

Académie de Montpellier

Université Montpellier II

- Sciences et Techniques du Languedoc -

Habilitation à Diriger des Recherches

Des modèles de capture-recapture pour l'écologie évolutive

Olivier Gimenez

Soutenue le 29 Mars 2010, devant la commission d'examen :

Dominique Pontier	Professeur, Université Claude Bernard, Lyon	Rapporteur
Denis Couvet	Professeur, MNHN, Paris	Rapporteur
James D. Nichols	Directeur de recherche (équivalent), USGS, Etats-Unis	Rapporteur
Isabelle Olivieri	Professeur, UM2, Montpellier	Examineur
Thierry Boulinier	Directeur de recherche, CNRS, Montpellier	Examineur

Capture-recapture models for evolutionary ecology

In evolutionary ecology, the increasing availability of data collected in time allows to understand evolutionary processes in natural conditions better than ever. Biologists are more and more interested in testing empirically evolutionary forces acting on wild populations. While these studies are needed to investigate mechanisms in natural environments that are not easily reproduced in the lab, they raise methodological issues. In particular, measuring fitness in the field is a difficult task. Ideally, estimating fitness requires an exhaustive monitoring of all individuals of a population, from birth to death. In practice however, it is unrealistic to follow all the elements of a population. Individuals can be seen (or captured) or not at several occasions during their life, which raised the issue of detection probability < 1 .

While the issue of an imperfect detectability has long been realized in ecology and conservation biology, it is surprising to see how much it has been neglected in evolutionary ecology. Most often, it is assumed that the detection probability is 1 so that standard statistical methods (regression or survival analyses) can be used. However, because this assumption is equivalent to assume that an individual seen for the last time before the end of the study is dead, mortality (among other parameters) is underestimated if the detection probability is actually < 1 .

Capture-Mark-Recapture (CMR ; see Encadré 1) models explicitly incorporate the detection process in the estimation of demographic parameters (e.g. survival, breeding or dispersal probabilities). However, the potential of CMR methods has been little explored in evolutionary ecology, and several discussions with colleagues have convinced me that a didactical effort was needed to show that accounting for detectability was important. To do so, what's better than revisiting popular questions in evolutionary ecology using CMR modelling tools ? (see Gimenez et al. 2008 ; P40 and Figure 1) ?

Here is the Pandora's box open ! While demographic parameters can be influenced by several individual characteristics such as age, sexe, social status, phenotype or genotype¹, these various sources of variability are imperfectly incorporated in CMR models, if not just ignored. More annoying, heterogeneity among individuals is often considered as a nuisance that causes bias in the parameter estimates ; efforts to evaluate and characterize this heterogeneity are therefore motivated by the need to reduce this bias, and consequently most work in this area has been produced by statisticians mainly. But, isn't the detection and characterization of differences among individuals or groups of individuals one of the main objectives in evolutionary ecology, and in particular differences affecting fitness components (hence the demographic parameters) ?

In this document, I argue that individual variability should be emphasized while estimating and making inference with CMR models, and not just treated as a nuisance. To illustrate my point, I go through four important questions in evolutionary ecology (and four chapters) that can be fruitfully addressed using CMR models.

This motivation explains why I have asked to apply for an HDR in a Biology section rather than a Math section through which I got my PhD. My work is clearly at the interface of modeling, applied statistics and population biology : case studies generate biological questions which, in turn, motivate the development of models and statistical methods ; to me, it is also important to insure the transfer of knowledge and techniques by supervising students, teaching, workshops, developing and making computer programs available. I'd therefore like my work to be evaluated by biologists.

Without going into a psychanalysis, I must confess that it took me several years to understand and assume this place at work, my place. This is probably due to the French system which favors math studies a lot, and produces some sort of elite. My post-docs in the UK have helped me to make these barriers fall and to get rid of a kind of complex. I'm also deeply indebted to my lab, the Centre d'Ecologie Fonctionnelle et Evolutive, and to the ingeniors, technicians, admins, students, researchers and teachers who make it alive.

Enjoy your reading !

1. Demographic parameters can also be influenced by abiotic environmental factors (climate, temperature, rainfall) or biotic environmental factors (competition, predation, diseases). Studying the impact of abiotic factors on demographic parameters is also a major component of my research. Rather than being exhaustive, I have focused here on pieces of my work that look the most promising.

1

Describing senescence patterns

This chapter deals with methods to investigate senescence in survival. In particular, I show how to describe age-specific survival patterns while accounting for individual heterogeneity.

Section 1.2 is a short introduction where I explain how individual heterogeneity can mask senescence. Using simulated data, Figure 2A illustrates the issue with discrete heterogeneity, Figure 2B with continuous heterogeneity.

In section 1.3, I introduce mixture models (à la Shirley, see Encadré 3) to incorporate discrete heterogeneity and random effects models to incorporate continuous heterogeneity. The later is based on Gimenez and Choquet (2010 ; P1, see Encadré 2).

In section 1.4, I go through two case studies. The first one (section 1.4.1) is based on Marzolin, Charmantier and Gimenez (to be submitted ; S38) and uses a random effect model to incorporate a frailty on survival (Figure 3). The second one is based on Péron et al. (2010 ; P33) and uses mixture models to deal with heterogeneity in several parameters (Figure 4).

2

Investigating life-history tradeoffs

In this chapter, I show how to investigate life-history tradeoffs using mark-recapture data and two families of models (multistate models and state-space models).

In section 2.2, I introduce multistate models using a fake example with two states breeder and non-breeder to illustrate survival vs. reproduction and reproduction vs. reproduction tradeoffs. This is an excuse to talk about some methodological developments I did during my PhD thesis, namely goodness-of-fit tests in section 2.2.1 (based on Pradel et al. 2003 ; P12 and Pradel et al. 2005 ; P9), testing parameter redundancy (based on Gimenez et al. 2003 ; P11 and Gimenez et al. 2004 ; P10, see also Figure 5) and building confidence intervals (based on Gimenez et al. 2005 ; P8, see also Figure 7).

In section 2.3, I introduce state-space models to analyse mark-recapture data as I proposed in Gimenez et al. (2007 ; P6, see Encadré 4). Using this framework, I illustrate Mathieu Buoro's current work (one of my PhD students) on investigating life-history tradeoffs in the Atlantic salmon (Buoro et al. 2010 ; P34, see also Figure 7).

3

Exploring fitness surfaces

In this chapter, I show how to study natural selection acting on quantitative traits with mark-recapture data.

In section 3.2, I propose an extension of the Lande and Arnold method used to calculate selection gradients to the general case $p < 1$. This is illustrated with a single trait (section 3.2.1, see Table 1) and several traits (section 3.2.2, see Table 3).

In section 3.3, I introduce a new method (Gimenez et al. 2006 ; P7 and Gimenez and Barbraud 2009 ; CL29) based on splines to produce a visual representation of fitness surfaces. This is illustrated with a single trait (section 3.3.1, see Figure 8) and several traits (section 3.3.2, see Figure 9).

The case studies are based on Gimenez et al. (2006 ; P41) and Gimenez et al. (2009 ; P39).

4

Quantifying heritability in demographic parameters

In this chapter, I introduce a new state-space model combining pedigree information and mark-recapture data (Figure 10) to estimate heritability of demographic parameters. This work is based on a threshold model (Figure 11) and is illustrated with data on Blue tits (Figure 13). This is based on Papaix et al. (submitted ; S37).

Conclusions and thoughts

Which CMR models for evolutionary ecology ?

Running the risk of looking like a CMR « integrist », I have insisted a lot on the importance of accounting for the detection probability. However, if in some cases ignoring this parameter can lead to flawed inference (see introduction), in other cases the difference can be negligible (e.g. if detectability is high and constant through time, as in chapter 4). Nevertheless, it is difficult to give reliable rules to say whether the detectability issue can be ignored and standard statistical methods be used in place of CMR methods. As a consequence, I highly recommend using CMR methods. Regarding previous studies, I do not claim that they are all wrong, but rather that it is difficult to know whether we can trust the results without conducting a CMR reanalysis (see Nichols et al. 1997 who first made this point, and Gaillard et al. 1994 for an example).

My main motivation when writing this manuscript (besides the diploma...) was to put the emphasis on individual variability while estimating demographic parameters and making inference using CMR models. Have I succeeded in ? I hope so, at least in this document, through the examination of four important questions in evolutionary ecology. My contribution is mainly methodological via the treatment of heterogeneity in CMR models. In brief, if measurable indicators of intra-population heterogeneity are available, they can be integrated in the analyses (age, chapter 1 ; morphological traits, chapters 2 and 3). If the origin of heterogeneity is unknown and cannot be quantified using observable criteria (the concept of frailty introduced in chapter 1), then we need mixture models (chapter 1) or random effects models (all chapters). Both approaches are complementary (chapters 1 and 3).

Dealing with a frailty requires models with a hidden structure (hidden Markov or state-space models, Encadrés 3 and 4). While complex from a numerical or statistical point of view, I think these models correspond closely to the intuition of a biologist about what occurs on the field and

the mechanisms involved in the study system : focus on the way you think it works (the hidden process) rather than what you see (a noisy observation of the biological signal).

I'm convinced that one of the most promising avenue to incorporate a frailty is the development of CMR mixed models (Gimenez and Choquet 2010 ; P1). These models allow to account for individual heterogeneity by incorporating random effects¹ on top of the trend or the general pattern considered in generalized linear models with fixed effects (e.g. Bolker et al. 2009). CMR mixed models is the focus of my current research through Sarah Cubaynes's PhD who studies several case studies and methodological developments and Eleni Papadatou's post-doc on multi-species and multi-population CMR models (Papadatou et al. submitted ; S21 ; see also Grosbois et al. 2009 ; P24).

Towards eco-evo-statistics

Writing my HDR is an opportunity to sit down and think of the way I see my work. Am I a statistician ? Nop, I do not actually develop new statistical methods. Am I a biometrician then ? I don't like the word because it makes me think of the passport with the same name... Do I like biostatistician ? Even though it is the title of my PhD, we see that this term is monopolized by medical sciences and bioinformatics (just check the table of contents of the last issue of the journal *Biostatistics*). I'm tempted by the term ecological statistician or statistical ecologist used by my british colleagues (<http://www.ncse.org.uk/>), but this ignores evolution. To reflect my activity, I would use eco-evo statistician.

More seriously, and to conclude, I'd like to insist on my attachment to collaborations between biologists and methodologists to which I give much importance (Figure 14) ; these collaborations are articulated as follows.

- The development and popularization (papers, workshops, teaching, ...) of computer programs to analyze data and ensure the transfer of methods. This aspect is an important component of the work done in our team and led by our « informathematicians » (Rémi Choquet and Erika Nogué).
- The data and more precisely the long-term monitoring datasets which are so precious to biologists and methodologists. The importance of fieldwork should be emphasized here (in France, I have in mind the work done by the Museum National d'Histoire Naturelle, the

1. I think of genetic differences (chapter 4) among individuals but also of the influence of conditions during development (e.g. family or parental effects) or the conditions experienced later on in life (e.g. exposition to a pathogen).

Office National de la Chasse et de la Faune Sauvage, the Terres Australes et Antarctiques Françaises and the CEFE of course). I'd like to thank them all for their tremendous work !

- Interacting with students in the broad sense via teaching, supervising and workshops. What I can teach is mainly why, when and how to use and adapt existing methods and concepts to address questions in population biology. I also hope to inject some pragmatism in my teaching². In return, I'd like my students to be convinced that the time spent together is an inspiring source of learning for me (among others...).
- Last but not least, collaborations mean meeting with new people. The four chapters of this document and more generally my everyday work has been, is and will be a product of human interactions. They will recognize themselves.

2. The polemic bayesian vs. frequentist is a good illustration. Rather than opposing the two approaches, one can choose one method or the other according to what's needed (Gimenez 2008 ; P5), or even use them in combination (see Gimenez and Barbraud 2009 ; CL29 and Servanty et al. submitted ; S31 for examples).