} (p_d(t|q_i) + p_m(t|q_i))$. For stage 1,
211 $t_{\max} = t_c$. Death, or right-censor at t_c , in stage 1 imposes that $p_{ci} = 1$ for stage 2. Otherwise, for stage 2, t_{\max} is
212 $t_{\max}^{(i)} = t_c - t_{s_2}^{(i)}$, where $t_{s_2}^{(i)}$ is the time-step that individual i enters stage 2. The probabilities $p_d(t|q_i)$, $p_m(t|q_i)$ and
213 p_{ci} define the right-censored sojourn-mortality time distribution for individual i in a given stage. Individual-level
214 data were sampled from the categorical distribution,

$$(y_{iA}, y_{iB}) \sim \text{Categorical}(p_d(1|q_i), p_m(1|q_i), \dots, p_d(t_{\max}|q_i), p_m(t_{\max}|q_i), p_{ci}). \quad \text{eqn 11}$$

215 For each simulation and stage, the "true" mean ($\tilde{\mu}$) and standard deviation ($\tilde{\sigma}$) of maturation-times and the
216 probability to survive to maturation ($\tilde{\nu}$) were calculated (Appendix S4) from probabilities $p_d(t)$ and $p_m(t)$, using
217 the marginal $F_{\Delta}(\Delta|\alpha_{\Delta})$ to evaluate eqn 8.

218 **Estimation.** For each simulation, the posterior distribution

$$f(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\nu}, \mathbf{r}, \mathbf{q}, \rho | \mathbf{y}_{S_1}, \mathbf{y}_{S_2}) \propto f(\rho) f(\mathbf{q}) \prod_{s \in \{S_1, S_2\}} f(\mu_s) f(\kappa_s) f(\nu_s) f(r_s) \prod_{i=1}^N f(y_{is} | \mu_s, \kappa_s, \nu_s, r_s, \rho, q_i). \quad \text{eqn 12}$$

was sampled using the default block Metropolis-Hastings sampler in NIMBLE. Thinning was set to twice the minimum expected sample size obtained from pre-runs (Appendix S9). Thereafter, 10^4 thinned MCMC samples were generated and convergence diagnostics were performed using CODA (Plummer *et al.*, 2006).

Estimation performance. Posterior medians and 95% credibility intervals (CI_{95}) of the means ($\hat{\mu}$), standard deviations ($\hat{\sigma}$) and total survivals ($\hat{\nu}$) of the joint sojourn-mortality distribution were plotted against true values (Fig. 2). Medians were distributed evenly around the 1:1 line and uncertainty was sufficiently small to suggest that the sojourn-mortality distribution approximations were accurate given the sample size. Precision was greatest when $\tilde{\mu}$ and $\tilde{\sigma}$ were small. The CI_{95} s of estimated parameters enveloped true values in approximately 95% of simulations. Banding in the posteriors for $\tilde{\nu}$ arose from the limited possibilities regarding the number of individuals completing both stages.

In general, the CI_{95} s of estimated values for r_1 , r_2 , ρ and $\{q_1, \dots, q_N\}$ enveloped the true values (Figs. S1, S2). Uncertainty was larger for these parameters than for $\tilde{\mu}$, $\tilde{\sigma}$ and $\tilde{\nu}$. The largest CI_{95} s for qualities q_i were associated with individuals that died in stage 1, and the greatest precision was achieved when sojourn-times were right censored (Fig. S2). For fully developed individuals, quality estimates ranged greatly in precision. True vs. fitted values of ρ showed that, despite uncertainties in the q_i , relatively accurate estimates for ρ were obtained. The posterior median and CI_{95} s for resolution were clustered in horizontal bands (Fig. S1), suggesting that model fit was not sensitive to resolution so long as resolution was not too small. This implies that very large values of r can be computationally superfluous since even relatively low resolutions can yield sojourn-mortality distributions as accurate as can be supported by the data.

These results highlight that quality-conditioned IPLMs can successfully model development and survival in marked cohort studies. A two-stage study with $N = 50$ was used here to indicate what can be possible with a typical data set from a small experiment. Naturally, a greater number of individuals or stages would increase precision – an important consideration regarding the design of experiments to parameterise eco-evolutionary models. Most importantly, we show that even when within-stage development is unmeasurable, realistic IPM-based matrix models can be fit using maturation-time data.

244 4 Case study II: seasonal and transient dynamics of biting midges

245 Current methods in ecology fail to scale up realistically from laboratory development studies to field predictions
 246 of population dynamics. This is mainly because they do not provide valid transition probabilities for populations
 247 with vital rates sensitive to exogenous sources of variation. For example, poikilothermic Ecdysozoa (animals that
 248 shed exoskeletons) exposed to varying temperatures are rarely modelled using Lefkovich matrices because these
 249 models do not track stages accurately under time-varying vital rates. Furthermore, while IPM is regularly used
 250 to model the dynamics of wild vertebrate or plant populations, it is rarely used to model Ecdysozoan populations
 251 subject to time-varying parameters. Indeed, for many Ecdysozoa it can be prohibitively difficult to obtain appropriate
 252 within-stage development data to fit IPMs with time-lagged regression.

253 Classically, the dynamics of poikilotherms are modelled using degree-day accumulation (DDA), a physiological
 254 unit capturing cumulative metabolic responses to temperature (De Reaumur, 1735; Belehradek, 1935). Maximum
 255 likelihood estimators are available for a stochastic DDA model (Dennis *et al.*, 1986). However, that model neglects
 256 mortality, doesn't yield stage-specific parameterisation, requires developmental homogeneity at time zero, uses
 257 non-monotonic DDA and, as for most DDA models, assumes a linear temperature–development relationship.
 258 Although linearity works over small temperature ranges (Bonhomme, 2000), non-linearity becomes important
 259 when temperature fluctuations gain amplitude (Lobell *et al.*, 2011).

260 Non-linear degree-day models are available (Briere *et al.*, 1999) and are used to parameterise IBMs (Régnière
 261 and Powell, 2013). This framework emphasises fitting non-linear expected response curves and treats variance as
 262 a nuisance parameter. Often, proportionality between SDD mean and standard deviation is assumed (Sharpe and
 263 DeMichele, 1977), and the covariates or stochastic processes that generate variance are neglected. Moreover,
 264 proponents neglect that mortality modifies SDDs, and either estimate survival by neglecting SDD shape and
 265 variance (Régnière *et al.*, 2012) or neglect mortality entirely (Yurk and Powell, 2010). With IPLM models, SDD
 266 variance arises naturally from a stochastic development-mortality process. Furthermore, the assumptions used for
 267 estimation and simulation are identical, thereby eliminating potential bias arising from model mismatch.

268 Here, we fit a temperature-dependent IPLM to unmarked maturation-time data for biting midges of the genus
 269 *Culicoides* at fixed temperatures. We model IPM-kernel parameters as a function of temperature using non-
 270 parametric regression – the model is fit using biologically justified unimodal constraints, and unimodal spline
 271 interpolation determines parameters at unmeasured temperatures. The interpolated model is used to analyse
 272 asymptotic and transient dynamics under fixed and seasonal temperatures.

Culicoides IPLM model. *Culicoides* biting midges attract considerable interest as vectors of numerous viral diseases (Mellor *et al.*, 2000). Modelling has provided empirical descriptions of flight-trap data for phenology, bio-geography or epidemiological risk studies (Sanders *et al.*, 2011; Guis *et al.*, 2012; Searle *et al.*, 2012; Diarra *et al.*, 2015). But these approaches cannot provide all the vital rates required for incorporating vector life-cycle dynamics in mechanistic epidemiological models.

While insufficient *Culicoides* within-stage trait data (i.e. size, weight) exists for time-lagged regression, sufficient maturation-time data exist for fitting temperature-dependent IPLMs. Consider the female only egg-larva-pupa-adult model

$$\begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_t = \begin{bmatrix} \mathbf{W}_E & \mathbf{0} & \mathbf{0} & \mathbf{F}_A \\ \mathbf{B}_E & \mathbf{W}_L & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_L & \mathbf{W}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_P & \mathbf{G}_A \end{bmatrix}_t \begin{bmatrix} \mathbf{E} \\ \mathbf{L} \\ \mathbf{P} \\ \mathbf{A} \end{bmatrix}_{t-1}, \quad \text{eqn 13}$$

where $\mathbf{G}_A = \mathbf{W}_A + \mathbf{B}_A$ models multiple gonotrophic cycles. We fitted this model to cohort data from *C. variipennis* egg, pupae and combined larvae-pupae development studies (Mullens and Rutz, 1983; Vaughan and Turner, 1987), and to individual-level data from *C. nubeculosus* fecundity, gonotrophic cycle and egg stage-duration studies (Balenghien *et al.*, 2016). Details of each data set are given in Appendix S5. Note, these species share similar developmental responses across the 15°C-35°C range (Purse *et al.*, 2015).

For each stage, temperature-dependence was modelled using unimodal splines on survival $\nu(T)$, the 1st and 99th percentiles ($\mathcal{P}_1(T)$ and $\mathcal{P}_{99}(T)$) of the developmental rate distribution $f_\Delta(\Delta|\alpha)$, and (for adults) fecundity $F_A(T)$. For this, unimodality constraints on the responses to temperature of these parameters were incorporated into the MCMC. The unimodality constraints ensure an optimal temperature for each stage (Sharpe and DeMichele, 1977; Régnière *et al.*, 2012) and permit shape-constrained interpolation at unsampled temperatures. Interpolation was performed, for each line of MCMC output, using unimodal cubic Hermite splines (Appendix S10). This non-parameteric regression produces a smoothed unimodal curve analogous to (the piece-wise linear) multivariate adaptive regression splines (Friedman, 1991). Note, the spline modelling for Δ was performed on percentiles rather than μ and κ in order to enforce a unimodal response to temperature. Only two percentiles were needed to identify μ and κ and the 1st and 99th proved a practical choice.

Estimation. Details of likelihoods used for model fitting, and missing-value imputation steps, are given in Appendices S6 and S7 respectively. Posteriors were sampled using parallel tempering (Swendsen and Wang, 1986) (Appendices S8, S9). Ten thousand thinned MCMC samples were generated, with thinning and convergence

diagnostics performed as in case study I. Estimates of $\mu(T)$ and $\kappa(T)$ were obtained via back-transformation of the interpolated $\mathcal{P}_{01}(T)$ and $\mathcal{P}_{99S}(T)$ (Appendix S10).

Resolution. Posterior likelihoods were consistently poor at $r_s = 1$, and the CLM's posterior probability was negligible (Fig. 3). Small increases in r_s yielded large likelihood gains because the CLM's geometric sojourn-time distributions do not fit the data well and IPLM improves fitted sojourn-time distributions in all cases (Figs. S3, S4). Maximum *a posteriori* (MAP) estimates for resolution were $r_E^{(\text{MAP})} = 6$, $r_L^{(\text{MAP})} = 9$, $r_P^{(\text{MAP})} = 31$ and $r_A^{(\text{MAP})} = 8$, with variable levels of uncertainty.

Differential Responses to Temperature. Stages differed in their responses to temperature (Figs. 4, S5). Generally, the mean and variance of developmental rates increased with temperature. However, eggs, and to a lesser extent pupae, exhibited impaired development at high temperatures. Survival was low at the highest temperatures for eggs, pupae and adults. Larvae experienced relatively high survival at all temperatures and were the most resistant stage to cold – this concurs with field reports of over-wintering success being greatest for larvae (Kettle, 1962). Nevertheless, the stable stage distribution (dominant eigenvector) did not exhibit clear visual evidence of strong temperature-dependence within the range of experimental temperatures (Fig. S7).

Asymptotic dynamics at fixed temperatures. The asymptotic growth rate (dominant eigenvalue λ_1) over a 10°C-40°C range was similar for CLM and IPLM: both models suggested temperatures in the mid-twenties optimise growth, although CLM systematically predicted higher growth rates than IPLM (Fig. S6). Both models predicted population decline ($\lambda_1 < 1$) at high temperatures. The range of temperatures yielding $\lambda_1 > 1$, and uncertainties regarding growth–decline threshold temperatures, were greater for CLM than for IPLM.

Transient dynamics. To investigate potential effects of temperature perturbations, various indices of transient dynamics (Stott *et al.*, 2011) were quantified. The duration of transient dynamics is largely determined by the *damping ratio* of first and second eigenvalues $\rho = \lambda_1/|\lambda_2|$. Plots of ρ^{-t} indicated slower convergence for IPLM than for CLM at all temperatures (Fig. 5). Thus, the relative importance of λ_2 increased when within-stage developmental heterogeneity was included, and CLM underestimated the duration of transients at every temperature. Estimates of maximum amplification (amp_{\max}), maximum attenuation (att_{\max}) and associated inertias were all affected when within-stage developmental heterogeneity was excluded, and CLM consistently underestimated the amplitude of transient oscillations (Fig. 5).

Seasonal dynamics. Resolution effects on seasonal dynamics were explored by projecting daily growth rate, relative density (density divided by annual growth λ_Y), amp_{\max} and att_{\max} over two years. Seasonal temperatures were modelled as $T_t = v + w \cos\left(t \frac{2\pi}{365}\right)$, where v and w were set such that $\min(T_t) = 15^\circ\text{C}$ and $\max(T_t) = 25^\circ\text{C}$,

with $t = 0$ the coldest day of the year. The initial population was set to the stable distribution associated with 15°C . The amplitude of annual oscillations in λ_1 and relative density were greater for CLM than IPLM (Fig. 6). Trajectories of amp_{\max} and att_{\max} were more complex for IPLM and exhibited spring-time oscillations in the first year that were damped in the second year. This damping suggests that spring-time flux in the stable stage distribution was mild. Raising $\max(T_t)$ to 30° reduced both precision in λ_1 and the probability of $\lambda_1 > 1$ in mid-summer (Fig. 6). This possibility of negative summertime growth arose from uncertainty in adult survival at 30°C .

These analyses suggest that, for the chosen temperature profile, the importance of transient dynamics relative to asymptotic dynamics is small. Despite vital rates responding differentially to temperature, the perturbations generated by the chosen temperature profile only generate low-amplitude transient oscillations. The amplitude of transient oscillations is expected to increase as winter-summer, or day-to-day, temperature differences increase since cold winters exert strong differential mortality. Temperature transfer experiments (Régnière *et al.*, 2012) could help test this hypothesis.

5 Discussion

Integral projection Lefkovitch matrices (IPLMs) are a new tool for modelling the dynamics of stage-structured populations. They augment classic Lefkovitch matrices (CLMs) by modelling within-stage dynamics with integral projection models (IPMs). By doing so, stage-specific vital rates, and related metrics, can be parameterised with time-varying covariates to yield more realistic sojourn-mortality distributions. Moreover, IPLMs can provide valid transition probabilities when stage distributions are non-stable, reduce errors in transient or non-linear dynamics and can therefore improve predictive performance when exogenous factors differentially affect vital rates.

The kernel of any IPM-based model must synthesise the net effects of interacting endogenous and exogenous processes on vital rates. We have shown that maturation-time data provides an alternative to dynamic trait data for fitting IPM-kernels. This enables IPM parameterisation even when measuring within-stage development is unfeasible or impracticable.

Trait heterogeneity is fundamental to eco-evolutionary models, and static traits, such as *quality*, are often used to condition vital rates (Vindenes and Langangen, 2015). Yet, ecologists lack tools for tracking many important traits in natural populations (De Valpine *et al.*, 2014). We have shown how unmeasured individual traits can condition IPLM kernels to model correlated stage-durations. Although we call these individual traits *quality*,

these parameters can be interpreted to provide: group-level random effect when modelling laboratory data; spatial random effects conditioning phenotypic responses of sub-populations to unmeasured local factors; a synthetic index of genotypic traits that affect vital rates and fitness in eco-evolutionary models.

In our *Culicoides* analyses, neglecting within-stage heterogeneity generated a small systematic bias favouring over-estimation of growth rates and the probability of $\lambda_1 > 1$. Thus, neglecting within-stage heterogeneity can apparently affect predictions of potential ecological niche. It is increasingly recognised that perturbations and transient dynamics can be as important as asymptotic dynamics (Stott *et al.*, 2011). The CLM model underestimated the duration and amplitude of transient oscillations, the potential range of relative densities (att_{\max} , amp_{\max}) and associated inertias. Whereas this model yields just one pair of complex eigenvalues, the larger IPLM yields many more complex eigenvalues giving a richer characterisation of the transient oscillations that follow perturbation. These effects can have important consequences in wildlife management and other branches of ecology and evolution where perturbations limit the forecasting horizon of current methods.

Our *Culicoides* study represents the first time an IPM has been used to analyse temperature effects on within-stage development, transient dynamics and phenology for a poikilothermic Ecdysozoa. Although tracking within-stage development with temperature-dependent IPMs is analogous to tracking degree-day accumulation, IPMs and degree-day accumulation models have hitherto evolved in relative isolation. The use of IPLMs in the *Culicoides* study bridges a historic gap between these schools of ecological modelling, overcomes many of the limitations of the pioneering work of Dennis *et al.* (1986), and avoids popular, but unrealistic, linearity assumptions by readily accommodating non-linear responses to temperature. We modelled these responses at unmeasured temperatures using biologically justified unimodal spline interpolation. Alternatively, mechanistic link functions (Régnière *et al.*, 2012; Smallegange *et al.*, 2016) could have been used to obtain smoother development–temperature response curves. However, our non-parameteric approach provided greater parsimony in the *Culicoides* study. The model’s relative simplicity, and ability to exploit maturation-time data, suggest that IPLM provides a valuable tool for modelling the dynamics of many stage-structured populations.

Discretisation unavoidably introduces resolution parameters r . Simulation results suggested that, provided r is large enough, model fit can be relatively insensitive to r . By contrast, the *Culicoides* study showed that the degree of sensitivity of likelihoods to r is data dependent. Indeed, at low values, r functions as a shape parameter, suggesting that more flexible kernels should reduce sensitivity to r . Modelling development rate heterogeneity with beta distributions allowed us to demonstrate how the likelihoods of CLMs are greatly increased with just a few additional parameters. But, we do not expect this distribution to be optimal in all situations. Further research is

required to test alternatives, such as probability distributions with more parameters or semi-parametric methods.

Future research should address how parameter estimation could exploit alternative data sources. For example, could time series data from field studies reduce parameter uncertainty beyond the range of experimental conditions? This is challenging because such data can reflect overlapping generations and parts of a life cycle can be hidden. Bayesian data augmentation, where laboratory data informs the imputation of missing field data, can provide a powerful tool. However, the problem of calibrating population models to time series data lies at the cutting edge of statistical ecology (Andrieu *et al.*, 2010; Wood, 2010; Scranton *et al.*, 2014). Since, generic methods for fitting IPLM models to such data would provide a major step towards improving the forecasts of matrix models, we strongly encourage further research in that direction.

6 Conclusion

We have presented a new matrix model for stage-structured populations with non-stable stage distributions. A CLM is augmented with stage-specific IPMs that track within-stage development, permitting parsimonious parameterisation, even with time-varying covariates. The IPM-kernels are estimated from maturation-time data, enabling IPM methods to be used in many scenarios that were not previously possible. The resulting model is simple, reduces projection error, can be analysed using the powerful matrix model toolbox, and is expected to improve the forecast horizon of stage-structured models in many branches of ecology and evolution.

7 Acknowledgments

The authors would like to thank Bradley Mullens and members of ASTRE's entomology team, including Thomas Balenghien, Xavier Allène, Ignace Rakotoarivony, Jonathan Lhoir and Karine Huber for sharing their *Culicoides* data. We would like to thank Claire Guinant, Mathieu Poli, Renaud Lancelot, Jérémy Bouyer, Claire Garros, Jean Vaillant, Nonito Pages, Jonathan Gordon, Simon Carpenter and Elizabetha Vergu for fruitful discussions at various stages during the development of this work. We are particularly grateful to Daniel Turek and Christoper Parciorek for their technical support with NIMBLE. We are grateful to INRA's bioinformatics platform (<http://migale.jouy.inra.fr>), and to the Centre Commun de Calcul Intensif (<http://calamar.univ-ag.fr/c3i>), for providing computational resources, help and support. This work was partially funded by: CIRAD's doctoral fellowship program; the DGAL (Direction générale de l'alimentation) of the French Ministry of Agriculture; the "INVERNESS" project of INRA's

413 “Integrated Management of Animal Health” meta-program (www.gisa.inra.fr/en); the Institut Carnot Animal
 414 Health (ICSA) project “Integrated Vector Management: Innovating to Improve Control and Reduce Environmental
 415 Impacts” (<http://www6.jouy.inra.fr/gabi/Le-Partenariat/L-Institut-Carnot-Sante-Animale-ICSA>); and from EU grant
 416 FP7-261504 EDENext. The work is catalogued as EDENext12345 (<http://www.edenext.eu>).

417

418 **References**

- 419 Andrieu, C, Doucet, A, and Holenstein, R (2010). Particle Markov chain Monte Carlo methods. *Journal of the*
 420 *Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.
- 421 Balenghien, T, Allene, X, Lhoir, J, Rakotoarivony, I, and Huber, K (2016). Data from: Lefkovitch matrices
 422 meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental
 423 processes. Dryad Digital Repository. doi:123456789/dryad.123456789 [UPLOAD PENDING].
- 424 Belehradek, J (1935). *Temperature and Living Matter*. Borntraeger, Berlin.
- 425 Berezansky, L, Braverman, E, and Idels, L (2010). Nicholson’s blowflies differential equations revisited: main
 426 results and open problems. *Applied Mathematical Modelling*, 34(6), 1405–1417.
- 427 Bierzychudek, P (1999). Looking backwards: Assessing the projections of a transition matrix model. *Ecological*
 428 *Applications*, 9, 1278–1287.
- 429 Birt, A, Feldman, R.M, Cairns, D.M, Coulson, R.N, Tchakerian, M, Xi, W, *et al.* (2009). Stage-structured matrix
 430 models for organisms with non-geometric development times. *Ecology*, 90, 57–68.
- 431 Blythe, S, Nisbet, R, and Gurney, W (1984). The dynamics of population models with distributed maturation
 432 periods. *Theoretical Population Biology*, 25, 289–311.
- 433 Bolnick, D.I, Amarasekare, P, Araújo, M.S, Bürger, R, Levine, J.M, Novak, M, *et al.* (2011). Why intraspecific
 434 trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- 435 Bonhomme, R (2000). Bases and limits to using degree-day units. *European Journal of Agronomy*, 13, 1–10.

- 436 Briere, J.F, Pracros, P, Le Roux, A.Y, and Pierre, J.S (1999). A novel rate model of temperature-dependent
437 development for arthropods. *Environmental Entomology*, 28(1), 22–29.
- 438 Caswell, H (2006). *Matrix Population Models*. 2nd Edition. Sinauer Associates, Sunderland, MA.
- 439 Coulson, T, MacNulty, D.R, Stahler, D.R, vonHoldt, B, Wayne, R.K, and Smith, D.W (2011). Modeling effects of
440 environmental change on wolf population dynamics, trait evolution, and life history. *Science*, 334, 1275–1278.
- 441 Cushing, J.M, Costantino, R.F, Dennis, B, Desharnais, R, and Henson, S.M (2002). *Chaos in Ecology:
442 Experimental Nonlinear Dynamics*. Elsevier, New York.
- 443 De Reaumur, R.A.F (1735). Observations du thermomètre, faites á Paris pendant l’année 1735, comparés avec
444 celles qui ont été faites sous la ligne, á l’Isle de France, á Alger et en quelques-unes de nos isles de l’Amérique.
445 *Paris: Mémoires de l’Académie des Sciences*.
- 446 De Valpine, P (2009). Stochastic development in biologically structured population models. *Ecology*, 90, 2889–
447 2901.
- 448 De Valpine, P and Knappe, J (2015). Estimation of general multistage models from cohort data. *Journal of
449 Agricultural, Biological, and Environmental Statistics*, 20, 140–155.
- 450 De Valpine, P, Scranton, K, Knappe, J, Ram, K, and Mills, N.J (2014). The importance of individual developmental
451 variation in stage-structured population models. *Ecology Letters*, 17, 1026–1038.
- 452 De Valpine, P, Turek, D, Paciorek, C.J, Anderson-Bergman, C, Lang, D.T, and Bodik, R (2016). Programming with
453 models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational
454 and Graphical Statistics*, (accepted), 1–28.
- 455 Dennis, B, Kemp, W.P, and Beckwith, R.C (1986). Stochastic model of insect phenology: estimation and testing.
456 *Environmental Entomology*, 15, 540–546.
- 457 Diarra, M, Fall, M, Lancelot, R, Diop, A, Fall, A.G, Dicko, A, Seck, M.T, Garros, C, Allène, X, Rakotoarivony,
458 I, *et al.* (2015). Modelling the abundances of two major *Culicoides* (Diptera: Ceratopogonidae) species in the
459 Niayes area of Senegal. *PloS One*, 10(6), e0131021.
- 460 Easterling, M.R, Ellner, S.P, and Dixon, P.M (2000). Size-specific sensitivity: applying a new structured population
461 model. *Ecology*, 81, 694–708.

- 462 Ellner, S.P and Rees, M (2007). Stochastic stable population growth in integral projection models: theory and
463 application. *Journal of Mathematical Biology*, 54, 227–256.
- 464 Friedman, J.H (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67.
- 465 Gelman, A, Carlin, J.B, Stern, H.S, and Rubin, D.B (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC,
466 Boca Raton, FL.
- 467 Guis, H, Caminade, C, Calvete, C, Morse, A.P, Tran, A, and Baylis, M (2012). Modelling the effects of past
468 and future climate on the risk of bluetongue emergence in Europe. *Journal of the Royal Society Interface*, 9,
469 339–350.
- 470 Hoeting, J.A, Tweedie, R.L, and Olver, C.S (2003). Transform estimation of parameters for stage-frequency data.
471 *Journal of the American Statistical Association*, 98.
- 472 Hougaard, P (2012). *Analysis of Multivariate Survival Data*. Springer Science & Business Media, New York.
- 473 Kettle, D (1962). The bionomics and control of *Culicoides* and *Leptoconops* (Diptera,
474 Ceratopogonidae=Heleidae). *Annual Review of Entomology*, 7, 401–418.
- 475 Kruskal, W.H (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53,
476 814–861.
- 477 Lefkovitch, L (1965). The study of population growth in organisms grouped by stages. *Biometrics*, pages 1–18.
- 478 Leslie, P.H (1945). On the use of matrices in certain population mathematics. *Biometrika*, pages 183–212.
- 479 Lobell, D.B, Bänziger, M, Magorokosho, C, and Vivek, B (2011). Nonlinear heat effects on African maize as
480 evidenced by historical yield trials. *Nature Climate Change*, 1, 42–45.
- 481 Longstaff, B.C (1984). An extension of the Leslie matrix model to include a variable immature period. *Australian*
482 *Journal of Ecology*, 9, 289–293.
- 483 Manly, B.F (1990). *Stage-Structured Populations: Sampling, Analysis and Simulation*. Springer, Netherlands.
- 484 Mellor, P, Boorman, J, and Baylis, M (2000). *Culicoides* biting midges: their role as arbovirus vectors. *Annual*
485 *Review of Entomology*, 45, 307–340.

- Merow, C, Dahlgren, J.P, Metcalf, C.J.E, Childs, D.Z, Evans, M.E, Jongejans, E, *et al.* (2014). Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution*, 5, 99–110.
- Mullens, B and Rutz, D (1983). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae) at constant laboratory temperatures. *Annals of the Entomological Society of America*, 76, 747–751.
- Murtaugh, P.A, Emerson, S.C, McEvoy, P.B, and Higgs, K.M (2012). The statistical analysis of insect phenology. *Environmental Entomology*, 41, 355–361.
- NIMBLE Development Team (2016). *NIMBLE Users Manual, Version 0.5*.
- Nisbet, R.M (1997). Delay-differential equations for structured populations. In *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*, pages 89–118. Springer, US.
- Ozgul, A, Coulson, T, Reynolds, A, Cameron, T.C, and Benton, T.G (2012). Population responses to perturbations: the importance of trait-based analysis illustrated through a microcosm experiment. *The American Naturalist*, 179, 582–594.
- Petchey, O.L, Pontarp, M, Massie, T.M, Kéfi, S, Ozgul, A, Weilenmann, M, Palamara, G.M, Altermatt, F, Matthews, B, Levine, J.M, *et al.* (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- Plummer, M, Best, N, Cowles, K, and Vines, K (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Purse, B, Carpenter, S, Venter, G, Bellis, G, and Mullens, B (2015). Bionomics of temperate and tropical *Culicoides* midges: knowledge gaps and consequences for transmission of *Culicoides*-borne viruses. *Annual Review of Entomology*, 60, 373–392.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rees, M, Childs, D.Z, and Ellner, S.P (2014). Building integral projection models: a user's guide. *Journal of Animal Ecology*, 83, 528–545.

- 511 Régnière, J, Powell, J, Bentz, B, and Nealis, V (2012). Effects of temperature on development, survival and
 512 reproduction of insects: experimental design, data analysis and modeling. *Journal of Insect Physiology*, 58,
 513 634–647.
- 514 Régnière, J and Powell, J.A (2013). Animal life cycle models (poikilotherms). In *Phenology: An Integrative*
 515 *Environmental Science*, pages 295–316. Springer, Netherlands.
- 516 Sanders, C.J, Shortall, C.R, Gubbins, S, Burgin, L, Gloster, J, Harrington, R, Reynolds, D.R, Mellor, P.S, and
 517 Carpenter, S (2011). Influence of season and meteorological parameters on flight activity of *Culicoides* biting
 518 midges. *Journal of Applied Ecology*, 48, 1355–1364.
- 519 Scranton, K, Knape, J, and de Valpine, P (2014). An approximate Bayesian computation approach to parameter
 520 estimation in a stochastic stage-structured population model. *Ecology*, 95, 1418–1428.
- 521 Searle, K, Blackwell, A, Falconer, D, Sullivan, M, Butler, A, and Purse, B (2012). Identifying environmental
 522 drivers of insect phenology across space and time: *Culicoides* in Scotland as a case study. *Bulletin of*
 523 *Entomological Research*, 1, 1–16.
- 524 Sharpe, P.J and DeMichele, D.W (1977). Reaction kinetics of poikilotherm development. *Journal of Theoretical*
 525 *Biology*, 64, 649–670.
- 526 Smallegange, I.M, Caswell, H, Toorians, M.E, and Roos, A.é.M (2016). Mechanistic description of population
 527 dynamics using dynamic energy budget theory incorporated into integral projection models. *Methods in Ecology*
 528 *and Evolution*.
- 529 Smallegange, I.M, Deere, J.A, and Coulson, T (2014). Correlative changes in life-history variables in response to
 530 environmental change in a model organism. *The American Naturalist*, 183(6), 784–797.
- 531 Stott, I, Townley, S, and Hodgson, D.J (2011). A framework for studying transient dynamics of population
 532 projection matrix models. *Ecology Letters*, 14, 959–970.
- 533 Swendsen, R.H and Wang, J.S (1986). Replica Monte Carlo simulation of spin-glasses. *Physical Review Letters*,
 534 57, 2607.
- 535 Van Tienderen, P.H (1995). Life cycle trade-offs in matrix population models. *Ecology*, pages 2482–2489.

- 536 Vaughan, J and Turner, E (1987). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae)
 537 from Saltville, Virginia, at constant laboratory temperatures. *Journal of Medical Entomology*, 24, 390–395.
- 538 Vindenes, Y and Langangen, Ø (2015). Individual heterogeneity in life histories and eco-evolutionary dynamics.
 539 *Ecology Letters*, 18, 417–432.
- 540 Wilson, A.J and Nussey, D.H (2010). What is individual quality? An evolutionary perspective. *Trends in Ecology*
 541 *and Evolution*, 25, 207–214.
- 542 Wood, S.N (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466, 1102–1104.
- 543 Yamanaka, T, Nelson, W.A, Uchimura, K, and Bjørnstad, O.N (2012). Generation separation in simple structured
 544 life cycles: Models and 48 years of field data on a tea tortrix moth. *The American Naturalist*, 179(1), 95–109.
- 545 Yurk, B.P and Powell, J.A (2010). Modeling the effects of developmental variation on insect phenology. *Bulletin*
 546 *of Mathematical Biology*, 72(6), 1334–1360.

547 SUPPORTING INFORMATION

548 Additional details may be found online in the Supporting Information tab for this article.

549 8 Figures

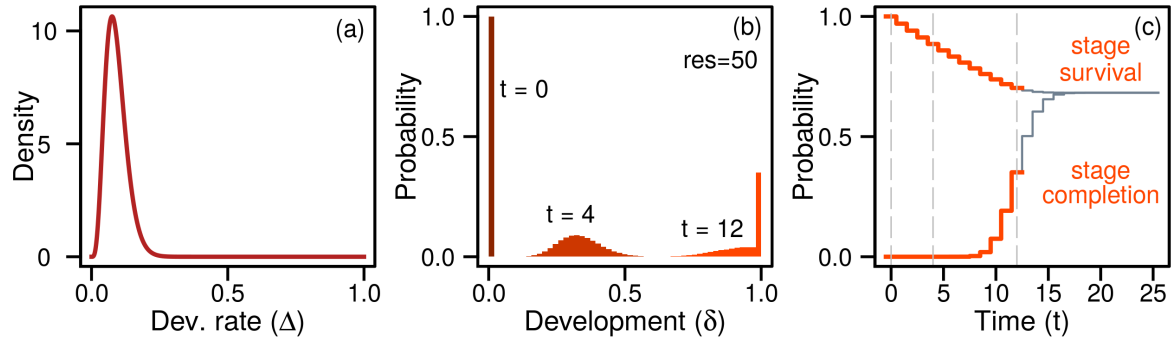


Figure 1. A basic integral projection model (IPM) for within-stage dynamics. The stage-specific IPM-kernel K_{Θ} is defined as the product of survival probability ν and a probability density function (PDF) for development increments Δ . Here, $\nu = 0.97$ and the PDF is $\text{Beta}(\alpha_1 = 5, \alpha_2 = 50)$ (a). A population, initialised ($t = 0$) in the first sub-stage, is projected forward through a series of ($r = 50$) discrete developmental increments (b). The accumulation of density in the final sub-stage, and the loss of density over all sub-stages, generates the probabilities to complete the stage, or die, per time-step. The cumulative probabilities are shown with the interval $t = 0$ to $t = 12$ coloured orange (c). Dashed lines (c) correspond to the developmental distributions (b) at $t = 0$, $t = 4$ and $t = 12$.

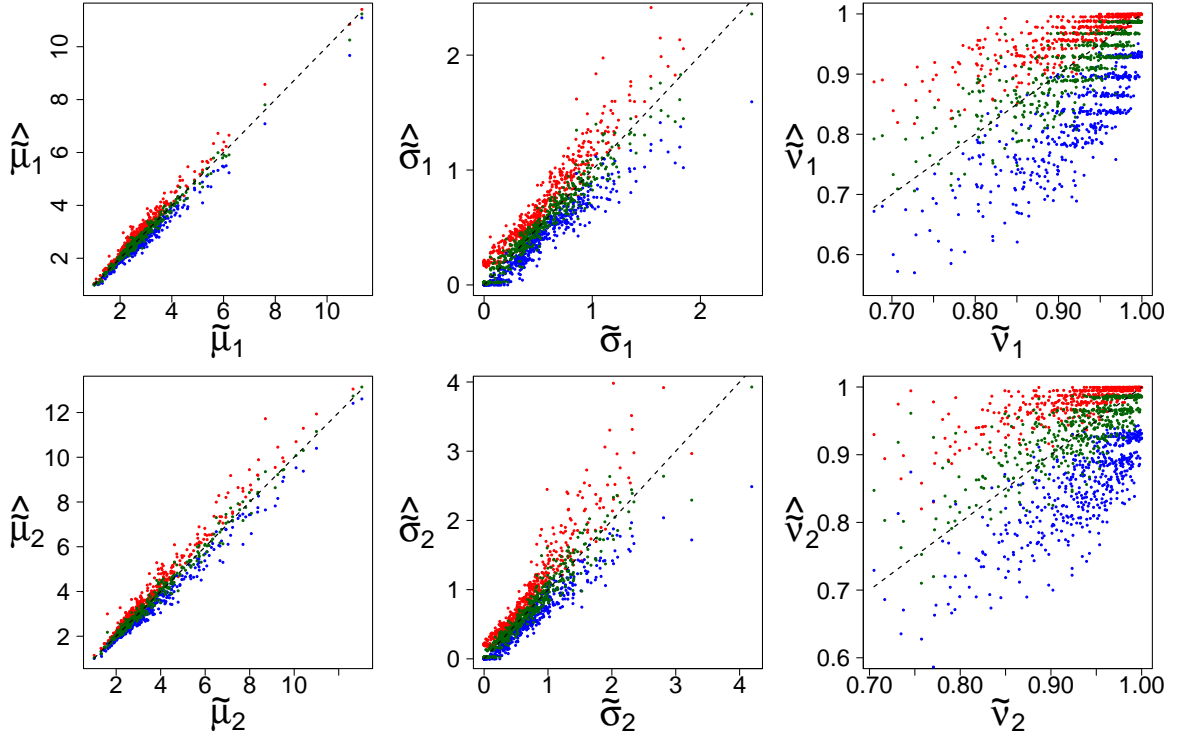


Figure 2. Estimated (y-axes) vs. true (x-axes) mean ($\tilde{\mu}$), standard deviation ($\tilde{\sigma}$) and total survival probability ($\tilde{\nu}$) of sojourn-mortality distributions from 500 simulations of a two-stage quality-conditioned IPLM model with $N = 50$ individuals. For each simulation, the median (green) and upper (red) and lower (blue) CI_{95} bounds of the posterior distribution are presented. The one-one (dashed) line is shown for reference. The number of outliers where CI_{95} s failed to envelope the true parameters were: 25 ($\tilde{\mu}_1$), 32 ($\tilde{\mu}_2$), 20 ($\tilde{\sigma}_1$), 29 ($\tilde{\sigma}_2$), 20 ($\tilde{\nu}_1$) and 28 ($\tilde{\nu}_2$).

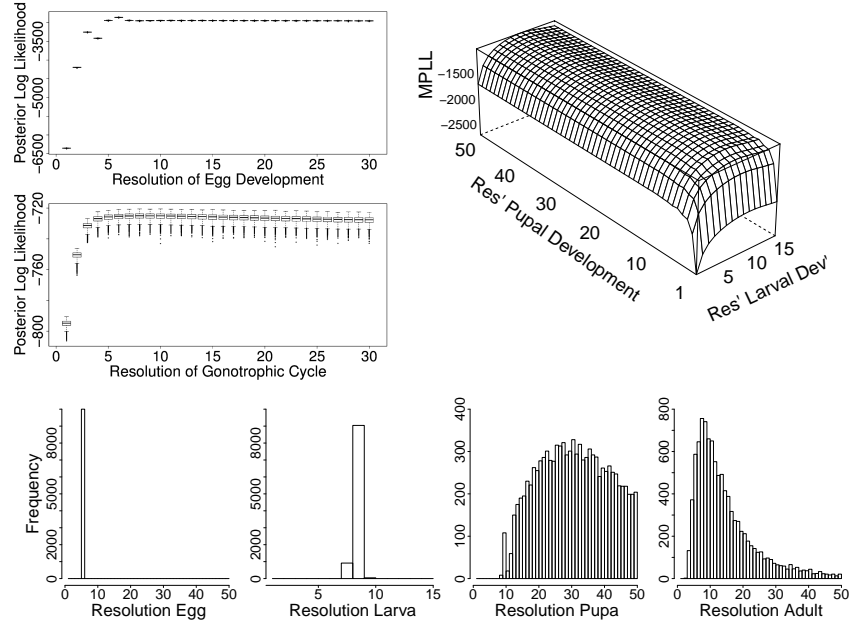


Figure 3. Posterior log-likelihood profiles w.r.t. resolution (top) and the distribution of estimates for each resolution (bottom) for an egg, larvae, pupae and adult IPLM model fitted to *Culicoides* biting midge data. Boxplots (top left) summarise the distribution of posterior log-likelihoods and the wireplot (right) shows mean posterior log-likelihoods (MPLL) calculated from 10^4 MCMC samples per resolution or combination.

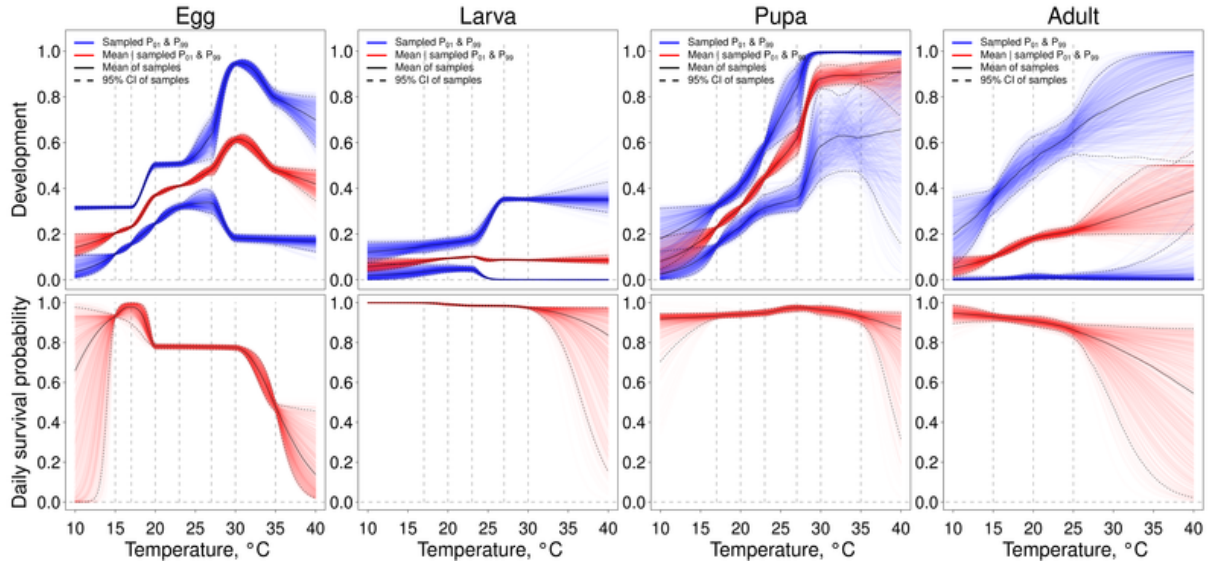


Figure 4. Non-linear responses to temperature of development (top) and survival (bottom) at most likely resolutions ($r_E^{MAP} = 6$, $r_L^{MAP} = 9$, $r_P^{MAP} = 31$, $r_A^{MAP} = 8$) for egg, larvae, pupae and adult midges. Results from 1000 MCMC samples are plotted with unimodal spline interpolation. Experimental temperatures are indicated (dashed vertical lines). Red and blue lines show median and 1% tail percentiles (P_{01} , P_{99}) of development kernel $f(\Delta|\theta)$ (top row). Expected values (black line) and CI_{95} s for each parameter are shown (dashed lines).

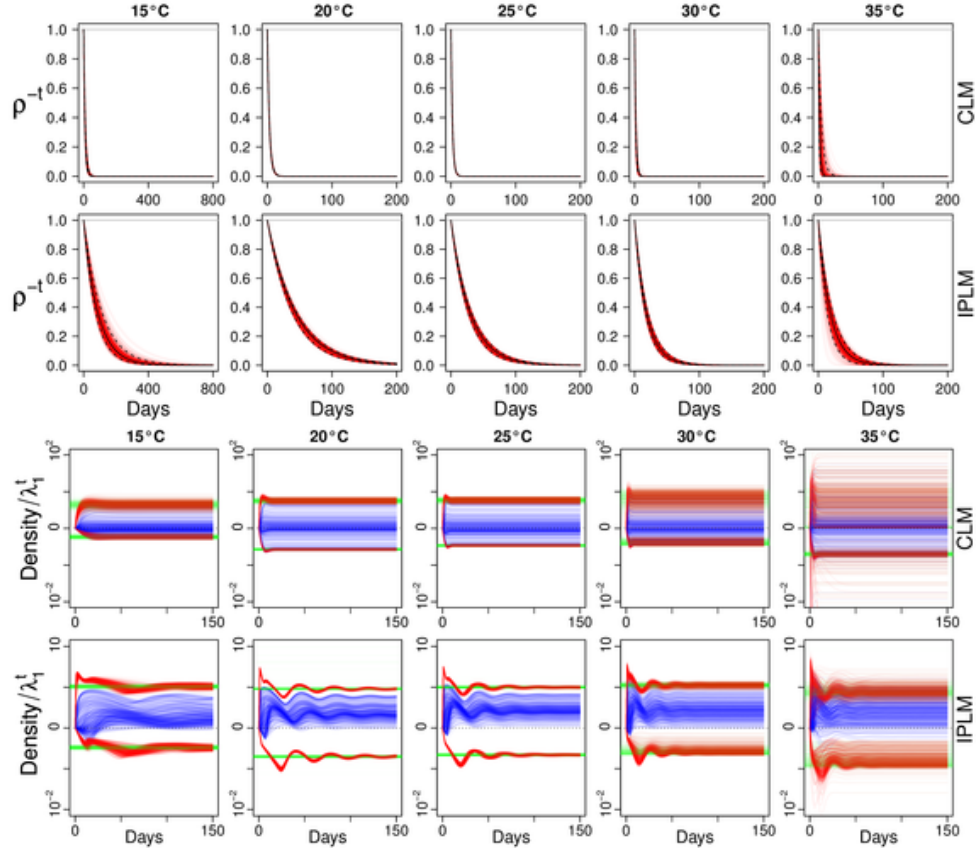


Figure 5. Indices of transient dynamics for *Culicoides* CLM and IPLM models at fixed temperatures. Geometric projections of inverted damping ratios ($\rho = \lambda_1/|\lambda_2|$) from 1000 MCMC samples (red lines), their means (black line) and CI_{95} s (dashed lines). Projected trajectories of relative densities (blue lines, third and forth rows), with initial values set to the 10°C stable stage distribution. Maximum amplification, maximum attenuation (upper and lower red lines respectively) and associated inertias (green lines) are shown.

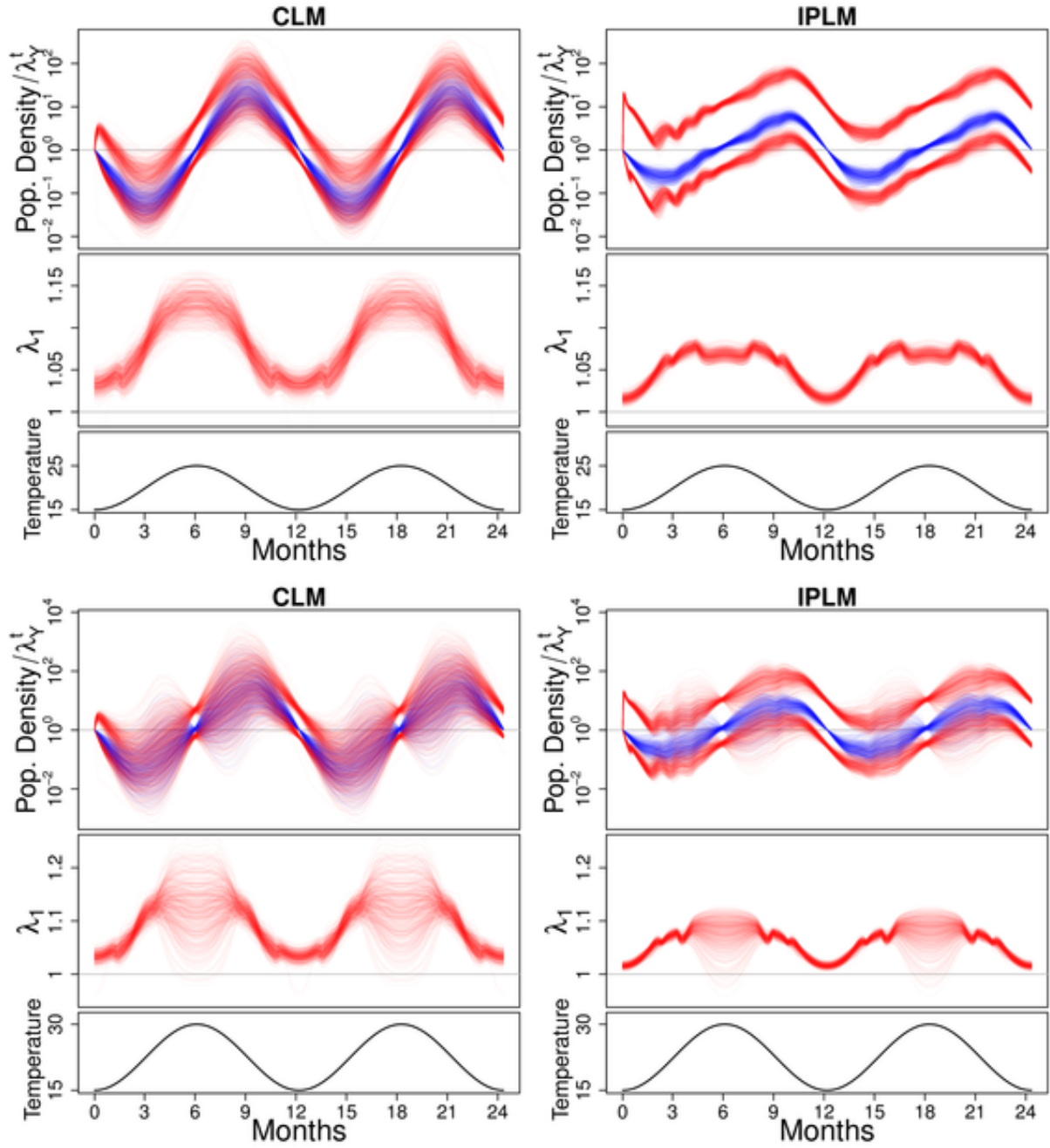


Figure 6. Growth rate (λ_1) and standardised density projections from *Culicoides* CLM (left) and IPLM (right) models forced with annual temperature fluctuations with ranges 15°C-25°C (top) and 15°C-30°C (bottom). Relative density projections were initialised at the 15°C stable stage distribution and standardised using annual growth λ_Y (blue lines). Maximum amplification and maximum attenuation trajectories (upper and lower red lines) are shown. Results from 1000 MCMC samples are presented.

SUPPORTING INFORMATION

Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes

María Soledad Castaño, Perry de Valpine, H       Guis and David R.J. Pleydell.

S1 Discretising integral projection models

The main text describes an integral projection model (IPM) for characterising the dynamics of within-stage development δ (eqn 3). In this model, an individual's developmental status δ progresses via increments $\Delta = \delta' - \delta$ with PDF $f(\Delta|\theta)$, CDF $F(\Delta|\theta)$ and parameters θ . For parsimony, we assume both $f(\Delta|\theta)$ and survival ν are independent of δ . The IPM-kernel is therefore $K_{\Theta}(\Delta) = \nu f(\Delta|\theta)$, where the PDF $f(\Delta|\theta)$ accounts for developmental rate heterogeneity. Throughout this work, we assume $f(\Delta|\theta)$ is the PDF of a beta distribution with mean μ and variance $\kappa\mu(1 - \mu)$. The full parameter set for a given stage is $\Theta = \{\nu, \theta\} = \{\nu, \mu, \kappa\}$.

A matrix approximation of this IPM for within-stage development is obtained by discretising the within-stage developmental status δ into r equally sized sub-stages. We assume individuals always start a new stage in the first sub-stage and develop by increments of up to r sub-stages each time-step. Within-stage transition probabilities for discrete developmental increments of l sub-stages are given by

$$p_l = \int_{\frac{l}{r+1}}^{\frac{l+1}{r+1}} f(\Delta|\theta) d\Delta, \quad (1)$$

where $p_r = 1 - \sum_{l=0}^{r-1} p_l$ gives the probability to complete the entire stage in just one time-step. These probabilities are used to define the projection matrix \mathbf{W}_S for each stage S (eqn 4 in main text). The full parameter set for a given stage in this discretised IPM-approximation is $\{\nu, \mu, \kappa, r\}$.

S2 Quality-dependent development with Gaussian copulas

Copulas are tools for generating multivariate distributions from an arbitrary set of marginal distributions (Kruskal, 1958; Nelsen, 2006). In the simulation-estimation study, Gaussian copulas were used to model correlation between individual quality q and development increments Δ . Here, we outline the details required to condition development

kernels on individual quality. We assume q is fixed through an individual's lifespan and conditions the distribution of increments at each time-step.

Marginal distributions of both Δ and q were described by beta distributions with densities

$$f_{\Delta}(\Delta|\alpha_1, \alpha_2) = \frac{\Delta^{\alpha_1-1}(1-\Delta)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)}, \quad (2)$$

$$f_q(q|\xi_1, \xi_2) = \frac{q^{\xi_1-1}(1-q)^{\xi_2-1}}{B(\xi_1, \xi_2)}, \quad (3)$$

and with corresponding cumulative distribution functions F_q and F_{Δ} . A standard uniform distribution is obtained for q by setting $\xi_1 = \xi_2 = 1$. Correlation between Δ and q was established via random variables x_{Δ} and x_q , which have a bi-variate Gaussian density with correlation coefficient ρ and standard normal marginal densities $\phi(\cdot)$ with distribution functions $\Phi(\cdot)$. These variables are linked via the following probability integral transformations:

$$F_q(q) = u_q = \Phi(x_q), \quad (4)$$

$$F_{\Delta}(\Delta) = u_{\Delta} = \Phi(x_{\Delta}), \quad (5)$$

where u_q and u_{Δ} follow standard uniform distributions. The joint density of Δ and q was therefore

$$f_{\Delta,q}(\Delta, q|\alpha_1, \alpha_2, \xi_1, \xi_2, \rho) = f_{x_{\Delta}, x_q}(x_{\Delta}, x_q|\alpha_1, \alpha_2, \xi_1, \xi_2, \rho) \left| \begin{array}{cc} \frac{\partial \Delta}{\partial x_{\Delta}} & \frac{\partial \Delta}{\partial x_q} \\ \frac{\partial q}{\partial x_{\Delta}} & \frac{\partial q}{\partial x_q} \end{array} \right|^{-1}, \quad (6)$$

where $f_{x_{\Delta}, x_q}(\cdot, \cdot|\alpha_1, \alpha_2, \xi_1, \xi_2, \rho)$ is the bi-variate normal density with correlation parameter ρ and the Jacobian determinant provides a change of variables correction for the transformations.

The above specification gives the following conditional density of Δ given q :

$$f_{\Delta|q}(\Delta|\alpha_1, \alpha_2, q, \rho) = \frac{f_{x_{\Delta}|x_q}(x_{\Delta}|\alpha_1, \alpha_2, x_q, \rho)f_{\Delta}(\Delta)}{\phi(x_{\Delta})}, \quad (7)$$

where $f_{x_{\Delta}|x_q}(x_{\Delta}|\alpha_1, \alpha_2, x_q, \rho)$ is normal with mean $\mu_{\Delta} = \rho x_q$ and variance $\sigma_{\Delta}^2 = 1 - \rho^2$. The corresponding conditional distribution function is given by the identity

$$F_{\Delta|q}(\Delta|\alpha_1, \alpha_2, q, \rho) = \int_0^{\Delta} f_{\Delta|q}(y|\alpha_1, \alpha_2, q, \rho)dy = \int_{-\infty}^{x_{\Delta}} f_{x_{\Delta}|x_q}(z|\alpha_1, \alpha_2, x_q, \rho)dz = F_{x_{\Delta}|x_q}(x_{\Delta}|\alpha_1, \alpha_2, x_q, \rho). \quad (8)$$

S3 Generating *in silico* data

For each of the five hundred simulations, *in silico* data for two successive stages was generated as follows. Sample size was fixed as $N = 50$ individuals and the duration of the maturation experiment was set to $t_c = 21$ time steps. For each simulation, parameters were sampled and data were simulated using a rejection sampler (algorithm S3.1) that ran until the following constraints were satisfied: 1) all parameters α_1 and α_2 for the development rate distributions F_Δ of each stage were greater than one, and 2) daily survival parameters $\{\nu_1, \nu_2\}$ were sufficiently large that the number of individuals completing both stages (n_{d_2} , see below) was at least $N_{d_2}^{\min}$. Throughout we used $N_{d_2}^{\min} = 35$. Qualities $\{q_i\}_{i=1}^{i=N}$, correlation coefficient ρ and initial parameters $\{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2\}$ were initialised with draws from a standard uniform distribution. These parameters were rejected and resampled if, for either stage, the alternative parameterisation $\alpha_1 = \mu \frac{(1-\kappa)}{\kappa}$ and $\alpha_2 = (1-\mu) \frac{(1-\kappa)}{\kappa}$ yielded $\alpha_1 \leq 1$ or $\alpha_2 \leq 1$.

Following each simulation (main text, section 3, subsection "Simulation"), the number of individuals that got censored (n_{c_1}, n_{c_2}), died (n_{m_1}, n_{m_2}) or developed (n_{d_1}, n_{d_2}) in each stage was recorded. If $n_{d_2} < N_{d_2}^{\min}$ and $n_{c_1} + n_{c_2} > n_{m_1} + n_{m_2}$, all parameters were resampled from their priors. Otherwise, if $n_{d_2} < N_{d_2}^{\min}$ and $n_{c_1} + n_{c_2} \leq n_{m_1} + n_{m_2}$, the lowest of the two survival probabilities was resampled from a uniform prior truncated at the current value of ν . The rejection sampler was stopped once $n_{d_2} \geq N_{d_2}^{\min}$ (see Algorithm S3.1).

S4 IPLM generated sojourn-mortality probabilities

In the IPLM approach, the fate of every individual in a given stage can be described by two pieces of information: 1) the time-to-event, $y_A \in \{1, \dots, t_c\}$ (where t_c is the time beyond which data are right censored); and 2) the event-type, $y_B \in \{\text{stage completion, mortality, censored}\}$. Probabilities associated with combinations of these two pieces of information can be obtained as follows. Given stage-specific parameters $\{\mu, \kappa, \nu, r\}$, construct the IPM-approximation (eqn 5, main text). Project the unit density pulse vector $(\mathbf{n}^T, c_s)_0 = (1, 0, \dots, 0)^T$ and, for each time-step $t \in \{0, \dots, t_c\}$, record the probabilities to complete a stage, $p_d(t)$, or die, $p_m(t)$. The probability of right censor beyond t_c is $p_c = 1 - \sum_{t=1}^{t_c} (p_d(t) + p_m(t))$.

Mean and variance of sojourn-time distribution The probabilities p_d , p_m and p_c define the right censored sojourn-mortality time distribution for individuals in a given stage. These probabilities provide the basis for calculating the likelihood of IPLM parameters given the observed data. The mean and variance of the sojourn

Algorithm S3.1: DATA SIMULATION WITH CONSTRAINTS (inputs= $N, N_{d_2}^{\min}$)

comment: Initialise all parameters.

$\{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, q_1, \dots, q_N, \rho\} \sim \text{Uniform}(0, 1)$

comment: Determine alternative parameters.

$\{\mu_1, \kappa_1\} \rightarrow \{\alpha_{11}, \alpha_{12}\}$

$\{\mu_2, \kappa_2\} \rightarrow \{\alpha_{21}, \alpha_{22}\}$

comment: Simulate cohort data $n_{c_1}, n_{c_2}, n_{m_1}, n_{m_2}, n_{d_2}$

while $n_{d_2} < N_{d_2}^{\min}$ **or** $\min\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}\} < 1$

comment: Use rejection until constraints are satisfied.

if $(n_{c_1} + n_{c_2} > n_{m_1} + n_{m_2})$ **or** $\min\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}\} < 1$

do $\left\{ \begin{array}{l} \{\mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, \rho, q_1, \dots, q_N\} \sim \text{Uniform}(0, 1) \\ \{\mu_1, \kappa_1\} \rightarrow \{\alpha_{11}, \alpha_{12}\} \\ \{\mu_2, \kappa_2\} \rightarrow \{\alpha_{21}, \alpha_{22}\} \\ \text{Simulate cohort data} \end{array} \right.$

do $\left\{ \begin{array}{l} \text{else if } (n_{c_1} + n_{c_2} \leq (n_{m_1} + n_{m_2})) \\ \quad \text{comment: Stepping-in avoids high rejection rates.} \\ \quad \text{if } \nu_1 = \min\{\nu_1, \nu_2\} \\ \quad \quad \text{do } \nu_1 \sim \text{Uniform}(\nu_1, 1) \\ \quad \quad \text{else} \\ \quad \quad \text{do } \nu_2 \sim \text{Uniform}(\nu_2, 1) \\ \quad \text{Simulate cohort data} \end{array} \right.$

return $(n_{c_1}, n_{c_2}, n_{m_1}, n_{m_2}, n_{d_2}, \mu_1, \mu_2, \kappa_1, \kappa_2, \nu_1, \nu_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, q_1, \dots, q_N, \rho)$

time distribution can be calculated as

$$\tilde{\mu} = \lim_{t_c \rightarrow \infty} \frac{\sum_{t=1}^{t_c} t p_d(t)}{\sum_{t=1}^{t_c} p_d(t)} \quad (9)$$

and

$$\tilde{\sigma}^2 = \lim_{t_c \rightarrow \infty} \frac{\sum_{t=1}^{t_c} (t - \tilde{\mu})^2 p_d(t)}{\sum_{t=1}^{t_c} p_d(t)}. \quad (10)$$

In practice, these quantities are approximated by setting t_c large enough that p_c become negligibly small.

S5 *Culicoides* life cycle data

Laboratory data from two *Culicoides* species were used to parameterise development, survival and fecundity for a complete life cycle. Note, these species share similar developmental responses across the 15°C-35°C range (Purse et al., 2015). Our own insectarium (ASTRE, Montpellier) provided individual-level data on gonotrophic cycle durations, number of gonotrophic cycles and number of eggs laid for *C. nubeculosus* females at 15°C, 20°C and 25°C (Balenghien et al., 2016). The laboratory also provided egg maturation and survival data at 15°C. Two similar *C. variipennis* studies (Mullens and Rutz, 1983; Vaughan and Turner, 1987) provided maturation time data for immature stages (egg, larva and pupa) – these data were available either as individual-level maturation times or summary statistics (sample means and standard deviations). The study of Vaughan and Turner (1987) provided developmental data at 20°, 23°, 27°, 30° and 35°C whilst that of Mullens and Rutz (1983) provided developmental data at 17°, 20°, 23°, 27°, 30° and 35°C. In the latter case, published data were complimented with original notes from the author’s lab-book which gave initial sample sizes and individual-level data for pupal and composite larval-pupal stage studies. Published larvae-only data from both studies were not used due to ambiguity regarding sample sizes. The likelihoods functions used to analyse data type are given in section S6.

S6 Likelihood functions

Likelihood in case study I. When data is available at the individual-level (i.e. for each observation the total number of counts in one), the probabilities p_d , p_m and p_c can be plugged into a categorical likelihood function, as described in case study I (main text).

Temp	DEVELOPMENT, SURVIVAL & FECUNDITY DATA								
	<i>C. nubeculosus</i> Balenghien <i>et al.</i> (2016)			<i>C.variipennis</i> Mullens <i>et al.</i> (1983)					Vaughan <i>et al.</i> (1987)
	gonotrophic cycle	egg	fecundity	egg	larva & pupa	pupa	egg	pupa	
	15	1k	1c	3					
	17				2u [†]	1c	2u		
20	1k		3	2u [†]	1c	2u	2p [†]	2p [†]	
23				2u [†]	1c	2u	2p [†]	2p [†]	
25	1k		3						
27				2u [†]	1c	2u	2p [†]	2p [†]	
30				2u [†]	1c	2u	2p [†]	2p [†]	
35							2p [†]	2p [†]	

Table S1. Available laboratory data for estimating *Culicoides* life cycle parameters. Data sets marked “1” provide sojourn time frequency distributions with “k” indicating that mortality date frequency distributions are *known* and “c” indicating *clumping* of mortality and right censored data. Data sets marked “2” provide the mean & standard deviation of observed sojourn times with “p” indicating the *proportion* surviving at each temperature is known and “u” indicating survival related information was *unreported*. The data set marked “3” provides the number of eggs laid per female in the gonotrophic cycle study. The notation † indicates missing sample size data requiring imputation steps described in section S7.

S6.1 Likelihood functions for the *Culicoides* study

For every column of table S1, data at different temperature are assumed to be independent. Thus, for a given stage and data set, eqn 9 in the main text is

$$f(\mu_1, \dots, \mu_K, \kappa_k, \dots, \kappa_K, \nu_1 \dots \nu_K, r | \mathbf{y}) \propto \prod_{k=1}^K f(y_k | \mu_k, \kappa_k, \nu_k, r) f(\mu_k, \kappa_k, \nu_k, r), \quad (11)$$

where $\mathbf{y} = \{y_1, \dots, y_K\}$ and y_k is data collected at the k^{th} temperature. The Bayesian framework easily permits to extend expression (11) for the cases where models were fitted using multiple data sources – they are assumed independent each other.

Details of the various expressions used for the likelihood function $f(y_k | \mu_k, \kappa_k, \nu_k, r)$, which depend on the statistical information provided in data $\mathbf{y} = \{y_1, \dots, y_K\}$ (see table S1), are presented below. The following expressions for each data type are valid at every empirical temperature and implicitly dependent of r , thus r and subscripts k are dropped for brevity.

Modelling sojourn-mortality time frequency data. Where data provide counts for the number of individuals that can be associated with $p_d(t)$, $p_m(t)$ or p_c at each time step, then a multinomial likelihood function can be defined. When such data is available at the individual-level – i.e. for each observation the total number of counts

in one – then a categorical likelihood function can be used. This was the case for adult females in the *Culicoides* study – note, data sets marked "1k" in table S1 provide the sojourn-mortality time frequency distribution for each gonotrophic cycle of individual females. In that study, the sojourn-time frequency distribution was used to model the time required to complete each gonotrophic cycle, where the completion of a gonotrophic cycle was indicated by the completion of egg-laying.

The fate of each individual i in a given stage (or gonotrophic cycle) is described by time-to-event $y_{iA} \in \{1, \dots, t_c\}$ and event-type $y_{iB} \in \{\text{development, mortality, censored}\}$ data. Each observation can therefore be modelled by assuming

$$(y_{iA}, y_{iB}) \sim \text{Categorical}(p_d(1), p_m(1), \dots, p_d(t_c), p_m(t_c), p_c). \quad (12)$$

Sojourn time frequency data with clumped mortality – right-censor information. Sometimes it is difficult or impossible to know precisely when mortality has occurred or even how many individuals have died prior to right censor time t_c . In such scenarios, the inability to distinguish dead from censored individuals requires clumping of the probabilities $p_m(t)$ and p_c to match the clumping of the associated data. Once data and probabilities are correctly clumped, then the likelihood is derived following a similarly procedure to the previous case.

In the *Culicoides* study, data sets marked "1c" in table S1 provide sojourn time frequency distribution data but do not distinguish the number of dead individuals from immature individuals still alive after t_c . Let y_t be the number of individuals completing maturation in t days, let y_{mc} be the number of dead or right censored individuals and y^{total} be the sample size. Likelihoods for such data are simply obtained by assuming

$$y_1, \dots, y_{t_c}, y_{mc} \sim \text{Multinomial}\left(y^{\text{total}}, p_d(1), \dots, p_d(t_c), 1 - \sum_{t=1}^{t_c} p_d(t)\right), \quad (13)$$

where probabilities $p_d(t)$ are obtained as described previously.

Likelihood for mean and standard deviation of sojourn times. Sometimes the only data available are summary statistics, typically means and standard deviations, obtained from publications. In such cases, we need the likelihood of the available summary statistics given the parameters.

For data sets marked "2u" and "2p" in table S1, data y consists of sojourn-time mean, μ^{obs} , and standard deviation, σ^{obs} . Ideally the sample size N is available too. Assuming normality, the following likelihood can be

written for this data:

$$f(\mu^{\text{obs}}, \sigma^{\text{obs}} | \mu, \kappa, \nu) = \frac{S^{\text{Post}}}{\sqrt{2\pi}\tilde{\sigma}} \exp \left\{ \frac{-\sigma^{\text{obs}2} + (\mu^{\text{obs}} - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right\}, \quad (14)$$

where $\{\mu, \kappa, \nu\}$ are the model parameters, and $\tilde{\mu}$ and $\tilde{\sigma}$ are moments of the joint sojourn-mortality time distribution estimated from probabilities $p_d(t)$ (Appendix S4), calculated with t_c high enough that p_c falls below a precision threshold and could be assumed negligible. Note, in equation (S14) above, μ^{obs} and σ^{obs} are not independent of mortality ν . However, greater precision in mortality ν is possible when survival data are available too. Sometimes, detailed survival data are not available, but the number, or proportion, of individuals surviving until t_c is reported. Data sets marked " $2p^\dagger$ " (in table S1) reported the proportion of individuals surviving each experimental temperature, π , plus the total sample size at the start of each experiment, $S_{\text{total}}^{\text{Pre}}$. The proportions π were used in the imputation of S^{Post} (S7). The posterior likelihood of daily survival ν , given $\{\mu, \kappa\}$, was given by the beta-binomial model

$$f(\nu | \mu, \kappa, S^{\text{Pre}}, S^{\text{Post}}) \propto \hat{\pi}^{S^{\text{Post}}} (1 - \hat{\pi})^{S^{\text{Pre}} - S^{\text{Post}}}, \quad (15)$$

where $\hat{\pi} = 1 - \sum_{t=1}^{t_c} p_m(t)$.

S6.2 Bayesian model for expected fecundity per gonotrophic cycle

Fecundity data obtained from the gonotrophic cycle experiment (marked "3" in table S1) provided the number of eggs laid at the end of each gonotrophic cycle. The posterior likelihood for the expected fecundity given this data was obtained using the following approach. Expected fecundity at each temperature was estimated from oviposition data using a Poisson model with Jeffrey's prior, resulting in posterior distributions

$$F_A(T_k) \sim \text{Gamma} \left(\text{shape} = \frac{1}{2} + \sum_{i=1}^{N_{T_k}} n_{i_{T_k}}, \text{rate} = N_{T_k} \right) \quad (16)$$

where, at temperature T_k , N_{T_k} is the number of observed ovipositions and $n_{i_{T_k}}$ is the fecundity of the i^{th} oviposition. As for survival and development, a unimodal constraint with respect to temperature was used to facilitate unimodal spline interpolation (see section S10).

S7 Imputation of missing data

Three of the data sets used for modeling the *Culicoides* life cycle presented missing sample size data (see table S1). Bayesian imputation steps to account for associated uncertainties are described here.

Sample sizes at experimental temperatures T_k were typically reported in one of the two forms : the sample size at the start of maturation experiment k , S_k^{Pre} , and the sample size at the end of a maturation experiment k , S_k^{Post} . It was assumed that any differences between S_k^{Pre} and S_k^{Post} could be accounted for by mortality and right censoring.

In Vaughan *et al.*'s egg and pupae studies, the total sample size $S_{\text{total}}^{\text{Pre}}$ was published, but how those numbers were divided among the five experimental temperatures was missing data. Moreover, neither S_k^{Post} or $S_{\text{total}}^{\text{Post}}$ were published, although the proportion that survived each experiment, π_k , was available. The duration of the experiment was not published, curtailing the possibility to account for potential right censor. It was assumed *a priori* that the expected sample sizes in each of the n_T experimental temperatures were equivalent. Dropping the k notation for brevity, this gave the following prior for each experimental group,

$$S_1^{\text{Pre}}, \dots, S_{n_T}^{\text{Pre}} \sim \text{Multinomial}\left(S_{\text{total}}^{\text{Pre}}, \frac{1}{n_T}, \dots, \frac{1}{n_T}\right). \quad (17)$$

The proportions π_k that survived at each experimental temperature T_k permitted imputed values for S_k^{Post} to be determined as

$$S_k^{\text{Post}} = \pi_k S_k^{\text{Pre}}. \quad (18)$$

In the Mullens and Rutz egg data, a total sample size at the end of the maturation experiment was reported, but how those sample sizes were distributed among the different temperature groups was not reported. However, those authors did note in their paper that those values were roughly equivalent. Thus, we adopted the following prior

$$S_1^{\text{Post}}, \dots, S_{n_T}^{\text{Post}} \sim \text{Multinomial}\left(S_{\text{total}}^{\text{Post}}, \frac{1}{n_T}, \dots, \frac{1}{n_T}\right). \quad (19)$$

S8 Posteriors for the *Culicoides* study

For stages $S = \{\text{egg}, \text{adult}\}$, posterior densities of parameters at each experimental temperature were of the form

$$f(\mu_S, \kappa_S, \nu_S, r_S, \hat{\mathbf{y}}_S | \mathbf{y}_S) \propto f(\mathbf{y}_S | \mu_S, \kappa_S, \nu_S, r_S, \hat{\mathbf{y}}_S) f_{\mathcal{U}}(\mu_S, \kappa_S) f_{\mathcal{U}}(\nu_S) f(r_S) f(\hat{\mathbf{y}}_S), \quad (20)$$

where \mathbf{y}_S represents the full set of data, from different various sources, for stage S at a given empirical temperature, \mathcal{U} indicates the unimodality constraints, and $\hat{\mathbf{y}}_S$ represents imputed missing data. Recall, unimodal constraint \mathcal{U} includes: 1) unimodal response of parameters ν , \mathcal{P}_{01} and \mathcal{P}_{99} to temperature, and 2) unimodality in $f(\Delta|\alpha_1, \alpha_2)$, which is ensured when both α_1 and α_2 are greater than one. Posterior densities for larvae and pupae parameters per experimental temperature were

$$\begin{aligned} f(\mu_L, \mu_P, \kappa_L, \kappa_P, \nu_L, \nu_P, r_L, r_P, \hat{\mathbf{y}}_P | \mathbf{y}_P, \mathbf{y}_{LP}) &\propto f_{\mathcal{U}}(\mu_L, \kappa_L) f_{\mathcal{U}}(\nu_L) f(r_L) f_{\mathcal{U}}(\mu_P, \kappa_P) f_{\mathcal{U}}(\nu_P) f(r_P) f(\hat{\mathbf{y}}_P) \\ &\times f(\mathbf{y}_P | \mu_P, \kappa_P, \nu_P, r_P, \hat{\mathbf{y}}_P) \\ &\times f(\mathbf{y}_{LP} | \mu_L, \kappa_L, \nu_L, \mu_P, \kappa_P, \nu_P, r_L, r_P). \end{aligned} \quad (21)$$

Likelihoods for data \mathbf{y}_E , \mathbf{y}_P and \mathbf{y}_A were calculated from sojourn-mortality distributions obtained by projecting the unit pulse vector $(\mathbf{n}^T, c_S)_0 = (1, 0, \dots, 0)^T$ with eqn 5, main text). The likelihood for \mathbf{y}_{LP} was calculated similarly using the following two-stage projection matrix for the IPM-approximation:

$$\mathcal{M}_{LP} = \begin{bmatrix} \mathbf{W}_L & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_L & \mathbf{W}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_P^1 & 1 \end{bmatrix}. \quad (22)$$

S9 MCMC strategy

Parameters μ , κ and ν are bounded on $(0, 1)$. We adopted the strategy of transforming these parameters to the logit scale to enable sampling on unbounded domains. Thus, we sampled the logit transformed parameters

$$\mu' = \log\left(\frac{\mu}{1-\mu}\right), \quad (23)$$

$$\kappa' = \log\left(\frac{\kappa}{1-\kappa}\right), \quad (24)$$

$$\nu' = \log\left(\frac{\nu}{1-\nu}\right). \quad (25)$$

A simple change of variables correction reveals the prior densities of these transformed parameters to be

$$f(\mu') = \frac{1}{(1 + e^{\mu'})(1 + e^{-\mu'})}, \quad (26)$$

$$f(\kappa') = \frac{1}{(1 + e^{\kappa'})(1 + e^{-\kappa'})}, \quad (27)$$

$$f(\nu') = \frac{1}{(1 + e^{\nu'})(1 + e^{-\nu'})}. \quad (28)$$

In the *Culicoides* study, target posteriors distributions were sampled using a parallel tempering algorithm (Swendsen and Wang, 1986; Liu, 2008; Łacki and Miasojedow, 2015) adapted from NIMBLE's library of functions for adaptive MCMC. Tempering was not applied to the transformed priors in equations (26-28) since the prior on the original scale is already flat, neither was tempering applied during the imputation steps. Depending on the analysis, the tempering was performed with temperature ladders of 10 to 15 different temperatures (n_{Temps}). Temperature ladders were initialised as $\mathbf{T} = e^{\log(0 \times 10)}, e^{\log(1 \times 10)}, \dots, e^{\log(n_{\text{Temps}} \times 10)}$ and were adjusted to target a 0.234 acceptance rate using techniques described in (Łacki and Miasojedow, 2015) during an adaptive burn-in phase.

A burn-in period, consisting of a series of runs of 10^4 iterations (10^3 in the simulation-estimation study), was iterated until $\overline{LP}_{\text{run}} < 2 + \overline{LP}_{\text{run}-1}$, where $\overline{LP}_{\text{run}}$ is the mean log posterior density of the model over a given short run. Expected sample size (ESS) (Plummer et al., 2006) was calculated for each parameter from the final pre-run. Thinning was set to $2 \times \min(\text{ESS})$, to remove much of the auto-correlation from subsequent samples. Thereafter, 10^4 thinned post-adaption MCMC samples were generated per model and convergence diagnostics were performed using CODA.

To avoid mixing difficulties in the *Culicoides* study, the above sampling strategy was applied with resolution parameters fixed at values (or combinations of values for larvae-pupae) given by $r_E, r_P, r_A \in \{1, \dots, 50\}$ and $r_L \in \{1, \dots, 15\}$. Integration over each $r_S \in \{r_E, r_A\}$ was achieved in a post-MCMC step by sampling among lines of model output with weights

$$w_l(r_S) = \frac{f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}} | \mathbf{y}, r_S)}{\sum_{r'_S=1}^{50} f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{s}, \hat{\mathbf{y}} | \mathbf{y}, r'_S)}, \quad (29)$$

where $l \in \{1, \dots, 10^4\}$ indicates the MCMC output line. For the larvae-pupae analysis, equation (29) was adjusted

to include integration over both r_L and r_P such that

$$w_l(r_L, r_P) = \frac{f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{s}, \hat{\mathbf{y}} | \mathbf{y}, r_L, r_P)}{\sum_{r'_L=1}^{15} \sum_{r'_P=1}^{50} f_l(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{s}, \hat{\mathbf{y}} | \mathbf{y}, r'_L, r'_P)}. \quad (30)$$

These post-MCMC sampling steps generated a unique MCMC output file per stage and further CODA diagnostics were performed on those outputs. NIMBLE and R scripts used in these analyses are available on github https://github.com/scastano/IPLM_code.

S10 Unimodal cubic Hermite spline interpolation

Posterior estimates of $\text{logit}(\mathcal{P}_{01S})$, $\text{logit}(\mathcal{P}_{99S})$, $\text{logit}(\nu_S)$ and $\log(F_A)$ at unsampled temperatures were obtained via interpolation with unimodal (i.e. up to two piece-wise monotonic) cubic Hermite splines. Recall, given n points (x_k, y_k) , where $k \in \{1, \dots, n\}$ and $x_k < x_{k+1}$ for all k , a cubic Hermite spline between two successive points is defined

$$f_{\text{interpolated}}(t) = y_k h_{00}(t) + (x_{k+1} - x_k) m_k h_{10}(t) + y_{k+1} h_{01}(t) + (x_{k+1} - x_k) m_{k+1} h_{11}(t), \quad (31)$$

where m_k is the gradient at point k , $t = \frac{x - x_k}{x_{k+1} - x_k}$ and h_{ii} are the cubic Hermite spline basis functions $h_{00}(t) = (1 + 2t)(1 - t)^2$, $h_{10}(t) = t(1 - t)^2$, $h_{01}(t) = t^2(3 - 2t)$ and $h_{11}(t) = t^2(t - 1)$. Equation (31) is available in R as `splinefunH` in the `stats` package.

The Fritsch-Carlson method (Fritsch and Carlson, 1980) provides a deterministic algorithm for setting gradients m_k such that the fitted spline is piece-wise monotonic. Pseudo-code for a classic implementation of the Fritsch-Carlson method is given in Algorithm S10.1.

We use a stochastic variation of the Fritsch-Carlson algorithm that permits uncertainty in gradients m_k to be explored within the piece-wise monotonic constraint $\alpha_k^2 + \beta_k^2 < 9$. We assume throughout that the set of points $(x_1, y_1), \dots, (x_n, y_n)$ contains at most one local maximum – a condition used as a constraint during MCMC. Moreover, we initialise each m_k with a draw from the following unconstrained conditional (on \mathbf{y}) priors,

$$\arctan(m_k) \sim \begin{cases} \text{Uniform}(\arctan(\Delta_k), \arctan(\Delta_{k-1})) & \text{if } y_{k-1} < y_k \text{ and } y_k > y_{k+1}, \text{ else} \\ \text{Uniform}(0, \arctan(3\Delta_{k-1})) & \text{if } y_{k-1} < y_k, \text{ else} \\ \text{Uniform}(\arctan(3\Delta_{k-1}), 0) & \text{if } k \text{ s.t. } y_{k-1} > y_k \end{cases} \quad (32)$$

Algorithm S10.1: FRITSCH-CARLSON METHOD(\mathbf{x}, \mathbf{y})

comment: Set slopes of secant lines
 $\Delta_k = \frac{x - x_k}{x_{k+1} - x_k} \quad \forall k \in \{1, n-1\}$
comment: Initialise tangents
 $m_1 = \Delta_1$ and $m_n = \Delta_{n-1}$
for $k \in \{2, n-1\}$
 $\left\{ \begin{array}{l} m_k = \frac{\Delta_{k-1} + \Delta_k}{2} \\ \text{if } \text{sign}(\Delta_{k-1}) \neq \text{sign}(\Delta_k) \\ \quad \text{then } m_k = 0 \\ \text{if } \Delta_k = 0 \\ \quad \text{then } m_k = m_{k+1} = 0 \end{array} \right.$
comment: Derive parameters α and β
 $\alpha_k = m_k / \Delta_k \quad \forall k \in \{1, n-1\}$
 $\beta_k = m_{k+1} / \Delta_k \quad \forall k \in \{1, n-1\}$
comment: Reset gradients where monotonic constraint is violated
for any k s.t. $\alpha_k^2 + \beta_k^2 > 9$
 $\left\{ \begin{array}{l} m_k = \frac{3\alpha_k \Delta_k}{\sqrt{\alpha_k^2 + \beta_k^2}} \\ m_{k+1} = \frac{3\beta_k \Delta_k}{\sqrt{\alpha_k^2 + \beta_k^2}} \end{array} \right.$

where $\Delta_0 = \Delta_1$ and impose the constraint that $\alpha_k^2 + \beta_k^2 < 9$, which truncates these priors, to avoid “overshoot”. We use a stepping-in algorithm to re-sample $\arctan(m_k)$ wherever the unimodality constraint is violated. Pseudo-code for this sampling strategy is given in Algorithm S10.2.

Finally, for each stage S (suffix dropped for brevity), development kernel parameters α_1 and α_2 at any temperatures T absent in the set of experimental temperatures for S were derived from interpolated values $\mathcal{P}_{01}(T)$ and $\mathcal{P}_{99}(T)$ by numerical minimisation. Thus, α_{1T} and α_{2T} were calculated as

$$\alpha_{1T}, \alpha_{2T} = \arg \min_{\alpha_1, \alpha_2} \sum_{p \in \{1, 99\}} \left(\hat{\mathcal{P}}_p(\alpha_1, \alpha_2) - \mathcal{P}_p(T) \right)^2, \quad (33)$$

where $\hat{\mathcal{P}}_p(\alpha_1, \alpha_2)$ was calculated with R function `qbeta` and minimisation was performed using the Nelder-Mead algorithm of R function `optim`.

Algorithm S10.2: STOCHASTIC UNIMODAL CUBIC SPLINE(x, y)

comment: Set slopes of secants and identify mode

$$\Delta_k = \frac{x - x_k}{x_{k+1} - x_k} \quad \forall k \in \{1, n-1\} \text{ and } k_{max} = \arg \max_k (y_k)$$

comment: Initialise bounds and sample gradients

for $k \leftarrow 1$ **to** n

do $\left\{ \begin{array}{l} \text{if } k = k_{max} \text{ and } k \notin \{1, n\} \text{ then } L_k = \Delta_k \text{ and } U_k = \Delta_{k-1} \\ \text{else if } y_k < y_{k+1} \text{ then } L_k = 0 \text{ and } U_k = 3\Delta_{k-1} \\ \text{else if } y_k > y_{k+1} \text{ then } L_k = 3\Delta_{k-1} \text{ and } U_k = 0 \\ \arctan(m_k) \sim \text{Uniform}(\arctan(L_k), \arctan(U_k)) \end{array} \right.$

$$\alpha_k = m_k / \Delta_k \text{ and } \beta_k = m_{k+1} / \Delta_k \quad \forall k \in 1, \dots, n-1$$

while $\max\{\alpha_k^2 + \beta_k^2 : k \in 1, \dots, n-1\} > 9$

do $\left\{ \begin{array}{l} \kappa = \arg \max_k (\alpha_k^2 + \beta_k^2) \\ \text{if } \beta_\kappa > \alpha_\kappa \text{ then } \kappa = \kappa + 1 \\ \text{if } \kappa = k_{max} \text{ and } m_\kappa < 0 \text{ then } L_\kappa = m_\kappa \\ \text{else if } \kappa = k_{max} \text{ and } m_\kappa > 0 \text{ then } U_\kappa = m_\kappa \\ \text{else if } \kappa < k_{max} \text{ then } U_\kappa = m_\kappa \\ \text{else if } \kappa > k_{max} \text{ then } L_\kappa = m_\kappa \\ \arctan(m_\kappa) \sim \text{Uniform}(\arctan(L_\kappa), \arctan(U_\kappa)) \\ \alpha_k = m_k / \Delta_k \text{ and } \beta_k = m_{k+1} / \Delta_k \quad \forall k \in 1, \dots, n-1 \end{array} \right.$

S11 Supplementary Figures

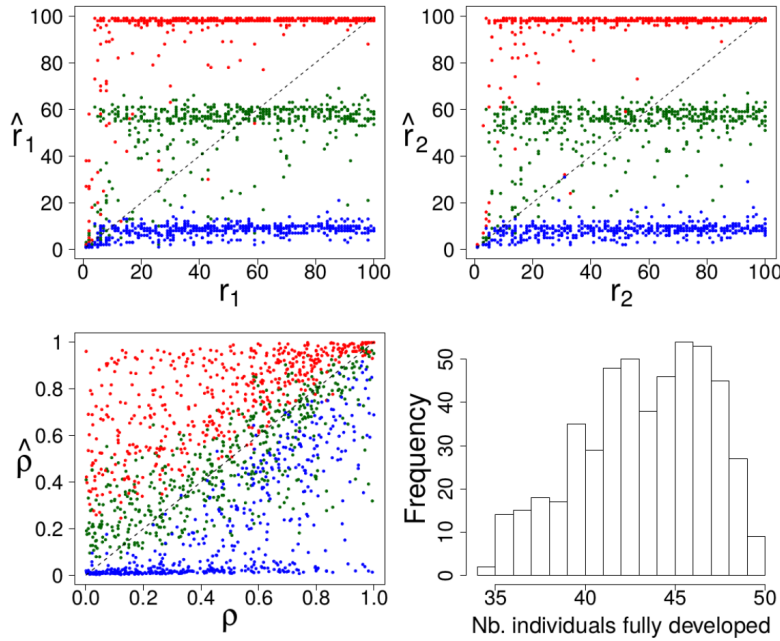


Figure S1. Estimated (y-axes) vs. true (x-axes) values for resolutions r_1 , r_2 and correlation ρ parameters, and frequency distribution of individuals completing both stages (bottom right), from 500 simulations with a two-stage quality-conditioned IPLM model with a sample size of $N = 50$. For each simulation, the median (green) and the upper (red) and lower (blue) bounds of the 95% credibility interval obtained from MCMC sampling of the posterior distribution are shown. The one-one line (dashed) is shown for reference.

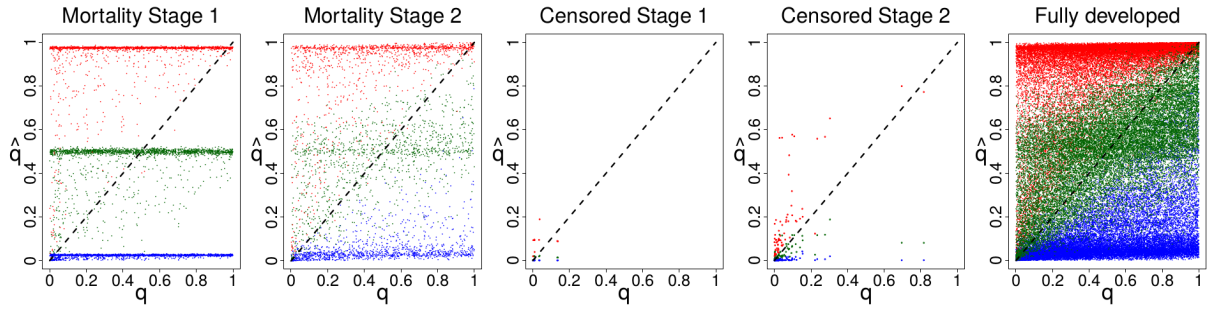


Figure S2. Estimated individual qualities \hat{q} versus their corresponding true value q , from 500 simulations of a two-stage quality-conditioned IPLM model with a sample size of $N = 50$. For each simulation, the median (green) and the upper (red) and lower (blue) bounds of the 95% credibility interval obtained from MCMC sampling of the posterior distribution are shown for five classes of individuals: died in stage 1; died in stage 2; censored in stage 1; censored in stage 2; and completed both stages. The one-one line (dashed) is shown for reference.

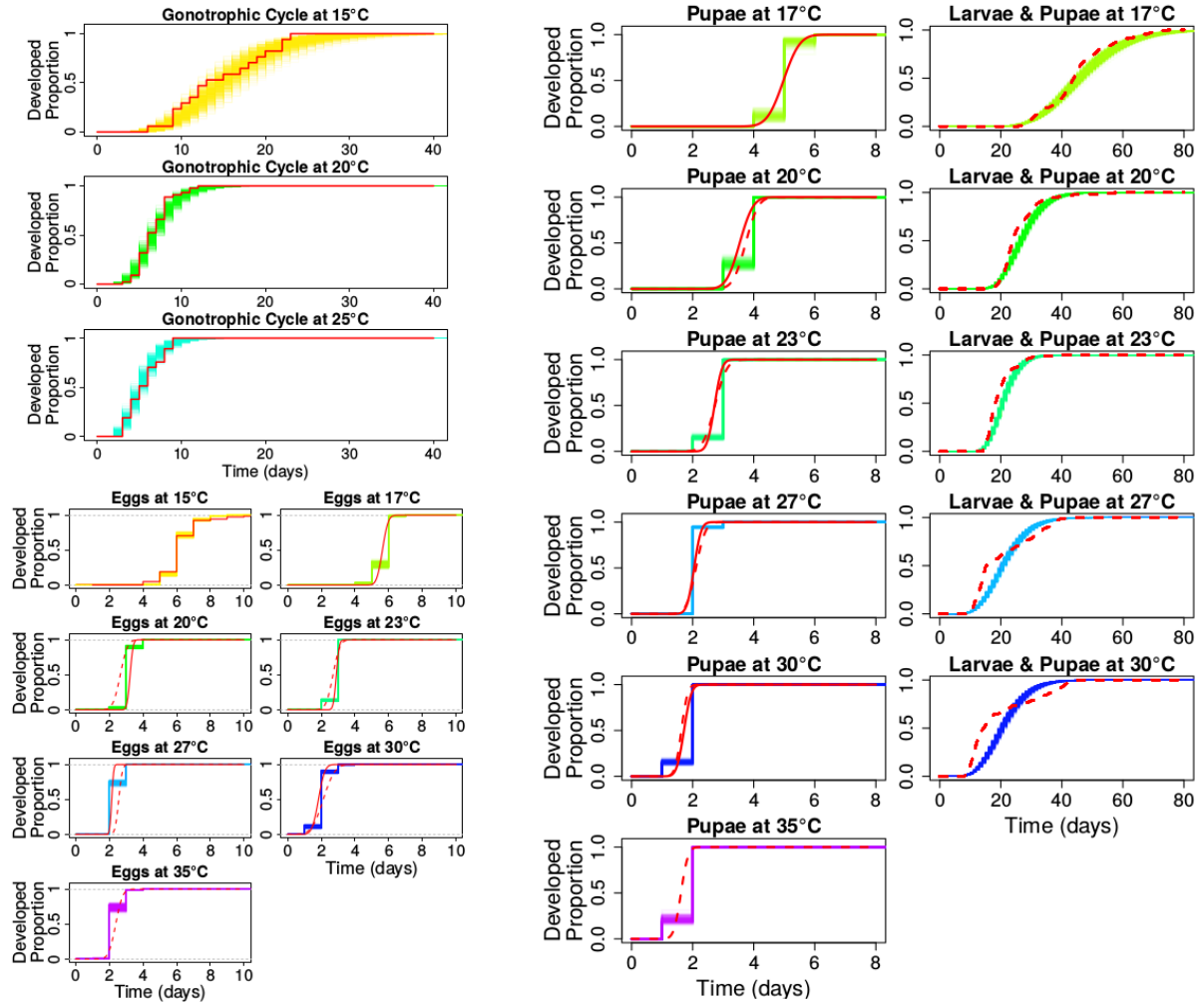


Figure S3. Posterior cumulative distributions of within-stage sojourn times of a temperature-dependent IPLM plotted with empirical *Culicoides* data (red lines). Where two data sets are shown, continuous red lines indicate Mullens *et al.* data, and dashed red lines indicate Vaughan *et al.* data. The fitted sojourn time distributions contrast markedly to those obtained with CLM (see Fig. S4)

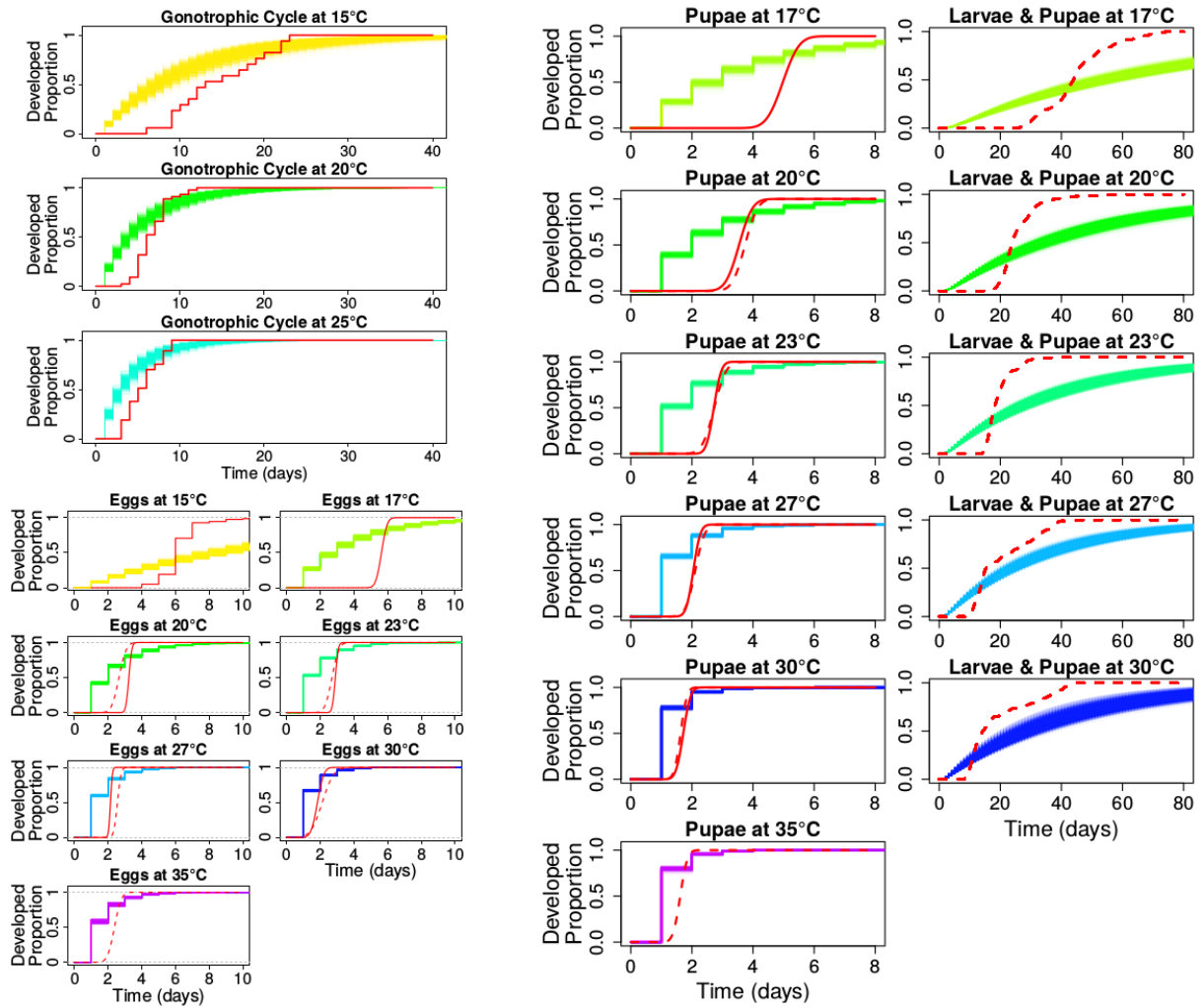


Figure S4. Posterior cumulative distributions of within-stage sojourn times of a temperature-dependent CLM plotted with empirical *Culicoides* data (red lines). Where two data sets are shown, continuous red lines indicate Mullens *et al.* data, and dashed red lines indicate Vaughan *et al.* data. The fitted sojourn time distributions contrast markedly to those obtained with IPLM (see Fig. S3).

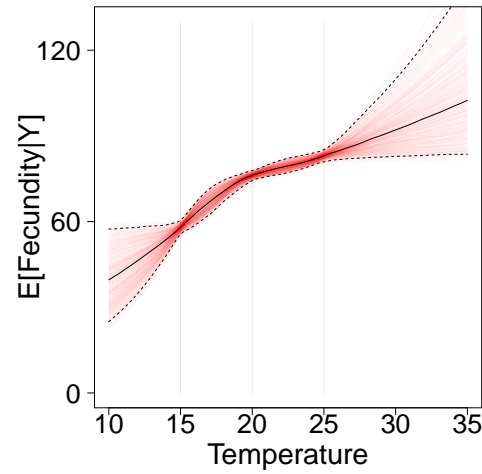


Figure S5. Expected fecundity from a Poisson-Jeffreys model fitted to *Culicoides* oviposition data collected at three temperatures (vertical grey lines). Posterior estimates from 1000 MCMC samples are shown with extrapolation over the range $10^{\circ} - 35^{\circ}\text{C}$. The data suggest a non-linear response of fecundity to temperature.

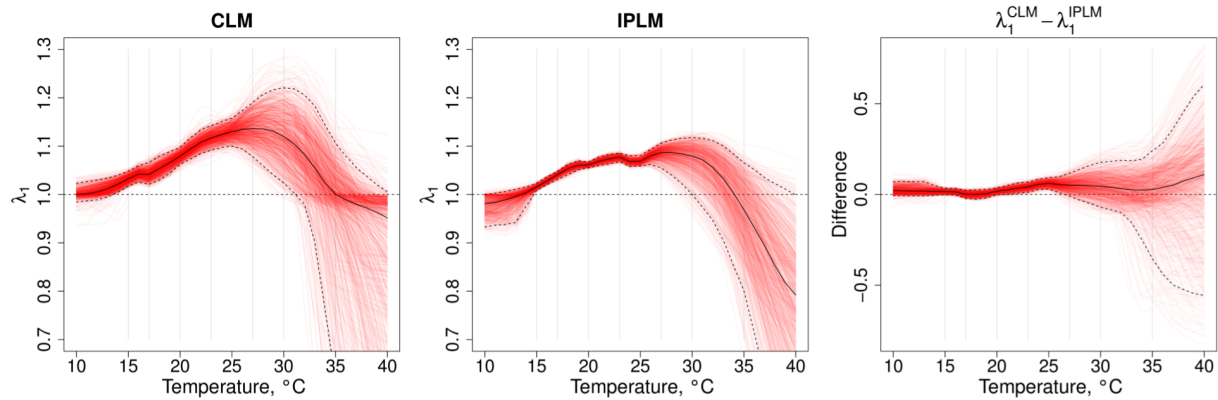


Figure S6. Temperature responses of asymptotic growth rate (dominant eigenvalue, λ_1) of CLM (left) and IPLM (center). Expected growth rates were higher for CLM than for IPLM over much of the temperature range, and the 95% CIs for this difference excluded zero over a range of approximately $22^{\circ}\text{C} - 26^{\circ}\text{C}$ (right). Both models predict population decline ($\lambda_1 < 1$) at higher temperatures. However, CLM predicted $\lambda_1 > 1$ over a greater range of temperatures than IPLM. Expected values (black line) with 95% CIs (dashed lines) from 1000 MCMC samples (red lines) are shown.

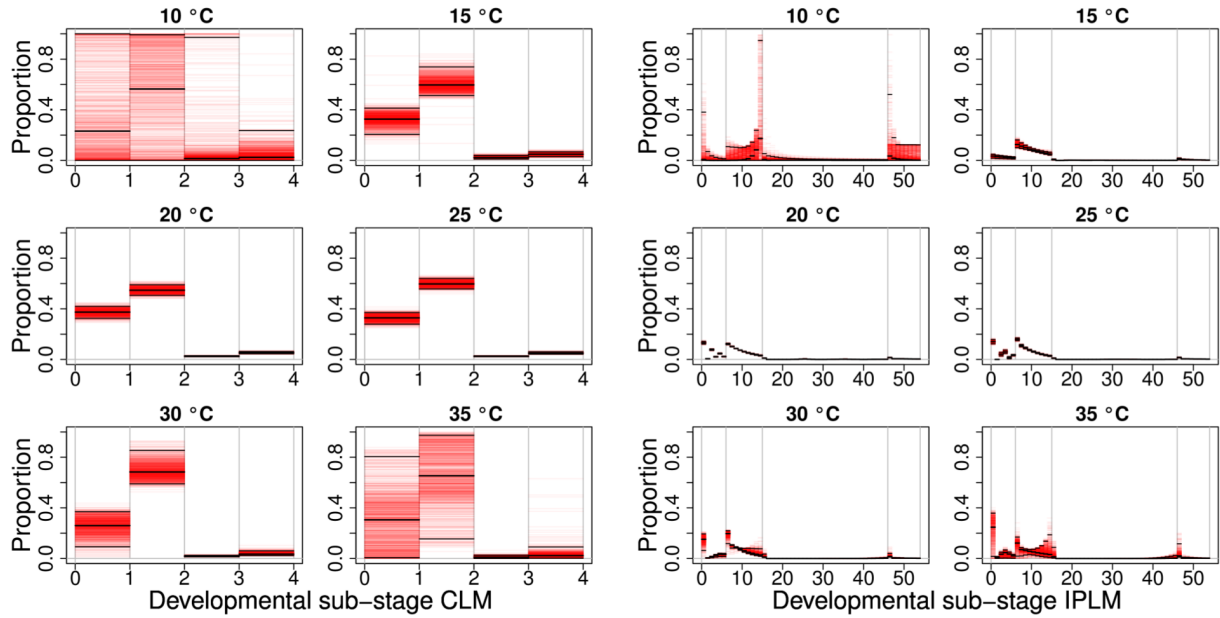


Figure S7. Stable stage distributions at several temperatures for *Culicoides* CLM (left) and IPLM (right) models. Results from 1000 MCMC samples (red lines) with posterior means and 95% CIs (black lines) are shown. For the IPLM model, resolutions were set to their maximum *a posteriori* (MAP) estimates: $r_E^{MAP} = 6$, $r_L^{MAP} = 9$, $r_P^{MAP} = 31$ and $r_A^{MAP} = 8$. Vertical lines separate stages.

References

- Balenghien, T., Allene, X., Lhoir, J., Rakotoarivony, I., and Huber, K. (2016). Data from: Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes. Dryad Digital Repository. doi:123456789/dryad.123456789 [UPLOAD PENDING].
- Fritsch, F. N. and Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17:238–246.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53:814–861.
- Łacki, M. K. and Miasojedow, B. (2015). State-dependent swap strategies and automatic reduction of number of temperatures in adaptive parallel tempering algorithm. *Statistics and Computing*, pages 1–14.
- Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer.

- Mullens, B. and Rutz, D. (1983). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae) at constant laboratory temperatures. *Annals of the Entomological Society of America*, 76:747–751.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. 2nd Edition. Springer-Verlag, New York.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11.
- Purse, B., Carpenter, S., Venter, G., Bellis, G., and Mullens, B. (2015). Bionomics of temperate and tropical *Culicoides* midges: knowledge gaps and consequences for transmission of *Culicoides*-borne viruses. *Annual Review of Entomology*, 60:373–392.
- Swendsen, R. H. and Wang, J.-S. (1986). Replica Monte Carlo simulation of spin-glasses. *Physical Review Letters*, 57:2607.
- Vaughan, J. and Turner, E. (1987). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae) from Saltville, Virginia, at constant laboratory temperatures. *Journal of Medical Entomology*, 24:390–395.

Bibliography

- Aguinaldo, A. M. A., Turbeville, J. M., Linford, L. S., Rivera, M. C., *et al.* (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 387(6632), 489.
- Aiello, W. G. and Freedman, H. (1990). A time-delay model of single-species growth with stage structure. *Mathematical biosciences*, 101(2), 139–153.
- Akey, D. H., Potter, H. W., and Jone, R. H. (1978). Effects of rearing temperature and larval density on longevity, size, and fecundity in the biting *Culicoides variipennis*. *Annals of the Entomological Society of America*, 71(3), 411–418.
- Alekseev, A. N. (2004). Pathogen transmission by arthropods. In *Encyclopedia of Entomology*, pages 1643–1645. Springer.
- Anderson, C. R., Spence, L., Downs, W. G., Aitken, T. H., *et al.* (1961). Oropouche virus: a new human disease agent from Trinidad, West Indies. *American Journal of Tropical Medicine and Hygiene*, 10(4), 574–8.
- Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10), 535–544.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.
- Ansari, M. A., Pope, E. C., Carpenter, S., Scholte, E.-J., and Butt, T. M. (2011). Entomopathogenic fungus as a biological control for an important vector of livestock disease: the *Culicoides* biting midge. *PloS one*, 6(1), e16108.
- Ardia, D. R. (2005). Individual quality mediates trade-offs between reproductive effort and immune function in tree swallows. *Journal of Economic Entomology*, 74, 517–524.
- Aubry, A., Bécart, E., Davenport, J., and Emmerson, M. C. (2010). Estimation of survival rate and extinction probability for stage-structured populations with overlapping life stages. *Population ecology*, 52(3), 437–450.
- Baisley, K. J., Watts, D. M., Munstermann, L. E., and Wilson, M. L. (1998). Epidemiology of endemic Oropouche virus transmission in upper Amazonian Peru. *The American journal of tropical medicine and hygiene*, 59(5), 710–716.
- Balenghien, T., Allene, X., Lhoir, J., Rakotoarivony, I., and Huber, K. (2016). Data from: Lefkovitch matrices meet integral projection models: quantifying the effects of individual heterogeneity in hidden developmental processes. Dryad Digital Repository. doi:123456789/dryad.123456789 [UPLOAD PENDING].
- Battisti, D. S. and Naylor, R. L. (2009). Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, 323, 240–244.
- Bayes, T., Price, R., and Canton, J. (1763). *An essay towards solving a problem in the doctrine of chances*. C. Davis, Printer to the Royal Society of London.
- Beaty, B. J., Marquardt, W. C., *et al.* (1996). *The biology of disease vectors*. University Press of Colorado.
- Beer, M., Conraths, F., and Van der Poel, W. (2013). 'Schmallenberg virus' – a novel orthobunyavirus emerging in Europe. *Epidemiology and infection*, 141(01), 1–8.

- Belehradek, J. (1935). *Temperature and Living Matter*. Borntraeger, Berlin.
- Bellows Jr, T. and Birley, M. (1981). Estimating developmental and mortality rates and stage recruitment from insect stage-frequency data. *Researches on Population Ecology*, 23, 232–244.
- Bentley, M. D. and Day, J. F. (1989). Chemical ecology and behavioral aspects of mosquito oviposition. *Annual review of entomology*, 34(1), 401–421.
- Bernardo, J. M. and Smith, A. (2000). *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.
- Biek, R., Funk, W. C., Maxell, B. A., and Mills, L. S. (2002). What is missing in amphibian decline research: insights from ecological sensitivity analysis. *Conservation Biology*, 16(3), 728–734.
- Bierzychudek, P. (1999). Looking backwards: Assessing the projections of a transition matrix model. *Ecological Applications*, 9, 1278–1287.
- Birt, A., Feldman, R. M., Cairns, D. M., Coulson, R. N., Tchakerian, M., Xi, W., *et al.* (2009). Stage-structured matrix models for organisms with non-geometric development times. *Ecology*, 90, 57–68.
- Blythe, S., Nisbet, R., and Gurney, W. (1984). The dynamics of population models with distributed maturation periods. *Theoretical Population Biology*, 25, 289–311.
- Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., *et al.* (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, 26, 183–192.
- Bonhomme, R. (2000). Bases and limits to using degree-day units. *European Journal of Agronomy*, 13, 1–10.
- Bonizzoni, M., Gasperi, G., Chen, X., and James, A. A. (2013). The invasive mosquito species *aedes albopictus*: current knowledge and future perspectives. *Trends in parasitology*, 29(9), 460–468.
- Borkent, A. and Wirth, W. W. (1997). *World species of biting midges (Diptera: Ceratopogonidae)*. American Museum of Natural History.
- Breteler, W. K., Schogt, N., and Van Der Meer, J. (1994). The duration of copepod life stages estimated from stage-frequency data. *Journal of Plankton Research*, 16(8), 1039–1057.
- Briere, J.-F., Pracros, P., Le Roux, A.-Y., and Pierre, J.-S. (1999). A novel rate model of temperature-dependent development for arthropods. *Environmental Entomology*, 28(1), 22–29.
- Bruno, J. F., Ellner, S. P., Vu, I., Kim, K., and Harvell, C. D. (2011). Impacts of aspergillosis on sea fan coral demography: modeling a moving target. *Ecological Monographs*, 81(1), 123–139.
- Burgin, L., Gloster, J., Sanders, C., Mellor, P., Gubbins, S., and Carpenter, S. (2012). Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transboundary and emerging diseases*, 60(3), 263–272.
- Cappé, O., Godsill, S. J., and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95, 899–924.
- Cardinale, B. J., Harvey, C. T., Gross, K., and Ives, A. R. (2003). Biodiversity and biocontrol: emergent impacts of a multi-enemy assemblage on pest suppression and crop yield in an agroecosystem. *Ecology Letters*, 6(9), 857–865.
- Carpenter, S., Groschup, M. H., Garros, C., Felipe-Bauer, M. L., and Purse, B. V. (2013). *Culicoides* biting midges, arboviruses and public health in Europe. *Antiviral Research*, 100, 102–113.
- Carpenter, S., Mellor, P., and Torr, S. (2008). Control techniques for *Culicoides* biting midges and their application in the UK and northwestern Palaearctic. *Medical and veterinary entomology*, 22(3), 175–187.
- Caswell, H. (1983). Phenotypic plasticity in life-history traits: demographic effects and evolutionary consequences. *American Zoologist*, 23, 35–46.

- Caswell, H. (2006). *Matrix Population Models*. 2nd Edition. Sinauer Associates, Sunderland, MA.
- Cavalerie, L., Charron, M. V., Ezanno, P., Dommergues, L., Zumbo, B., and Cardinale, E. (2015). A stochastic model to study Rift Valley fever persistence with different seasonal patterns of vector abundance: New insights on the endemicity in the tropical island of Mayotte. *PloS One*, 10, e0130838.
- Chakraborty, S. and Newton, A. C. (2011). Climate change, plant diseases and food security: an overview. *Plant Pathology*, 60(1), 2–14.
- Chen, K.-C., Wang, T.-Y., Tseng, H.-H., Huang, C.-Y. F., and Kao, C.-Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics*, 21(12), 2883–2890.
- Chevalier, V., Pépin, M., Plee, L., and Lancelot, R. (2010). Rift valley fever-a threat for Europe? *Euro Surveillance*, 15(10).
- Childs, D. Z., Rees, M., Rose, K. E., Grubb, P. J., and Ellner, S. P. (2003). Evolution of complex flowering strategies: an age- and size-structured integral projection model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270, 1829–1838.
- Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). SMC2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 397–426.
- Clements, B. and Rogers, A. (1968). Tests of larvicides for control of salt-marsh sand flies (*Culicoides*). *Mosquito News*, 28(4), 529–534.
- Cloudsley-Thompson, J. (1962). Microclimates and the distribution of terrestrial arthropods. *Annual Review of Entomology*, 7(1), 199–222.
- Clutton-Brock, T. and Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, 25(10), 562–573.
- Coll, C., Horvitz, C. C., and McElderry, R. (2012). Stage-structured periodic population model for the florida leafwing. *Int. J. Complex Systems in Science*, 2(1), 1–5.
- Cooke, K. L. and Grossman, Z. (1982). Discrete delay, distributed delay and stability switches. *Journal of mathematical analysis and applications*, 86(2), 592–627.
- Coulson, T., MacNulty, D. R., Stahler, D. R., vonHoldt, B., Wayne, R. K., and Smith, D. W. (2011). Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history. *Science*, 334, 1275–1278.
- Cowles, M. K. (2013). *Applied Bayesian statistics: with R and OpenBUGS examples*, volume 98. Springer Science & Business Media.
- Crone, E. E., Ellis, M. M., Morris, W. F., Stanley, A., Bell, T., Bierzychudek, P., *et al.* (2013). Ability of matrix models to explain the past and predict the future of plant populations. *Conservation Biology*, 27, 968–978.
- Cushing, J. M., Costantino, R. F., Dennis, B., Desharnais, R., and Henson, S. M. (2002). *Chaos in Ecology: Experimental Nonlinear Dynamics*. Elsevier, New York.
- da Silva Azevedo, R. d. S., Nunes, M. R. T., Chiang, J. O., Bensabath, G., Pinto, N., Martins, L. C., de Oliveira Monteiro, H. A., Rodrigues, S. G., and da Costa Vasconcelos, P. F. (2007). Reemergence of Oropouche fever, northern Brazil. *Emerg Infect Dis*, 13(6), 912.
- Dalgleish, H. J., Koons, D. N., Hooten, M. B., Moffet, C. A., and Adler, P. B. (2011). Climate influences the demography of three dominant sagebrush steppe plants. *Ecology*, 92, 75–85.
- Daszak, P., Cunningham, A. A., and Hyatt, A. D. (2000). Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science*, 287(5452), 443–449.

- De Liberato, C., Purse, B., Goffredo, M., Scholl, F., and Scaramozzino, P. (2003). Geographical and seasonal distribution of the bluetongue virus vector, *Culicoides imicola*, in central Italy. *Medical and veterinary entomology*, 17, 388–394.
- De Reaumur, R. A. F. (1735). Observations du thermomètre, faites à Paris pendant l'année 1735, comparés avec celles qui ont été faites sous la ligne, à l'Isle de France, à Alger et en quelques-unes de nos isles de l'Amérique. *Paris: Mémoires de l'Académie des Sciences*.
- De Valpine, P. (2009). Stochastic development in biologically structured population models. *Ecology*, 90, 2889–2901.
- De Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs*, 72(1), 57–76.
- De Valpine, P. and Knappe, J. (2015). Estimation of general multistage models from cohort data. *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 140–155.
- De Valpine, P., Scranton, K., Knappe, J., Ram, K., and Mills, N. J. (2014). The importance of individual developmental variation in stage-structured population models. *Ecology Letters*, 17, 1026–1038.
- De Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2016). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, (accepted), 1–28.
- Del Moral, P. (1996). Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4), 555–581.
- Dennis, B., Kemp, W. P., and Beckwith, R. C. (1986). Stochastic model of insect phenology: estimation and testing. *Environmental Entomology*, 15, 540–546.
- Diarra, M., Fall, M., Lancelot, R., Diop, A., Fall, A. G., Dicko, A., Seck, M. T., Garros, C., Allène, X., Rakotoarivony, I., et al. (2015). Modelling the abundances of two major *Culicoides* (Diptera: Ceratopogonidae) species in the Niayes area of Senegal. *PLoS One*, 10(6), e0131021.
- Donnet, S. and Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Française de Statistique*, pages to–appear.
- Eagles, D., Melville, L., Weir, R., Davis, S., Bellis, G., Zalucki, M. P., Walker, P. J., and Durr, P. A. (2014). Long-distance aerial dispersal modelling of *Culicoides* biting midges: case studies of incursions into Australia. *BMC veterinary research*, 10(1), 1.
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000a). Climate extremes: observations, modeling, and impacts. *Science*, 289(5487), 2068–2074.
- Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000b). Size-specific sensitivity: applying a new structured population model. *Ecology*, 81, 694–708.
- Ellner, S. P. and Rees, M. (2006). Integral projection models for species with complex demography. *The American Naturalist*, 167, 410–428.
- Ellner, S. P. and Rees, M. (2007). Stochastic stable population growth in integral projection models: theory and application. *Journal of Mathematical Biology*, 54, 227–256.
- Eurich, C. W., Thiel, A., and Fahse, L. (2005). Distributed delays stabilize ecological feedback systems. *Physical review letters*, 94(15), 158104.
- Evans, E. W., Carlile, N. R., Innes, M. B., and Pitigala, N. (2014). Infestation of grain fields and degree-day phenology of the cereal leaf beetle (Coleoptera: Chrysomelidae) in Utah: long-term patterns. *Journal of Economic Entomology*, 107, 240–249.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343–367.

- Faraji, A. and Gaugler, R. (2015). Experimental host preference of diapause and non-diapause induced *Culex pipiens pipiens* (Diptera: Culicidae). *Parasites & vectors*, 8(1), 389.
- Fasiolo, M. and Wood, S. N. (2015). Approximate methods for dynamic ecological models. *arXiv preprint*, arXiv:1511.02644.
- Fontenille, D., Lagneau, C., Lecollinet, S., and Robin, R. L. (2013). *La lutte antivectorielle en France*. IRD Orstom.
- Franco, A., Hill, J. K., Kitchke, C., Collingham, Y. C., Roy, D. B., Fox, R., Huntley, B., and Thomas, C. D. (2006). Impacts of climate warming and habitat loss on extinctions at species' low-latitude range boundaries. *Global Change Biology*, 12(8), 1545–1553.
- Frank, T. and Beek, P. (2001). Stationary solutions of linear stochastic delay differential equations: Applications to biological systems. *Physical Review E*, 64(2), 021917.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67.
- Frigola, R., Lindsten, F., Schön, T. B., and Rasmussen, C. E. (2013). Bayesian inference and learning in Gaussian process state-space models with particle MCMC. In *Advances in Neural Information Processing Systems*, pages 3156–3164.
- Fritsch, F. N. and Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17, 238–246.
- Fujiwara, M. and Caswell, H. (2001). Demography of the endangered north atlantic right whale. *Nature*, 414(6863), 537–541.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL.
- Gerry, A. C., Mullens, B. A., Maclachlan, N. J., and Mecham, J. O. (2001). Seasonal transmission of bluetongue virus by *Culicoides sonorensis* (Diptera: Ceratopogonidae) at a southern California dairy and evaluation of vectorial capacity as a predictor of bluetongue virus transmission. *Journal of medical entomology*, 38(2), 197–209.
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, 327(5967), 812–818.
- Gouno, E., Courtrai, L., and Fredette, M. (2011). Estimation from aggregate data. *Computational Statistics & Data Analysis*, 55(1), 615–626.
- Græsbøll, K., Enøe, C., Bødker, R., and Christiansen, L. E. (2014). Optimal vaccination strategies against vector-borne diseases. *Spatial and spatio-temporal epidemiology*, 11, 153–162.
- Gu, W. and Novak, R. J. (2006). Statistical estimation of degree days of mosquito development under fluctuating temperatures in the field. *Journal of Vector Ecology*, 31, 107–112.
- Gubbins, S., Carpenter, S., Baylis, M., Wood, J. L., and Mellor, P. S. (2008). Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of the Royal Society Interface*, 5, 363–371.
- Gubler, D. J. (2002a). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends in microbiology*, 10(2), 100–103.
- Gubler, D. J. (2002b). The global emergence/resurgence of arboviral diseases as public health problems. *Archives of medical research*, 33(4), 330–342.
- Guichard, S., Guis, H., Tran, A., Garros, C., Balenghien, T., and Kriticos, D. J. (2014). Worldwide niche and future potential distribution of *Culicoides imicola*, a major vector of bluetongue and African horse sickness viruses. *PloS One*, 9(11), e112491.

- Guis, H., Caminade, C., Calvete, C., Morse, A. P., Tran, A., and Baylis, M. (2012). Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *Journal of the Royal Society Interface*, 9, 339–350.
- Haines, A., Kovats, R. S., Campbell-Lendrum, D., and Corvalán, C. (2006). Climate change and human health: impacts, vulnerability and public health. *Public health*, 120(7), 585–596.
- Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo methods*, volume 1. Springer.
- Hanski, I. (1994). Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology & Evolution*, 9(4), 131–135.
- Harris, K. F. and Maramorosch, K. (2014). *Aphids as virus vectors*. Elsevier.
- Hartemink, N., Purse, B., Meiswinkel, R., Brown, H. E., De Koeijer, A., Elbers, A., Boender, G.-J., Rogers, D., and Heesterbeek, J. (2009). Mapping the basic reproduction number (R_0) for vector-borne diseases: a case study on bluetongue virus. *Epidemics*, 1(3), 153–161.
- Hartemink, N., Vanwambeke, S. O., Purse, B. V., Gilbert, M., and Van Dyck, H. (2015). Towards a resource-based habitat approach for spatial modelling of vector-borne disease risks. *Biological Reviews*, 90, 1151–1162.
- Hartig, F. and Dormann, C. F. (2013). Does model-free forecasting really outperform the true model? *Proceedings of the National Academy of Sciences*, 110(42), E3975–E3975.
- Heffernan, J., Smith, R., and Wahl, L. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4), 281–293.
- Hickling, R., Roy, D. B., Hill, J. K., Fox, R., and Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Global change biology*, 12(3), 450–455.
- Higley, L. G., Pedigo, L. P., and Ostlie, K. R. (1986). DEGDAY: a program for calculating degree-days, and assumptions behind the degree-day approach. *Environmental Entomology*, 15, 999–1016.
- Hill, J. K., Griffiths, H. M., and Thomas, C. D. (2011). Climate change and evolutionary adaptations at species' range margins. *Annual review of entomology*, 56, 143–159.
- Hoch, A., Roberts, D., Pinheiro, F., *et al.* (1990). Host-seeking behavior and seasonal abundance of *Culicoides paraensis* (Diptera: Ceratopogonidae) in Brazil. *Journal of the American Mosquito Control Association*, 6(1), 110–114.
- Hodgson, J. A., Thomas, C. D., Oliver, T. H., Anderson, B. J., Brereton, T., and Crone, E. (2011). Predicting insect phenology across space and time. *Global Change Biology*, 17(3), 1289–1300.
- Hoeting, J. A., Tweedie, R. L., and Olver, C. S. (2003). Transform estimation of parameters for stage-frequency data. *Journal of the American Statistical Association*, 98.
- Hoffmann, B., Scheuch, M., Höper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., Eschbaumer, M., Goller, K. V., Wernike, K., Fischer, M., *et al.* (2012). Novel orthobunyavirus in cattle, Europe, 2011. *Emerg Infect Dis*, 18(3), 469–472.
- Holmes, P. and Birley, M. (1987). An improved method for survival rate analysis from time series of haematophagous dipteran populations. *The Journal of Animal Ecology*, pages 427–440.
- Horvitz, C. C. and Schemske, D. W. (1995). Spatiotemporal variation in demographic transitions of a tropical understory herb: projection matrix analysis. *Ecological Monographs*, 65(2), 155–192.
- Hougaard, P. (2012). *Analysis of Multivariate Survival Data*. Springer Science & Business Media, New York.
- Howerth, E. W., Stallknecht, D. E., and Kirkland, P. D. (2001). Bluetongue, epizootic hemorrhagic disease, and other orbivirus-related diseases. *Infectious Diseases of Wild Mammals*, 3, 77–97.
- Hunter, C. M. and Caswell, H. (2005). Selective harvest of sooty shearwater chicks: effects on population dynamics and sustainability. *Journal of animal ecology*, 74(4), 589–600.

- Isaev, V. (1974). Photoperiodic induction of the diapause in the egg phase of the blood-sucking midge, *Culicoides pulicaris punctatus* Mg. (Diptera: Ceratopogonidae). *Parazitologiya*, 9, 501–506.
- Isaev, V. (1976). Effect of environmental factors on the formation and overcoming of diapause in the larvae of *Culicoides odibilis* Austen (Diptera: Ceratopogonidae). *Meditinskaya Parazitologiya i Parazitarnye Bolezni*, 46, 209–214.
- Jacquet, S., Huber, K., Guis, H., Setier-Rio, M.-L., Goffredo, M., Allène, X., Rakotoarivony, I., Chevillon, C., Bouyer, J., Baldet, T., *et al.* (2016). Spatio-temporal genetic variation of the biting midge vector species *Culicoides imicola* (Ceratopogonidae) Kieffer in France. *Parasites & Vectors*, 9, 1.
- Johnson, M. P. (2000). The influence of patch demographics on metapopulations, with particular reference to successional landscapes. *Oikos*, 88(1), 67–74.
- Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990–993.
- Jongejans, E., Shea, K., Skarpaas, O., Kelly, D., and Ellner, S. P. (2011). Importance of individual and environmental variation for invasive species spread: a spatial integral projection model. *Ecology*, 92(1), 86–97.
- Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82, 35–45.
- Kalos, M. H. and Whitlock, P. A. (2008). *Monte Carlo Methods*. John Wiley & Sons.
- Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press.
- Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., *et al.* (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324), 647–652.
- Kemp, W. P., Dennis, B., and Beckwith, R. C. (1986). Stochastic phenology model for the western spruce budworm (Lepidoptera: Tortricidae). *Environmental Entomology*, 15, 547–554.
- Kettle, D. (1962). The bionomics and control of *Culicoides* and *Leptoconops* (Diptera, Ceratopogonidae=Heleidae). *Annual Review of Entomology*, 7, 401–418.
- Kettle, D. (1984). *Ceratopogonidae (Biting Midges)*, volume Medical and Veterinary Entomology. Croom Helm, London.
- King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454, 877–880.
- Knape, J., Daane, K. M., and de Valpine, P. (2014). Estimation of stage duration distributions and mortality under repeated cohort censuses. *Biometrics*, 70, 346–355.
- Knies, J. L. and Kingsolver, J. G. (2010). Erroneous Arrhenius: modified Arrhenius model best explains the temperature dependence of ectotherm fitness. *The American Naturalist*, 176, 227.
- Koons, D. N., Grand, J. B., Zinner, B., and Rockwell, R. F. (2005). Transient population dynamics: relations to life history and initial population state. *Ecological modelling*, 185(2), 283–297.
- Koons, D. N., Holmes, R. R., and Grand, J. B. (2007). Population inertia and its sensitivity to changes in vital rates and population structure. *Ecology*, 88, 2857–2867.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53, 814–861.

- Kuang, Y. (1993). *Delay Differential Equations: with Applications in Population Dynamics*. Academic Press, Boston.
- Kuss, P., Rees, M., Ægisdóttir, H. H., Ellner, S. P., and Stöcklin, J. (2008). Evolutionary demography of long-lived monocarpic perennials: a time-lagged integral projection model. *Journal of Ecology*, 96(4), 821–832.
- Labruna, M. B., Amaku, M., Metzner, J. A., Pinter, A., and Ferreira, F. (2003). Larval behavioral diapause regulates life cycle of amblyomma cajennense (acari: Ixodidae) in southeast brazil. *Journal of medical entomology*, 40(2), 170–178.
- Łacki, M. K. and Miasojedow, B. (2015). State-dependent swap strategies and automatic reduction of number of temperatures in adaptive parallel tempering algorithm. *Statistics and Computing*, pages 1–14.
- Larska, M., Lechowski, L., Grochowska, M., and Żmudziński, J. F. (2013). Detection of the schmallenberg virus in nulliparous culicoides obsoletus/scoticus complex and c. punctatus—the possibility of transovarial virus transmission in the midge population and of a new vector. *Veterinary microbiology*, 166(3), 467–473.
- Law, R. (1983). A model for the dynamics of a plant population containing individuals classified by age and size. *Ecology*, 64(2), 224–230.
- Lefebvre, T., Bruyninckx, H., and De Schutter, J. (2004). Kalman filters for non-linear systems: a comparison of performance. *International Journal of Control*, 77, 639–653.
- Lefkovitch, L. (1965). The study of population growth in organisms grouped by stages. *Biometrics*, pages 1–18.
- Lehmann, K., Werner, D., Hoffmann, B., and Kampen, H. (2012). PCR identification of *Culicoides* biting midges (Diptera: Ceratopogonidae) of the Obsoletus complex including putative vectors of bluetongue and Schmallenberg viruses. *Parasite and Vectors*, 5, 213.
- Leslie, P. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, 35, 213–245.
- Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, pages 183–212.
- Lewis, O., Michels, G., Pierson, E., and Heinz, K. (2015). A predictive degree day model for the development of *Bactericera cockerelli* (Hemiptera: Trioziidae) Infesting Solanum tuberosum. *Environmental Entomology*, page nvv078.
- Lievaart-Peterson, K., Luttikholt, S., Van den Brom, R., and Vellema, P. (2012). Schmallenberg virus infection in small ruminants—First review of the situation and prospects in Northern Europe. *Small ruminant research*, 106(2), 71–76.
- Lillie, T., Kline, D., and Hall, D. (1985). The dispersal of *Culicoides mississippiensis* (Diptera: Ceratopogonidae) in a salt marsh near Yankeetown, Florida. *Journal of the American Mosquito Control Association*, 1(4), 463.
- Lindsey, J. C. and Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in medicine*, 17(2), 219–238.
- Ling, R. F. (1990). Principles of multivariate analysis.
- Linley, J., Hoch, A., and Pinheiro, F. (1983). Biting midges (Diptera: Ceratopogonidae) and human health. *Journal of Medical Entomology*, 20(4), 347–364.
- Linley, J. and Jordan, S. (1992). Effects of ultra-low volume and thermal fog malathion, Scourge and naled applied against caged adult *Culicoides furens* and *Culex quinquefasciatus* in open and vegetated terrain. *Journal of the American Mosquito Control Association*, 8(1), 69–76.
- Linley, J. R. (1985). Growth and survival of *Culicoides melleus* larvae (Diptera: Ceratopogonidae) on four prey organisms. *Journal of medical entomology*, 22(2), 178–189.
- Lipper, L., Thornton, P., Campbell, B. M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., et al. (2014). Climate-smart agriculture for food security. *Nature Climate Change*, 4(12), 1068–1072.

- Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer.
- Logan, J., Wollkind, D., Hoyt, S., and Tanigoshi, L. (1976). An analytic model for description of temperature dependent rate phenomena in arthropods. *Environmental Entomology*, 5(6), 1133–1140.
- Longstaff, B. C. (1984). An extension of the Leslie matrix model to include a variable immature period. *Australian Journal of Ecology*, 9, 289–293.
- Lotka, A. J. (1925). *Elements of physical biology*.
- Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4), 325–337.
- MacDonald, N. (1986). Two delays may not destabilize although either delay can. *Mathematical biosciences*, 82(2), 127–140.
- MacLachlan, N. J. (2011). Bluetongue: History, global epidemiology, and pathogenesis. *Preventive Veterinary Medicine*, 102, 107–111.
- MacLachlan, N. J. and Guthrie, A. J. (2010). Re-emergence of bluetongue, African horse sickness, and other orbivirus diseases. *Veterinary research*, 41(6), 35.
- Malthus, T. R. (1852). *Malthus: essai sur le principe de population*. Guillaumin.
- Manel, S. and Debouzie, D. (1997). Modeling insect development time of two or more larval stages in the field under variable temperatures. *Environmental Entomology*, 26(2), 163–169.
- Manetsch, T. J. (1976). Time-varying distributed delays and their use in aggregative models of large systems. *Systems, Man and Cybernetics, IEEE Transactions on*, pages 547–553.
- Manly, B. F. (1990). *Stage-Structured Populations: Sampling, Analysis and Simulation*. Springer, Netherlands.
- Mao, X., Yuan, C., and Zou, J. (2005). Stochastic differential delay equations of population dynamics. *Journal of Mathematical Analysis and Applications*, 304(1), 296–320.
- Mardulyn, P., Goffredo, M., Conte, A., Hendrickx, G., Meiswinkel, R., Balenghien, T., *et al.* (2013). Climate change and the spread of vector-borne diseases: using approximate Bayesian computation to compare invasion scenarios for the bluetongue virus vector *Culicoides imicola* in Italy. *Molecular Ecology*, 22, 2456–2466.
- May, R. M. *et al.* (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261, 459–467.
- Mayo, C. E., Mullens, B. A., Reisen, W. K., Osborne, C. J., Gibbs, E. P. J., Gardner, I. A., and MacLachlan, N. J. (2014). Seasonal and interseasonal dynamics of bluetongue virus infection of dairy cattle and *Culicoides sonorensis* midges in northern California – implications for virus overwintering in temperate zones. *PloS one*, 9(9), e106975.
- McKendrick, A. and Pai, M. K. (1912). The rate of multiplication of micro-organisms: A mathematical study. *Proceedings of the Royal Society of Edinburgh*, 31, 649–653.
- McMaster, G. S. and Wilhelm, W. (1997). Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology*, 87, 291–300.
- Meiswinkel, R., Gomulski, L., Delécolle, J., Goffredo, M., and Gasperi, G. (2004a). The taxonomy of *Culicoides* vector complexes—unfinished business. *Vet Ital*, 40(3), 151–159.
- Meiswinkel, R., Venter, G., Nevill, E., Coetzer, J., Tustin, R., *et al.* (2004b). Vectors: *Culicoides* spp. *Infectious diseases of livestock*, 1(Ed. 2), 93–136.
- Mellor, P., Boorman, J., and Baylis, M. (2000). *Culicoides* biting midges: their role as arbovirus vectors. *Annual Review of Entomology*, 45, 307–340.

- Mellor, P. and Wittmann, E. (2002). Bluetongue virus in the Mediterranean Basin 1998–2001. *The Veterinary Journal*, 164(1), 20–37.
- Mellor, P. S. and Hamblin, C. (2004). African horse sickness. *Veterinary research*, 35(4), 445–466.
- Merow, C., Dahlgren, J. P., Metcalf, C. J. E., Childs, D. Z., Evans, M. E., Jongejans, E., *et al.* (2014). Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution*, 5, 99–110.
- Merrill, S. C., Gebre-Amlak, A., Armstrong, J. S., and Peairs, F. B. (2010). Nonlinear degree-day models for postdiapause development of the sunflower stem weevil (Coleoptera: Curculionidae). *Journal of Economic Entomology*, 103, 302–307.
- Metcalf, C. J. E., Graham, A. L., Martinez-Bakker, M., and Childs, D. Z. (2015). Opportunities and challenges of integral projection models for modelling host–parasite dynamics. *Journal of Animal Ecology*.
- Metcalf, C. J. E., Horvitz, C. C., Tuljapurkar, S., and Clark, D. A. (2009). A time to grow and a time to die: a new way to analyze the dynamics of size, light, age, and death of tropical trees. *Ecology*, 90, 2766–2778.
- Mills, J. N., Gage, K. L., and Khan, A. S. (2010). Potential influence of climate change on vector-borne and zoonotic diseases: a review and proposed research plan. *Environmental health perspectives*, 118(11), 1507.
- Morellato, L. P. C., Alberti, L., and Hudson, I. L. (2010). *Applications of circular statistics in plant phenology: a case studies approach*. Springer.
- Mullens, B., Gerry, A., Lysyk, T., and Schmidtmann, E. (2004). Environmental effects on vector competence and virogenesis of bluetongue virus in *Culicoides*: interpreting laboratory data in a field context. *Vet Ital*, 40(3), 160–166.
- Mullens, B., McDermott, E., and Gerry, A. (2015). Progress and knowledge gaps in *Culicoides* ecology and control. *Veterinaria Italiana*, 51, 313–323.
- Mullens, B. and Rutz, D. (1983). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae) at constant laboratory temperatures. *Annals of the Entomological Society of America*, 76, 747–751.
- Munholland, P. L. and Kalbfleisch, J. D. (1991). A semi-markov model for insect life history data. *Biometrics*, pages 1117–1126.
- Murray, J. D. (2002). *Mathematical biology I: an introduction, Vol. 17 of interdisciplinary applied mathematics*. Springer, New York, NY, USA,.
- Murray, K. A. and Daszak, P. (2013). Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence. *Current opinion in virology*, 3(1), 79–83.
- Murtaugh, P. A., Emerson, S. C., McEvoy, P. B., and Higgs, K. M. (2012). The statistical analysis of insect phenology. *Environmental Entomology*, 41, 355–361.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. 2nd Edition. Springer-Verlag, New York.
- Nelson, W. A., Bjørnstad, O. N., and Yamanaka, T. (2013). Recurrent insect outbreaks caused by temperature-driven changes in system stability. *Science*, 341, 796–799.
- Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., L, B. D., and Cole, D. J. (2014). *Modelling population dynamics: model formulation, fitting and assessment using state-space methods*. Springer.
- NIMBLE Development Team (2016). *NIMBLE Users Manual, Version 0.5*.
- Nisbet, R., Blythe, S., Gurney, W., and Metz, J. (1985). Stage-structure models of populations with distinct growth and development processes. *Mathematical Medicine and Biology*, 2(1), 57–68.
- Nisbet, R. and Gurney, W. (1983). The systematic formulation of population models for insects with dynamically varying instar duration. *Theoretical Population Biology*, 23(1), 114–135.

- Nisbet, R. M. (1997). Delay-differential equations for structured populations. In *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*, pages 89–118. Springer, US.
- Osawa, A., Shoemaker, C. A., and Stedinger, J. R. (1983). A stochastic model of balsam fir bud phenology utilizing maximum likelihood parameter estimation. *Forest Science*, 29, 478–490.
- Ozgul, A., Coulson, T., Reynolds, A., Cameron, T. C., and Benton, T. G. (2012). Population responses to perturbations: the importance of trait-based analysis illustrated through a microcosm experiment. *The American Naturalist*, 179, 582–594.
- Ozgul, A., Tuljapurkar, S., Benton, T. G., Pemberton, J. M., Clutton-Brock, T. H., and Coulson, T. (2009). The dynamics of phenotypic change and the shrinking sheep of St. Kilda. *Science*, 325(5939), 464–467.
- Parmesan, C. and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918), 37–42.
- Pearl, R. and Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences*, 6(6), 275–288.
- Pegram, R., Mwase, E., Zivkovic, D., and Jongejan, F. (1988). Morphogenetic diapause in amblyomma variegatum (acari: Ixodidae). *Medical and veterinary entomology*, 2(4), 301–307.
- Perretti, C. T., Munch, S. B., and Sugihara, G. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proceedings of the National Academy of Sciences*, 110(13), 5253–5257.
- Pertierra, J. P., Lleonart, J., and Lo, N. C. (1997). Application of a stage-specific matrix model and length-cohort based analysis to assess the anchovy fishery in Catalan coastal waters (NW Mediterranean Sea). *Fisheries Research*, 30(1-2), 127–137.
- Petchey, O. L., Pontarp, M., Massie, T. M., Kéfi, S., Ozgul, A., Weilenmann, M., Palamara, G. M., Altermatt, F., Matthews, B., Levine, J. M., *et al.* (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- Pinheiro, F. P., Hoch, A. L., Gomes, M., and Roberts, D. R. (1981a). Oropouche virus. iv. Laboratory transmission by *Culicoides paraensis*. *The American journal of tropical medicine and hygiene*, 30(1), 172–176.
- Pinheiro, F. P., Travassos da Rosa, A., Travassos da Rosa, J., *et al.* (1981b). Oropouche virus. i. a review of clinical, epidemiological, and ecological findings. *American Journal of Tropical Medicine and Hygiene*, 30(1), 149–160.
- Pinheiro, F. P., Travassos da Rosa, A., and Vasconcelos, P. F. (1998). An overview of Oropouche fever epidemics in Brazil and neighbour countries. In *An overview of arbovirology in Brazil and neighbouring countries*, pages 186–92. Instituto Evandro Chagas.
- Pioz, M., Guis, H., Crespín, L., Gay, E., Calavas, D., Durand, B., Abrial, D., and Ducrot, C. (2012). Why did bluetongue spread the way it did? environmental factors influencing the velocity of bluetongue virus serotype 8 epizootic wave in France. *PloS One*, 7, e43360.
- Plant, R. E. and Wilson, L. (1986). Models for age structured populations with distributed maturation rates. *Journal of Mathematical Biology*, 23, 247–262.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Plummer, M. *et al.* (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution*, 16(12), 1791–1798.

- Purse, B., Carpenter, S., Venter, G., Bellis, G., and Mullens, B. (2015). Bionomics of temperate and tropical *Culicoides* midges: knowledge gaps and consequences for transmission of *Culicoides*-borne viruses. *Annual Review of Entomology*, 60, 373–392.
- Purse, B., Tatem, A., Caracappa, S., Rogers, D., Mellor, P., Baylis, M., *et al.* (2004). Modelling the distributions of *Culicoides* bluetongue virus vectors in Sicily in relation to satellite-derived climate variables. *Medical and Veterinary Entomology*, 18, 90–101.
- Purse, B. V., Mellor, P. S., Rogers, D. J., Samuel, A. R., Mertens, P. P., and Baylis, M. (2005). Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology*, 3, 171–181.
- Randolph, S. E. (1997). Abiotic and biotic determinants of the seasonal dynamics of the tick *Rhipicephalus appendiculatus* in South Africa. *Medical and veterinary entomology*, 11(1), 25–37.
- Rasmussen, D. A., Ratmann, O., and Koelle, K. (2011). Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput Biol*, 7(8), e1002136.
- Read, K. and Ashford, J. (1968). A system of models for the life cycle of a biological organism. *Biometrika*, 55, 211–221.
- Régnière, J., Powell, J., Bentz, B., and Nealis, V. (2012). Effects of temperature on development, survival and reproduction of insects: experimental design, data analysis and modeling. *Journal of Insect Physiology*, 58, 634–647.
- Régnière, J. and Powell, J. A. (2013). Animal life cycle models (Poikilotherms). In *Phenology: An Integrative Environmental Science*, pages 295–316. Springer, Netherlands.
- Roberts, D. R., Hoch, A. L., Dixon, K. E., and Llewellyn, C. H. (1981). Oropouche virus. iii. Entomological observations from three epidemics in Para, Brazil, 1975. *The American journal of tropical medicine and hygiene*, 30(1), 165–171.
- Roberts, G. O. and Rosenthal, J. S. (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, pages 2123–2139.
- Rodriguez, M., Hooghuis, H., and Castaño, M. (1992). African horse sickness in Spain. *Veterinary Microbiology*, 33(1), 129–142.
- Routray, A., Pradhan, A. K., and Rao, K. P. (2002). A novel Kalman filter for frequency estimation of distorted signals in power systems. *Instrumentation and Measurement, IEEE Transactions on*, 51, 469–479.
- Russelle, M., Wilhelm, W., Olson, R., and Power, J. F. (1984). Growth analysis based on degree-days. *Crop Science*, 24, 28–32.
- Salguero-Gomez, R. and Plotkin, J. B. (2010). Matrix dimensions bias demographic inferences: implications for comparative plant demography. *The American Naturalist*, 176(6), 710–722.
- Sanders, C. J., Shortall, C. R., Gubbins, S., Burgin, L., Gloster, J., Harrington, R., Reynolds, D. R., Mellor, P. S., and Carpenter, S. (2011). Influence of season and meteorological parameters on flight activity of *Culicoides* biting midges. *Journal of Applied Ecology*, 48, 1355–1364.
- Schaalje, G. B. and van der Vaart, H. (1989). Relationships among recent models for insect population dynamics with variable rates of development. *Journal of Mathematical Biology*, 27(4), 399–428.
- Schoolfield, R., Sharpe, P., and Magnuson, C. (1981). Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. *Journal of Theoretical Biology*, 88(4), 719–731.
- Scranton, K., Knappe, J., and de Valpine, P. (2014). An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. *Ecology*, 95, 1418–1428.
- Searle, K., Blackwell, A., Falconer, D., Sullivan, M., Butler, A., and Purse, B. (2012). Identifying environmental drivers of insect phenology across space and time: *Culicoides* in Scotland as a case study. *Bulletin of Entomological Research*, 1, 1–16.

- Sellers, R., Pedgley, D., and Tucker, M. (1977). Possible spread of african horse sickness on the wind. *Journal of Hygiene*, 79(02), 279–298.
- Severini, M., Baumgärtner, J., and Limonta, L. (2003). Parameter estimation for distributed delay based population models from laboratory data: egg hatching of *Oulema duftschmidi* Redthenbacher (Coleoptera, Chrysomelidae) as an example. *Ecological Modelling*, 167(3), 233–246.
- Sharpe, P. J. and DeMichele, D. W. (1977). Reaction kinetics of poikilotherm development. *Journal of Theoretical Biology*, 64, 649–670.
- Shulgin, B., Stone, L., and Agur, Z. (1998). Pulse vaccination strategy in the SIR epidemic model. *Bulletin of mathematical biology*, 60(6), 1123–1148.
- Shyu, E. and Caswell, H. (2016). Frequency-dependent two-sex models: a new approach to sex ratio evolution with multiple maternal conditions. *Ecology and Evolution*, 6(19), 6855–6879.
- Smallegange, I. M., Caswell, H., Toorians, M. E., and Roos, A. é. M. (2016). Mechanistic description of population dynamics using dynamic energy budget theory incorporated into integral projection models. *Methods in Ecology and Evolution*.
- Smallegange, I. M., Deere, J. A., and Coulson, T. (2014). Correlative changes in life-history variables in response to environmental change in a model organism. *The American Naturalist*, 183(6), 784–797.
- Smith, A. F. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W., and Lunn, D. (1996). Bugs: Bayesian inference using gibbs sampling. *Version 0.5,(version ii)* <http://www.mrc-bsu.cam.ac.uk/bugs>, 19.
- Stott, I., Hodgson, D. J., and Townley, S. (2012). Beyond sensitivity: nonlinear perturbation analysis of transient dynamics. *Methods in Ecology and Evolution*, 3, 673–684.
- Stott, I., Townley, S., and Hodgson, D. J. (2011). A framework for studying transient dynamics of population projection matrix models. *Ecology Letters*, 14, 959–970.
- Swendsen, R. H. and Wang, J.-S. (1986). Replica Monte Carlo simulation of spin-glasses. *Physical Review Letters*, 57, 2607.
- Tabachnick, W. J. (1996). *Culicoides variipennis* and bluetongue-virus epidemiology in the United States. *Annual Review of Entomology*, 41, 23–43.
- Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145(2), 505–518.
- Taylor, F. (1981). Ecology and evolution of physiological time in insects. *American naturalist*, pages 1–23.
- Taylor, L. H., Latham, S. M., and Mark, E. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1411), 983–989.
- Telford, M. J., Bourlat, S. J., Economou, A., Papillon, D., and Rota-Stabelli, O. (2008). The evolution of the Ecdysozoa. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1496), 1529–1537.
- Tenhumberg, B., Tyre, A. J., and Rebarber, R. (2009). Model complexity affects transient population dynamics following a dispersal event: A case study with pea aphids. *Ecology*, pages 1878–1890.
- Tesh, R. B. (1994). The emerging epidemiology of Venezuelan hemorrhagic fever and Oropouche fever in tropical South America. *Annals of the New York Academy of Sciences*, 740(1), 129–137.
- Tuljapurkar, S., Steiner, U. K., and Orzack, S. H. (2009). Dynamic heterogeneity in life histories. *Ecology Letters*, 12, 93–106.

- Turner, J., Bowers, R. G., and Baylis, M. (2013). Two-host, two-vector basic reproduction ratio (R_0) for bluetongue. *PloS One*, 8, e53128.
- Van Tienderen, P. H. (1995). Life cycle trade-offs in matrix population models. *Ecology*, pages 2482–2489.
- Vasconcelos, P. F., Travassos da Rosa, A., Rodrigues, S. G., Travassos da Rosa, E. S., Dégallier, N., da Rosa, T., and Jorge, F. (2001). Inadequate management of natural ecosystem in the Brazilian Amazon region results in the emergence and reemergence of arboviruses. *Cadernos de Saúde Pública*, 17, S155–S164.
- Vaughan, J. and Turner, E. (1987). Development of immature *Culicoides variipennis* (Diptera: Ceratopogonidae) from Saltville, Virginia, at constant laboratory temperatures. *Journal of Medical Entomology*, 24, 390–395.
- Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a. *Quetelet*, 10, 113–121.
- Viennet, E., Garros, C., Gardes, L., Rakotoarivony, I., Allene, X., Lancelot, R., Crochet, D., Moulia, C., Baldet, T., and Balenghien, T. (2013). Host preferences of Palaearctic *Culicoides* biting midges: implications for transmission of orbiviruses. *Medical and Veterinary Entomology*, 27, 255–266.
- Vindenes, Y., Engen, S., and Sæther, B.-E. (2008). Individual heterogeneity in vital parameters and demographic stochasticity. *The American Naturalist*, 171(4), 455–467.
- Vindenes, Y. and Langangen, Ø. (2015). Individual heterogeneity in life histories and eco-evolutionary dynamics. *Ecology Letters*, 18, 417–432.
- Watts, D. M., Phillips, I., Callahan, J. D., Griebenow, W., Hyams, K. C., and Hayes, C. G. (1997). Oropouche virus transmission in the Amazon River Basin of Peru. *The American journal of tropical medicine and hygiene*, 56(2), 148–152.
- Watts, N., Adger, W. N., Agnolucci, P., Blackstock, J., Byass, P., Cai, W., Chaytor, S., Colbourn, T., Collins, M., Cooper, A., *et al.* (2015). Health and climate change: policy responses to protect public health. *The Lancet*, 386(10006), 1861–1914.
- Wearing, H. J., Rohani, P., Cameron, T. C., and Sait, S. M. (2004). The dynamical consequences of developmental variability and demographic stochasticity for host-parasitoid interactions. *The American Naturalist*, 164, 543–558.
- Wheeler, T. and Von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341(6145), 508–513.
- Wilson, A. J. and Nussey, D. H. (2010). What is individual quality? An evolutionary perspective. *Trends in Ecology and Evolution*, 25, 207–214.
- Wong, M., Lam, K., and Lo, E. (2005). Bayesian analysis of clustered interval-censored data. *Journal of dental research*, 84(9), 817–821.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466, 1102–1104.
- Yamanaka, T., Nelson, W. A., Uchimura, K., and Bjørnstad, O. N. (2012). Generation separation in simple structured life cycles: Models and 48 years of field data on a tea tortrix moth. *The American Naturalist*, 179(1), 95–109.
- Young, H. S., Parker, I. M., Gilbert, G. S., Guerra, A. S., and Nunn, C. L. (2016). Introduced species, disease ecology, and biodiversity–disease relationships. *Trends in Ecology & Evolution*.
- Yurk, B. P. and Powell, J. A. (2010). Modeling the effects of developmental variation on insect phenology. *Bulletin of Mathematical Biology*, 72(6), 1334–1360.
- Zuidema, P. A., Brienens, R. J., During, H. J., and Güneralp, B. (2009). Do persistently fast-growing juveniles contribute disproportionately to population growth? A new analysis tool for matrix models and its application to rainforest trees. *The American Naturalist*, 174, 709–719.