How to Analyze Data From Multiple Animals

John Fieberg, Associate Professor

Department of Fisheries, Wildlife and Conservation Biology

---

1. We've seen how to estimate parameters describing an individual's use of space
   - RSFs
   - SSFs
2. We may want to know what is driving any differences among animals
3. We may want to pool information across individuals to better understand population-level selection patterns

---

Methods for Modeling Data From Multiple Individuals

1. Fit models to pooled data, ignoring the fact that we have repeated measures, but use “robust SEs” (Generalized Estimating Equations or a “cluster-level bootstrap”) for inference
2. Fit models to individual animals and treat the estimates as data (two-step approach)
3. Mixed models (aka hierarchical models, random effect models): allow parameters to vary by animal

---

Individual Variability is Important

1. Fit models to pooled data, ignoring the fact that we have repeated measures, but use "robust SEs" (Generalized Estimating Equations or a "cluster-level bootstrap")
2. Fit models to individual animals and treat the estimates as data (two-step approach)
3. Mixed models, hierarchical models, random effect models: allow parameters to vary by animal

Causes of Individual Variability

Behavioral Ecology

Volume 26, Issue 2
March-April 2015

Sex-specific adjustments in habitat selection contribute to buffer mouflon against summer conditions

Pascal Marchand, Mathieu Garel, Gilles Bourgoi
Anne Loison Author Notes

Behavioral Ecology, Volume 2
https://doi.org/10.1093/behe

Received: 22 August 2016 | Revised: 12 December 2016 | Accepted: 22 December 2016
DOI: 10.1002/ece3.2764

WILEY Ecology and Evolution

Vol. 137, Supplement The American Naturalist June 1991

OIKOS 92: 542–554. Copenhagen 2001

Habitat selection by predators
asymmetrical intraguild predati

Michael R. Heithaus

Heithaus, M. R
asymmetrical ir

GENETIC VARIATION FOR HABITAT PREFERENCE: EVIDENCE AND EXPLANATIONS

John Jaenike and Robert D. Holt

Department of Biology, University of Rochester, Rochester, New York 14627;
Museum of Natural History and Department of Systematics and Ecology,
University of Kansas, Lawrence, Kansas 66045

Abstract.—Because adaptive shifts may often be initiated by evolutionary changes in behavior, it is of interest to determine the extent to which natural populations harbor genetic variation for ecologically important behaviors. Habitat preference is an especially significant behavior, because it determines the regime of natural selection acting on loci that affect adaptation to the environment. A survey of the literature reveals that genetic variation for habitat selection is common, especially in arthropods and mollusks, the groups that have been studied most frequently. Possible adaptive mechanisms by which this variation could be maintained within populations include a genetic correlation between density-independent fitness in a habitat and a preference for it; and soft selection, whereby density-dependent population regulation occurs independently in separate habitats. Several studies have documented a phenotypic correlation

Why do we care about individual variability?

Both Environment and Genetic Makeup Influence Behavior

By: Michael D. Breed (*Department of Ecology & Evolutionary Biology, University of Colorado at Boulder*) & Leticia Sanche
Ecology & Evolutionary Biology, University of Colorado at Boulder) © 2010 Nature Education

Citation: Breed, M. & Sanchez, L. (2010) Both Environment and Genetic Makeup Influence Behavior. *Nature Education Knowledge* 3(10):68

How do genes and the environment come together to shape animal behavior? E
important roles. Genes capture the evolutionary responses of prior populations
behavior. Environmental flexibility
their own lifetime.

50 YEARS WITH IMPACT Journal of Applied Ecology

Journal of Applied Ecology 2015, **52**, 496–504 doi: 10.1111/1365-2664.12400

Adjustments in habitat selection to changing availability induce fitness costs for a threatened ungulate

Chrystel L. Losier[1,2], Serge Couturier[1], Martin-Hugues St-Laurent[2,3,4], Pierre Drapeau[2,5],
niel Fortin[1,2*]

ECOLOG
Ecology Letters, (2018) 21: 665-6

LETTER Intragroup competi
specialisation in a

Catherine E. Sheppard,[1] Richard Inger,[2] Robbie A. McDonald,[2] Sam Barker,[2] Andrew L. Jackson,[3] Faye J. Thompson,[1] Emma I. K. Vitikainen,[1,4] Michael A. Cant[1*] and Harry H. Marshall[1,5*]

Abstract

to niche partitioning and greater individual foraging specialisation ('niche partitioning hypothesi
We tested these opposing predictions in wild, group-living banded mongooses (*Mungos mung*
using stable isotope analysis of banded mongoose whiskers to quantify individual and group f
aging niche. Individual foraging niche size declined with increasing group size, despite all grou
having a similar overall niche size. Our findings support the prediction that competition promo
niche partitioning within social groups and suggest that individual foraging specialisation m
play an important role in the formation of stable social groupings.

Keywords

Banded mongoose, competition, foraging behaviour, foraging niche, group-living, *Mungos mung*
social group, specialisation, stable isotope.

Ecology Letters (2018) **21**: 665–673

Oecologia (2017) 185:415–425
DOI 10.1007/s00442-017-3939-8

CrossMark

BEHAVIORAL ECOLOGY –ORIGINAL RESEARCH

Not accounting for interindividual variability can mask habitat selection patterns: a case study on black bears

Rémi Lesmerises[1] · Martin-Hugues St-Laurent[2]

PNAS

M. Tim Tinker*†, Gena Bentall‡, and James A. Estes§

Edited by Robert T. Paine, University of Washington, Seattle, WA, and approved November 28, 2007 (received for review September 28, 2007)

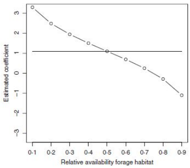
Dietary diversity often varies inversely with prey resource abundance. This pattern, although typically measured at the population level, is usually assumed to also characterize the behavior of individual animals within the population. However, the pattern might also be produced by changes in the degree of variation among individuals. Here we report on dietary and associated behavioral changes that occurred with the experimental translo-

their prey, and we have obtained longitudinal records of sea otter diet and foraging behavior from tagged individuals that span multiple years (8). Moreover, the experimental translocation of sea otters from central California (henceforth CC) to San Nicolas Island (henceforth SN) in the southern California Bight in 1987–1990 established a second population in a comparatively food-rich environment where the diversity of potential inverte-

Functional Response in Habitat Selection

Assume animals needed a constant amount of a particular resource (e.g., water).

What would you expect to see if you plotted animal-specific RSF parameters against availability of that resource?

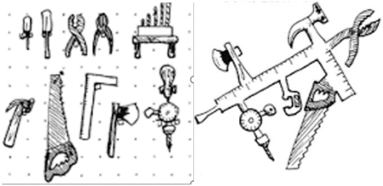


Aarts, G., Fieberg, J., Brasseur, S., & Matthiopoulos, J. (2013). Quantifying the effect of habitat availability on species distributions. Journal of animal ecology, 82(6), 1135-1145.

Why 2-Step Methods?

1. Fit models to pooled data, ignoring the fact that we have repeated measures, but use "robust SEs" (Generalized Estimating Equations or a "cluster-level bootstrap")
2. Fit models to individual animals and treat the estimates as data (two-step approach)
3. Mixed models, hierarchical models, random effect models: allow parameters to vary by animal

Two-step versus Mixed Effects Models



"If you can't explain it simply, you don't understand it well enough" - Albert Einstein

Two-step approach

Independence?

What if we treat all data as independent? What problems may we encounter?

Thought Exercise

Imagine trying to quantify the relative amount of time Americans and Europeans spend watching football.

Follow individuals for between 35 and 365 days. Record $y_i = 1$ if watched that day (0 otherwise).

Is the variable sample size problematic? If so, when?

Pool data assuming independence

Conventional wisdom:

- Non-independence may not bias parameter estimators, but. . .
- Estimates of uncertainty will be too small
- In reality, we need data to be 'missing completely at random' (MCAR)

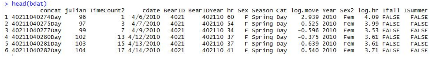
MCAR: n_i (sample size for each individual) does not depend on the response of interest. . . . a problem if those that like football tend to contribute more data!

Non-independence

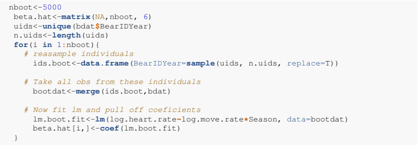
What about measures of uncertainty when assuming independence?

- Can use cluster-level bootstrap (resample individuals)
- Generalized estimating equations (robust, sandwich standard errors)

Code for cluster-level bootstrap



```
> head(bdat)
          concat julian TimeCount2     cdate BearID BearIDYear hr Sex Season Cat log.move Year Sex2 log.hr Ifall ISummer
1 40211040274Day     96          1  4/6/2010   4021     402110 60   F Spring Day    2.939 2010  Fem   4.09 FALSE   FALSE
2 40211040275Day     97          3  4/7/2010   4021     402110 54   F Spring Day    0.525 2010  Fem   3.99 FALSE   FALSE
3 40211040277Day     99          7  4/9/2010   4021     402110 34   F Spring Day   -0.596 2010  Fem   3.53 FALSE   FALSE
4 40211040280Day    102         13 4/12/2010   4021     402110 37   F Spring Day   -0.375 2010  Fem   3.61 FALSE   FALSE
5 40211040281Day    103         15 4/13/2010   4021     402110 37   F Spring Day   -0.639 2010  Fem   3.61 FALSE   FALSE
6 40211040282Day    104         17 4/14/2010   4021     402110 41   F Spring Day    0.540 2010  Fem   3.71 FALSE   FALSE
```



```
nboot<-5000
 beta.hat<-matrix(NA,nboot, 6)
 uids<-unique(bdat$BearIDYear)
 n.uids<-length(uids)
 for(i in 1:nboot){
   # reasample individuals
     ids.boot<-data.frame(BearIDYear=sample(uids, n.uids, replace=T))

   # Take all obs from these individuals
     bootdat<-merge(ids.boot,bdat)

   # Now fit lm and pull off coeficients
     lm.boot.fit<-lm(log.heart.rate~log.move.rate*Season, data=bootdat)
     beta.hat[i,]<-coef(lm.boot.fit)
 }
```

Generalized estimating equations

What are Generalized Estimating Equations (GEE)?

- Natural extension of generalized linear models to correlated data.

Generalized linear models

Assume Y comes from a distribution in the exponential family

- Gaussian
- Poisson (count data)
- Bernoulli (binary data)

Linear model applies to some transformation of the mean:

- $\eta(\mu) = \beta_0 + x_1\beta_1 + \ldots x_p\beta_p$
- Poisson $log(\mu)$
- Bernoulli: $logit(\mu) = log(\mu/(1-\mu))$

Generalized Linear Models

GLM: $\hat{\beta}_{MLE}$ solves: $\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} V_i^{-1}(Y_i - \mu_i) = 0$.

- $\mu_i = E[Y_i|X_i] = f(X_i, \beta)$
- $V_i = Var[Y_i|X_i] = g(\mu_i)$
- $\frac{\partial \mu_i}{\partial \beta}$ is a 1xp vector of derivatives of μ_i with respect to β [$p = dim(\beta)$].

Logistic regression:

- $\mu_i = E[Y_i|X_i] = \exp(X_i\beta)/[1 + \exp(X_i\beta)]$
- $V_i = Var[Y_i|X_i] = \mu_i(1 - \mu_i)$.

Generalized Estimating Equations (GEE)

GEE: $\hat{\beta}$ solves: $\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} V_i^{-1}(\alpha)(Y_i - \mu_i) = 0$.

- $Y_i = (Y_{i1}, Y_{i2}, \ldots Y_{im_i})$, a vector of responses for individual i
- $\frac{\partial \mu_i}{\partial \beta}$ is a $m_i xp$ matrix of first derivatives
- $V_i(\alpha) = A_i^{1/2} R_i(\alpha) A_i^{1/2}$ is the variance-covariance matrix for individual i
- A_i = variance model, typically based on exponential family (e.g., $\mu_i(1 - \mu_i)$ for binary data).
- $R_i(\alpha)$ = working correlation model that describes within subject correlation.

Fitting GEEs in R

library geepack:

geeglm(y $\sim$ x, family= , corstr =) Options:

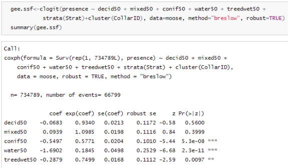
- family: poisson(), binomial(), gaussian(), Gamma(), quasipoisson(), quasibinomial()
- crostr: independence, exchangeable, ar1

<u>If data are MCAR:</u>

$\hat{\beta}$ will be asymptotically unbiased (think large no. of clusters) even when the correlation structure (and model of the variance) is mis-specified.

Works best with lots of similarly sized clusters.

Generalized Estimating Equations



```
gee.ssf<-clogit(presence ~ decid50 + mixed50 + conif50 + water50 + treedwet50 +
          strata(Strat)+cluster(CollarID), data=moose, method="breslow", robust=TRUE)
summary(gee.ssf)
```

```
Call:
coxph(formula = Surv(rep(1, 734789L), presence) ~ decid50 + mixed50 +
    conif50 + water50 + treedwet50 + strata(Strat) + cluster(CollarID),
    data = moose, robust = TRUE, method = "breslow")

  n= 734789, number of events= 66799

              coef exp(coef) se(coef) robust se     z Pr(>|z|)
decid50    -0.0683    0.9340   0.0213    0.1172 -0.58   0.5600
mixed50     0.0939    1.0985   0.0198    0.1116  0.84   0.3999
conif50    -0.5497    0.5771   0.0204    0.1010 -5.44  5.3e-08 ***
water50    -1.6902    0.1845   0.0498    0.2529 -6.68  2.3e-11 ***
treedwet50 -0.2879    0.7499   0.0168    0.1112 -2.59   0.0097 **
```

Notes: `cluster(CollarID), method="breslow", robust=TRUE` (and much larger SEs!)

Two-Step Approach

Step 1: fit models to individuals

$$f_i^u(s) \propto \exp(elev(s)\beta_i + popD(s)\gamma_i + forest(s)\tau_i)$$

Step 2: Do statistics on $(\hat{\beta}_i, \hat{\gamma}_i, \hat{\tau}_i)$

- calculate their variance/covariance (biased high due to sampling variability)
- relate coefficients to animal-specific characteristics (e.g., age, sex) using say `lm`
- plot coefficients against availability to explore functional responses

---

Fit models to individual animals

- Often a useful starting point (exploratory data analysis)
- Fewer parametric assumptions (no distributional assumptions about random effects)

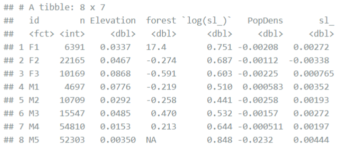
For justification of 2-step approach, see: Murtaugh, P. A. (2007). Simplicity and complexity in ecological data analysis. Ecology, 88(1), 56-62.

In the context of step-selection functions:

- Craiu, R. V., T. Duchesne, D. Fortin, and S. Baillargeon (2011). Conditional logistic regression with longitudinal follow-up and individual-level random coefficients: A stable and efficient two-step estimation method. Journal of Computational and Graphical Statistics 20, 767-784.
- Craiu, R. V., T. Duchesne, D. Fortin, and S. Baillargeon (2016). TwoStepCLogit: Conditional Logistic Regression: A Two-Step Estimation Method. R package version 1.2.5.

Two-step Approach

- Quick and easy using the `amt` package in conjunction with `tidyverse` in R
- See FisherRSF.R and FisherSSF.R (in Rscripts and Output folders) for examples.



```
## # A tibble: 8 x 7
##   id        n Elevation  forest `log(sl_)`   PopDens       sl_
##   <fct> <int>     <dbl>   <dbl>      <dbl>     <dbl>     <dbl>
## 1 F1     6391   0.0337   17.4        0.751 -0.00208   0.00272
## 2 F2    22165   0.0467   -0.274      0.687 -0.00112  -0.00338
## 3 F3    10169   0.0868   -0.591      0.603 -0.00225   0.000765
## 4 M1     4697   0.0776   -0.219      0.510  0.000583  0.00352
## 5 M2    10709   0.0292   -0.258      0.441 -0.00258   0.00193
## 6 M3    15547   0.0485    0.470      0.532 -0.00157   0.00272
## 7 M4    54810   0.0153    0.213      0.644 -0.000511  0.00197
## 8 M5    52303   0.00350  NA          0.848 -0.0232    0.00444
```

Mixed models

$$f_i^u(s) \propto \exp(elev(s)\beta_i + popD(s)\gamma_i + forest(s)\tau_i)$$

Further assume:

$$(\beta_i, \gamma_i, \tau_i) \sim N(\mu, \psi)$$

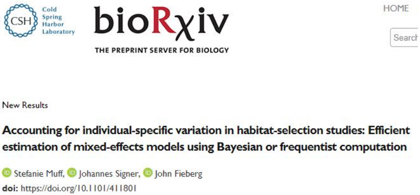
Similar 2-step approach, but assume the regression parameters come from a common normal distribution.

Advantages:

- inference at individual- and population-level with single model
- can "borrow strength" across individuals when estimating $(\beta_{1i}, \ldots, \beta_{2i})$

But. . . more assumptions, added complexity

Fitting Mixed RSFs and SSFs



CSH Cold Spring Harbor Laboratory

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

HOME

Search

New Results

Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation

Stefanie Muff, Johannes Signer, John Fieberg

doi: https://doi.org/10.1101/411801

One Step: Random Effects

- Random effects were proposed for RSFs over 10 years ago[1]

About 35,200 results (**0.11** sec)

[HTML] Application of **random effects** to the study of resource selection by animals

CS **Gillies**, M Hebblewhite, SE Nielsen... - Journal of Animal ..., 2006 - Wiley Online Library

2Most frequently, individual animals are monitored and pooled to estimate population-level **effects** without regard to group or individual-level variation. Pooling assumes that both observations and their errors are independent, and resource selection is constant given individual variation ...

- Majority of studies (80 % since 2016) only include random intercept and no random slope(s).

---

[1] Gillies et al. "Application of random effects to the study of resource selection by animals." Journal of Animal Ecology 75.4 (2006): 887-898.

RSFs: Random Intercept-Only Models

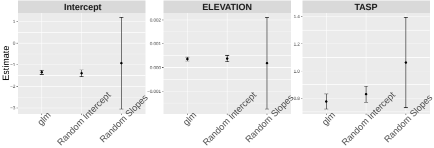
1. Intercept in RSFs is not of interest and depends heavily on the sampling ratio of used versus available points
2. Cannot (by definition) account for among-animal variation in the regression slopes (i.e., functional responses)!
3. SEs will be too small, particularly with lots of observations for each animal[5]

---

[5] Schielzeth, H. and W. Forstmeier (2009). Conclusions beyond support: Overconfident estimates in mixed models. Behavioral Ecology 20, 416-420.

Example: Goat RSFs[6]

```
glm <- glm(STATUS ~ TASP + ELEVATION, family=binomial(),
        data = goats)
glmer_int <- glmmTMB(STATUS ~ TASP + ELEVATION + (1|ID),
                family=binomial(),data = goats)
glmer_randcoef <- glmmTMB(STATUS ~ TASP + ELEVATION +
                    (1+TASP+ELEVATION|ID),
                    family=binomial(), data = goats)
```



---

[6] Lele & Keim, (2006) Weighted distributions and estimation of resource selection probability functions. Ecology 87, 3021–3028.

SSFs: Mixed Effects

Conditional logistic regression with random effects is computationally prohibitive for most data sets:

$$\mathrm{P}(y_{ntj} = 1 \mid \boldsymbol{x}_{ntj}) = \pi_{ntj} = \frac{\exp(\beta^\top \boldsymbol{x}_{ntj} + \boldsymbol{u}_n^\top \boldsymbol{z}_{ntj})}{\sum_i \exp(\beta^\top \boldsymbol{x}_{nti} + \boldsymbol{u}_n^\top \boldsymbol{z}_{nti})}\,, \qquad y_{ntj}|\boldsymbol{u}_n \sim \mathrm{B}(\pi_{ntj})$$

- Must integrate, numerically, over the distribution of $\boldsymbol{u}_n$ (no closed-formed solution)
- `coxme` for small numbers of strata
- Alternatively, can use `TwoStepCLogit::Ts.estim()`, a two-step approach

Mixed SSF Trick

Reformulation SSFs as a Poisson model with stratum-specific intercepts α_{nt}[5]

$$\mathrm{E}(y_{nti}) = \mu_{nti} = \exp(\mu + \alpha_{nt} + \boldsymbol{\beta}^\top \boldsymbol{x}_{nti} + \boldsymbol{u}^\top \boldsymbol{z}_{nti})\,, \qquad y_{nti}|\boldsymbol{u}_n \sim \mathrm{Po}(\mu_{nti})$$

- Same likelihood kernel as condition logistic regression likelihood, same $\hat{\beta}$, same $\mathrm{SE}(\hat{\beta})$
- But, then lots of intercepts to estimate
- Trick: $\alpha_{nt} \sim N(0, 10^6)$ (avoids shrinkage and explicit estimation)

---

[5] Armstrong et al. "Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis." BMC medical research methodology 14.1 (2014): 122.

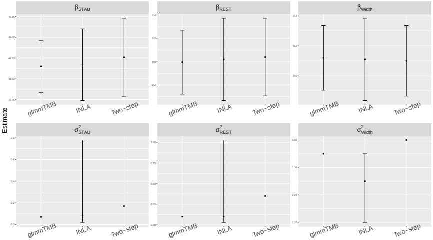
SSF: Otter Example

- 9 otter
- 4167 used locations
- 41670 total locations
- predictors: habitat type (REST1 if rest-water of a reservoir, STAU1 if in a reservoir), river width

Slope estimates	β_{STAU}	β_{REST}	β_{Width}
Fixed effects models			
clogit	-0.07 (0.07)	-0.38 (0.10)	0.16 (0.04)
cPois (INLA)	-0.07 (0.07)	-0.38 (0.10)	0.16 (0.04)
cPois (glmmTMB)	-0.07 (0.07)	-0.38 (0.10)	0.16 (0.04)

Weinberger, I. C., S. Muff, A. Kranz, and F. Bontadina (2016). Flexible habitat selection paves the way for a recovery of otter populations in the European Alps. Biological Conservation 199, 88–95.

Mixed Effects



Speed:
- TS.estim (1.75 seconds)
- `glmmTMB` (5 seconds)
- INLA (90 seconds)
- `coxme` (right-censor 24 hrs)

Summary: Methods for Inference

1. Fit models to pooled data, ignoring the fact that we have repeated measures
   - Use a cluster-level bootstrap or GEEs for inference
   - Can be reasonable if n_i is similar among animals and you are interested in population-level habitat selection patterns
2. Fit models to individual animals and treat the estimates as data (two-step approach)
   - Provides a simple way to explore among-animal variability
3. Mixed models, hierarchical models, random effect models: allow parameters to vary by animal
   - Similar to [2], but accomplish in 1-step

Practicals

Will see how to fit mixed RSFs and mixed SSFs using glmmTMB and INLA

- fisher data
- your own data?