

Resource selection studies with marked animals: making inference to average selection with GLMMs

Valentina La Morgia

Istituto Superiore per la Protezione e la Ricerca Ambientale

valentina.lamorgia@unito.it

March 25, 2024

Motivation

In the analysis of resource selection by animals, datasets are usually made up of a series of presence vs. (pseudo)absence data (binomial distribution)

Motivation

In the analysis of resource selection by animals, datasets are usually made up of a series of presence vs. (pseudo)absence data (binomial distribution)

- **Adequate quantities of usable resources are necessary to sustain animal populations.**

The need to document the availability of resources is critical in efforts to preserve endangered species and manage exploited populations - *old growth forest is vital to the continued existence of the spotted owl (*Strix occidentalis*)?*

Motivation

In the analysis of resource selection by animals, datasets are usually made up of a series of presence vs. (pseudo)absence data (binomial distribution)

- Determining which resources (or conditions) are selected more often than others provides fundamental information about the nature of animals and how they meet their requirements for survival

Motivation

In the analysis of resource selection by animals, datasets are usually made up of a series of presence vs. (pseudo)absence data (binomial distribution)

- Differential resource selection is one of the principal relationships which permit species to coexist - *evaluation of the effect of domestic animals on wild animal forage*

Motivation

In the analysis of resource selection by animals, datasets are usually made up of a series of presence vs. (pseudo)absence data (binomial distribution)

- ▶ The process of natural selection can occur when resource selectivity results in successful (e.g., breeding) and unsuccessful individuals - *food and oviposition selection play a significant role in evolution and speciation among some insects*

Basic concepts

It is often assumed that a species will select resources that are best able to satisfy its life requirements, and that high quality resources will be selected more than low quality ones.

The availability of various resources is not generally uniform in nature, and use may change as availability changes.

Therefore, *used resources should be compared to available (or unused) resources* in order to reach valid conclusions concerning resource selection.

Definitions

Usage

The usage of a resource is defined as that quantity of the resource that is utilized by an animal (or population of animals) in a fixed period of time.

Abundance and availability

The abundance of a resource is the quantity of the resource in the environment.

The availability of a resource is the quantity accessible to the animal (or population of animals) during that same period of time.

Selection and preference

When resources are used disproportionately to their availability, use is said to be selective.

Selection is the process in which an animal chooses a resource, and preference is the likelihood that a resource will be selected if offered on an equal basis with others.

A case study

Animal Behaviour 82 (2011) 791–799



Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav



To lay or not to lay: oviposition of *Maculinea arion* in relation to *Myrmica* ant presence and host plant phenology

Dario Patricelli*, Francesca Barbero, Valentina La Morgia, Luca P. Casacci, Magdalena Witek, Emilio Balletto, Simona Bonelli

Department of Animal and Human Biology, University of Turin

ARTICLE INFO

Article history:

Received 19 November 2010
Initial acceptance 12 January 2011
Final acceptance 1 July 2011
Available online 6 August 2011
MS. number: 10-00810R

Keywords:

host selection
Maculinea
myrmecophily
Origanum vulgare
oviposition
social parasite

The choices made by ovipositing female butterflies play a key role in the survival of their offspring and consequently in the persistence of butterfly populations. These choices are even more crucial in the case of obligate myrmecophilous organisms such as *Maculinea* butterflies with larvae that, after a phytophagous period, need to be adopted by *Myrmica* ants to complete their life cycle. Because the worker ants' foraging range is limited, selecting an 'ideal' oviposition site requires that both the phenological stage of the larval food plant (short-term larval fitness) and the presence of suitable host ants (long-term larval fitness) are taken into account. Whether the female's selection of a valuable oviposition plant is influenced by the closeness of a *Myrmica* nest is unclear. We studied the oviposition behaviour of a *Maculinea arion* population exploiting *Origanum vulgare* as a host plant. By following females, we collected phenological data on the visited plants that were either 'chosen' for oviposition or 'avoided' (flowers were visited and evaluated, but received no eggs), and we assessed the presence of *Myrmica* ants in the vicinity of each plant. Results suggest that plants are selected by *M. arion* females on the basis of their bud phenology and the presence of host ants and not of other environmental features. We thus hypothesize the evolution of an adaptive mechanism that affords females of this strictly myrmecophilous butterfly the ability to ensure the long-term survival of their brood by selecting host plants growing near a *Myrmica* nest.

© 2011 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

A case study



Data

```

1  2  3  4  5  6  7  8  9 10 12 13 14 15 16 17 18 19 20 21 22 2
4  2  5  9  9 12 13 19  3 27 31 10 12  2  5  9  9  2  7 17  4

```

```

'data.frame':      214 obs. of  13 variables:
 $ X                : int  1 2 3 4 5 6 7 8 9 10 ...
 $ id               : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Egg10            : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MyrmicaYN        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Noperaiepitfall  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Horigano         : num  50 50 30 60 70 37.5 50 45 25 55 ..
 $ HGrass           : int  10 10 10 40 15 15 10 15 15 5 ...
 $ Origanupatchdiameter: int  50 50 3 60 190 50 120 90 15 25 ...
 $ BV               : int  0 0 1 1 1 0 1 1 1 1 ...
 $ BVR              : int  0 0 0 0 0 0 0 0 0 0 ...
 $ BR               : int  1 1 0 0 0 1 0 0 0 0 ...
 $ BF               : int  1 1 0 0 0 1 0 0 0 0 ...
 $ NB               : int  0 0 0 0 0 0 1 0 1 0 ...

```

Study design

Three general study designs for evaluating resource selection have been identified in the literature - they differ with respect to the level at which resource use and availability are measured, at the population level or for each animal.

Design I

Measurement are made at the population level. Individual animals are not identified.

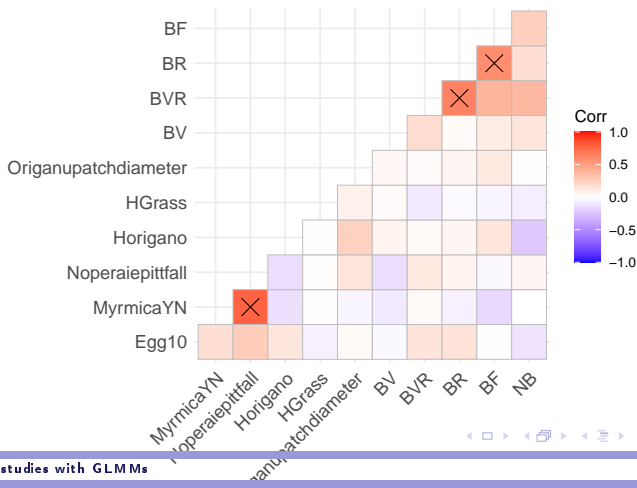
Design II

Individual animals are identified and the use of resources is measured for each, but availability is measured at the population level.

Design III

Individual animals are identified and both the use of resources and their availability are measured for each marked animal.

Data exploration - exercise 1



Model fitting: the base model - exercise 2

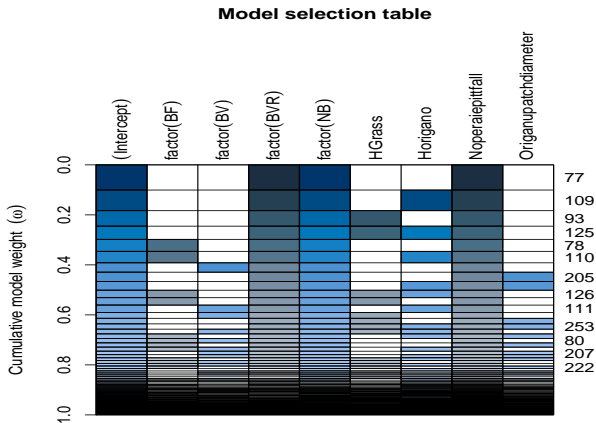
We use a typical fixed-effects exponential resource selection function:

$$\hat{w}(x) = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_n x_n)$$

Parameters are estimated from logistic regression

```
> # names(dati.glm)
> mod.glm <- glm(Egg10 ~ Noperaiepittfall +
+               Horigano + HGrass +
+               Origanupatchdiameter +
+               factor(BV) + factor(BVR) +
+               factor(BF) + factor(NB),
+               data=dati.glm,
+               family=binomial)
> summary(mod.glm)
>
```

Model selection: model averaging



Introducing the Specimen - exercise 3

```
> dati.GLMM <- read.csv("dati/dati.GLMM.csv")
```

```
> str(dati.GLMM)
```

```
'data.frame':      214 obs. of  14 variables:
```

```
$ X           : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
$ id          : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
$ Specimen    : int  1 6 4 5 5 6 8 8 8 8 ...
```

```
$ Egg10       : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ MyrmicaYN   : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ Noperaiepittfall : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ Horigano    : num  50 50 30 60 70 37.5 50 45 25 55 ..
```

```
$ HGrass      : int  10 10 10 40 15 15 10 15 15 5 ...
```

```
$ Origanupatchdiameter: int  50 50 3 60 190 50 120 90 15 25 ...
```

```
$ BV          : int  0 0 1 1 1 0 1 1 1 1 ...
```

```
$ BVR         : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ BR          : int  1 1 0 0 0 1 0 0 0 0 ...
```

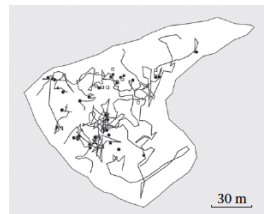
```
$ BF          : int  1 1 0 0 0 1 0 0 0 0 ...
```

```
$ ND          : int  0 0 0 0 0 0 1 0 1 0 ...
```

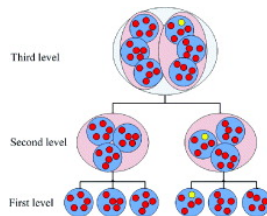

Repeated observations, are they a nuisance?

Repeated observations on the same individuals are often assumed to give rise to constant within group correlation structures.

Even if we assume independence among observations taken on different individuals, there will probably be a (constant?) **correlation between any two observations from the same individual.**



- ▶ constant correlation among observations from the same individual
- ▶ correlation to a lesser degree among observation from animals in the same herd
- ▶ independence among observations from animals in different herds



Correlation vs. autocorrelation

Correlation

It is due to the hierarchical nature of data, and it results in general correlation structures

Autocorrelation

Specifically, the within-animal correlation patterns (eg. correlations that change as a function of temporal and spatial distance between observations)

Correlation structure: assumptions

Animals are randomly sampled and independence is assumed among observations taken on different individuals



Issues

Correlation violates the assumption of independence

- ▶ biased estimates of coefficients
- ▶ deceptively low estimates of uncertainty
- ▶ overfit models
- ▶ spurious conclusion!!

In our previous analysis, the relocation was considered as the experimental unit, but we actually pooled data across animals.

POOLING HAS IMPORTANT CONSEQUENCES FOR INFERENCE

Issues

Pooling observations from several animals is often the **only option** when sample size is limited. But we are violating the assumption of independence. . . (this constitutes pseudo-replication)

Three main avenues:

1. data censoring - collect data a priori (or rarify through subsampling) to satisfy the assumption of independence
2. variance inflation - adopt a post hoc approach to adjust (inflate) standard errors of the parameters
3. explicit modelling of correlation

When sample size is sufficient (for both the no. of animals and no. of relocations per animal), **animals (not relocations) should be considered as the unit of replication if we want to infer average selection patterns for the entire population**

An intuitive approach

First stage

Regression models may be fit to data from individual animals and then averaged to determine population level responses.

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}$$

We obtain 22 estimates of α_i and β_i

Second stage

Sample means and variances are used in the second stage to characterize population means and variances.

Two-stage method: advantages

- ▶ unbiased estimators of $\bar{\beta}$ assuming individuals are independent
- ▶ can determine different covariate structures for each individual (model selection at individual level)
- ▶ model selection can be performed at the population level (makes more sense if we want to characterize typical selection patterns - note: likelihoods are additive given independence of the animals)

Two-stage method: drawbacks

- ▶ sufficient data must exist for fitting individual-level models

```
[1] 22
```

```
> table(dati.GLMM$Specimen)
```

```
 1  2  3  4  5  6  7  8  9 10 12 13 14 15 16 17 18 19 20 21 22 2
```

```
 4  2  5  9  9 12 13 19  3 27 31 10 12  2  5  9  9  2  7 17  4
```

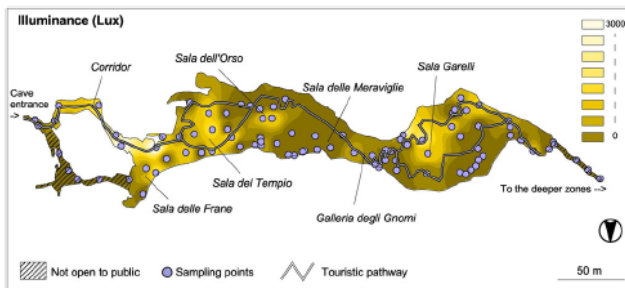
```
>
```

Are we going to drop individuals that are followed for only a short period of time?

Random effects - exercise 4

With the two-stage approach, the animal is treated as a **random effect**.

A variable is considered random when the investigator has not controlled explicitly for levels of the variable in the experimental design, but **has chosen a random sample of levels from the population**

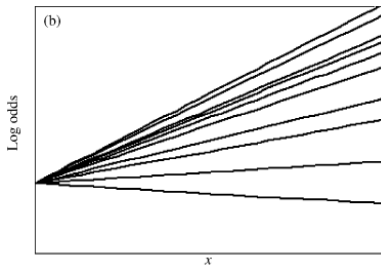


Mixed model formulation

By including a random effect for individuals, individual variability is identified explicitly and the scope of inference can be extended to the entire population

Mixed-effects models offer a powerful approach to modelling correlated data under the assumption that latent or unmeasured characteristics associated with individuals might induce correlation among repeated measurements on these individuals.

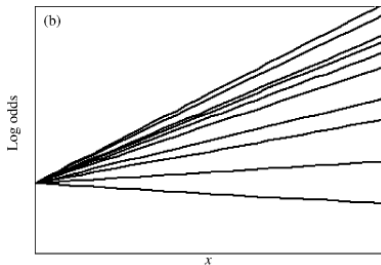
$$Y = X\beta + Zb + \epsilon$$



BUT...what about the number of parameters??

Random slope model: interpretation

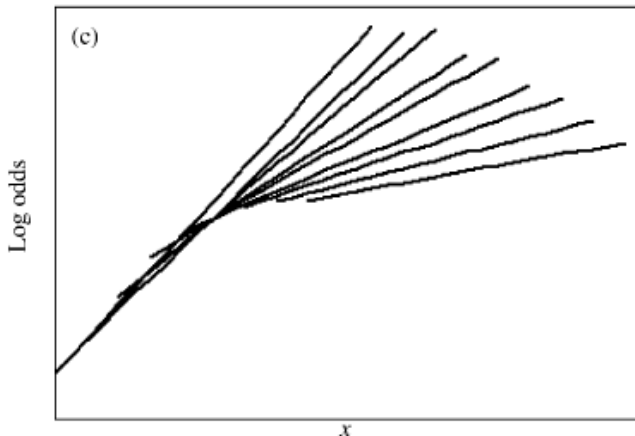
$$g(x) = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_n x_{nij} + \gamma_{nj} x_{nj}$$



If we fit models with random coefficients, we allow for animal-level specific estimates for a response (conditional estimates) and for an overall estimate (marginal or population-level estimator)

Random intercept and slope model: interpretation

$$g(x) = (\beta_0 + \gamma_{0j}) + \beta_1 x_{1ij} + \dots + \beta_n x_{nij} + \gamma_{nj} x_{nj}$$



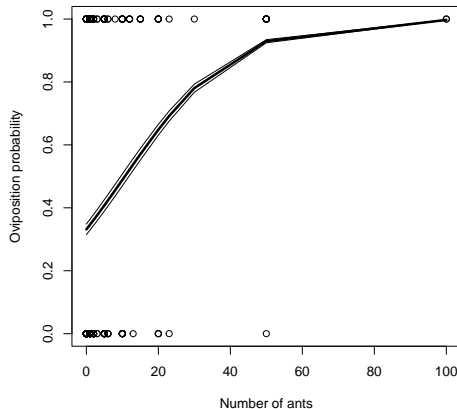
R packages

What methods are available to fit (estimate) GLMMs?

(adapted from Bolker et al TREE 2009)

Penalized quasi-likelihood	Flexible, widely implemented	Likelihood inference may be inappropriate; biased for large variance or small means	PROC GLIMMIX (SAS), GLMM (GenStat), glmmPQL (R:MASS), ASREML-R
Laplace approximation	More accurate than PQL	Slower and less flexible than PQL	glmer (R:lme4, lme4a), glmm.admb (R:glmmADMB), AD Model Builder, HLM
Gauss-Hermite quadrature	More accurate than Laplace	Slower than Laplace; limited to 2-3 random effects	PROC NLMIXED (SAS), glmer (R:lme4, lme4a), glmmML (R:glmmML), xtlogit (Statistica)
Markov chain Monte Carlo	Highly flexible, arbitrary number of random effects; accurate	Very slow, technically challenging, Bayesian framework	MCMCglmm (R:MCMCglmm), MCMCpack (R), WinBUGS/OpenBUGS (R interface: BRugs/R2WinBUGS), JAGS (R interface: rjags/R2jags), AD Model Builder (R interface: glmm.admb ¹ (R:glmmADMB))

Mixed model selection



Any other issue?

$Y =$ fixed part

$$\alpha + \beta_1 X_1 + \dots + \beta_q X_q$$

$$\alpha + f_1(X_1) + \dots + f_q(X_q)$$

$+$ random part

Heterogeneity

Nested data (random effects)

Temporal correlation

Spatial correlation

Random noise

Mixed models for resource selection: advantages

Mixed-effects models are appealing

- ▶ they match the way in which ecological data are structured
- ▶ they can effectively use data from individuals having few data points, eliminating the problem of sampling inequities among individuals
- ▶ they simultaneously fit individual- and population-level models, efficiently pooling information across individuals

Main references

- ▶ Manly et al 2002 Resource selection by animals. Statistical design and analysis for field studies. Kluwer Academic Publisher
- ▶ Millspaugh & Marzluff (eds) 2001 Radio tracking and animal populations. Academic Press.
- ▶ Gillies et al 2006 Application of random effects to the study of resource selection by animals. Journal of Animal Ecology 75:887-898
- ▶ Fieberg et al 2010 Correlation and studies of habitat selection: problem, red herring or opportunity? Philophical Transactions of the Royal Society B 365:2233-2244
- ▶ Burnham & Anderson 2002 Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach. Springer.
- ▶ Zuur et al 2009 Mixed effects models and extensions in ecology with R. Springer

Overview

Resource selection studies

- Motivation

- Definitions

Case study

- Case study description

- Study design

Generalized Linear Model

- Data exploration

- Model fitting and selection

Correlation structure

- Repeated observations

- Implications

Explicit modelling of correlation

- Two-stage methods

- Mixed-effects models