Advanced occupancy models

Diana Bowler

German Centre for Integrative Biodiversity Research

Gfoe workshop

29th August 2021

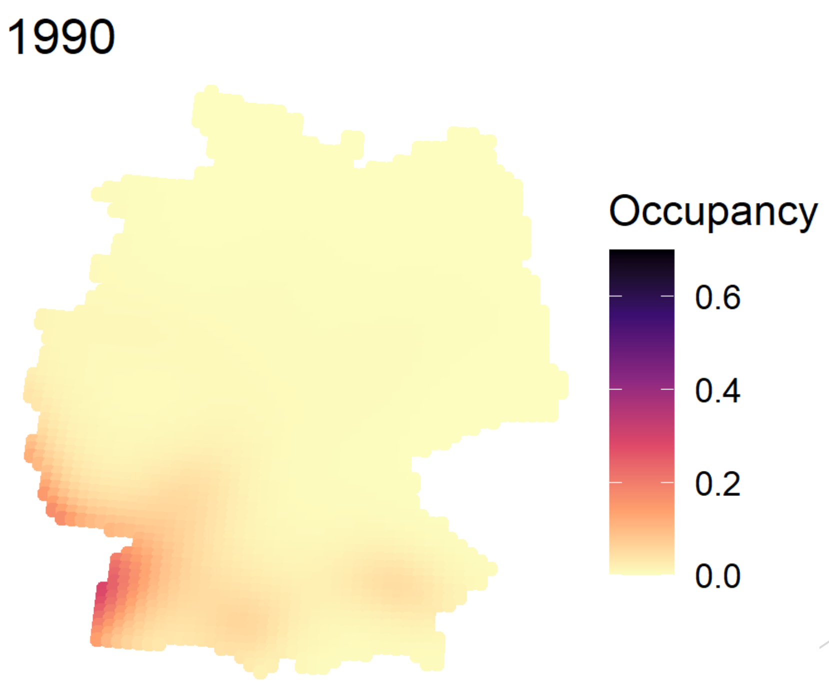
Outline

- Dynamic occupancy models
- Community occupancy models
  - Estimating species richness
- Dealing with false positives

Dynamic occupancy model

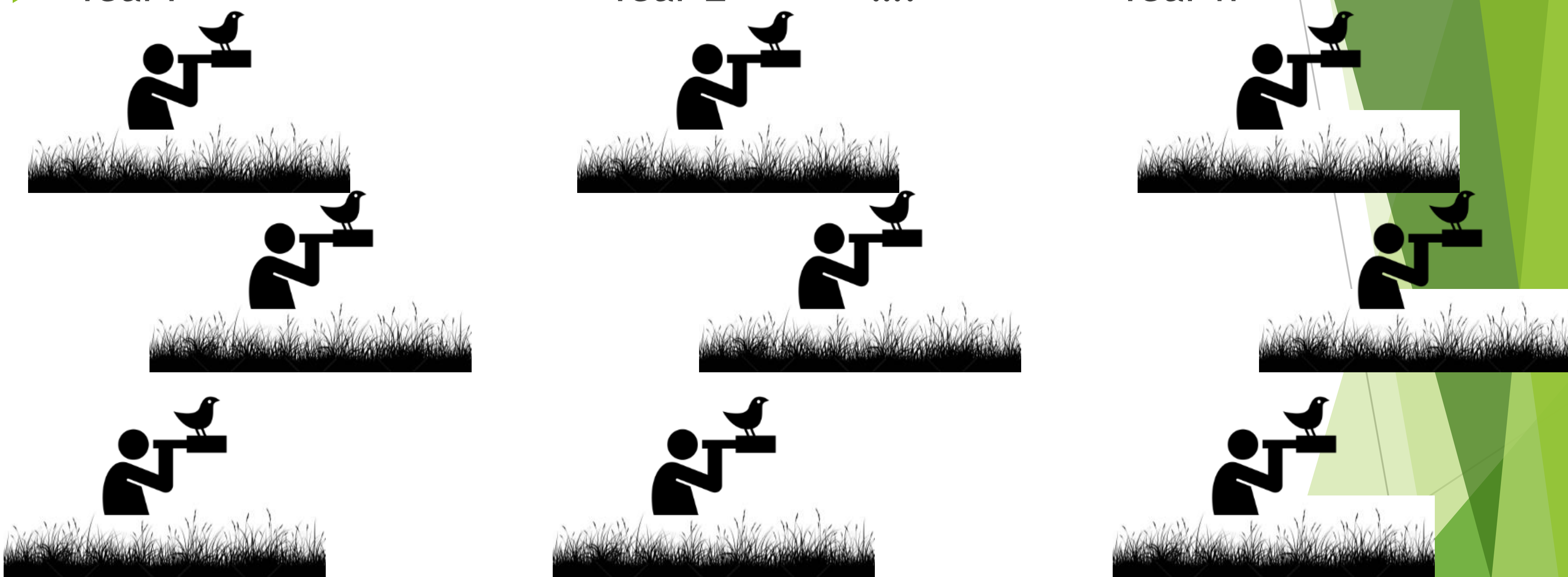
- Multi-season model
- Changes in occupancy among seasons (usually years)

- Temporal data!!



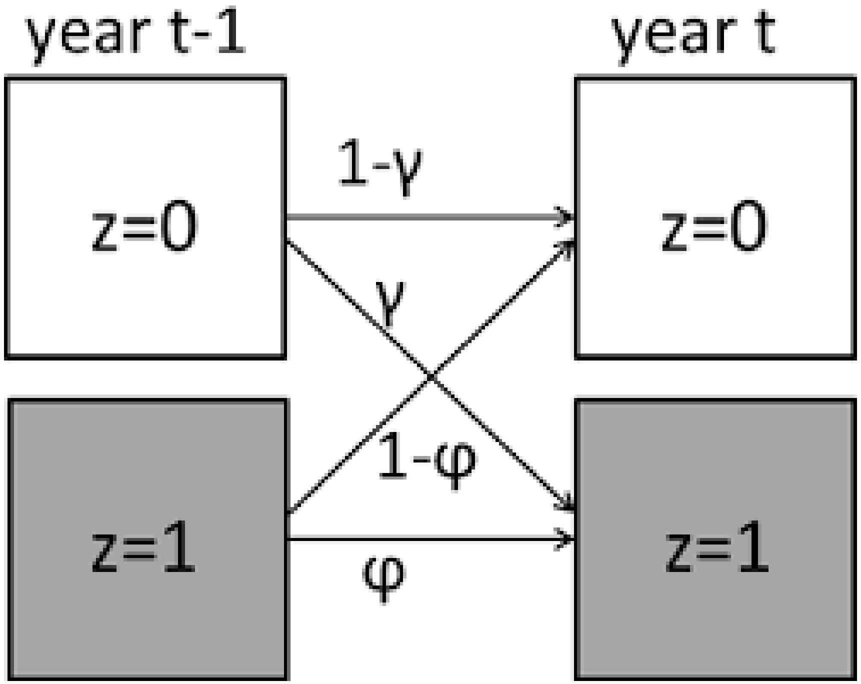
Study design

- Year1 Year 2 .... Year n



In each year, we need data from multiple sites – and some of these sites must have been visited at least twice in each year

Dynamic occupancy model



φ (phi) = persistence probability
γ (gamma) = colonization probability

1- φ = extinction probability

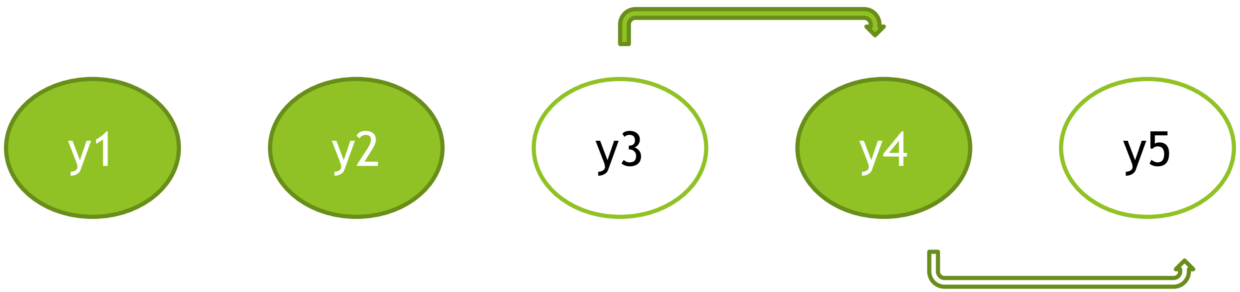
$$\psi_{it} = \psi_{i,t-1}\varphi_{t-1} + (1 - \psi_{i,t-1})\gamma_{t-1}$$

We can model variation in phi and gamma with covariates

Importance of accounting for imperfect detection

- If we ignore imperfect detection
  - Overestimate the number of colonizations and extinctions

Did a colonization even really happen?



Did an extinction even really happen?

Dynamic occupancy model: an example

Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data

Julie Louvrier, Christophe Duchamp, Valentin Lauret, Eric Marboutin, Sarah Cubaynes, Rémi Choquet, Christian Miquel and Olivier Gimenez

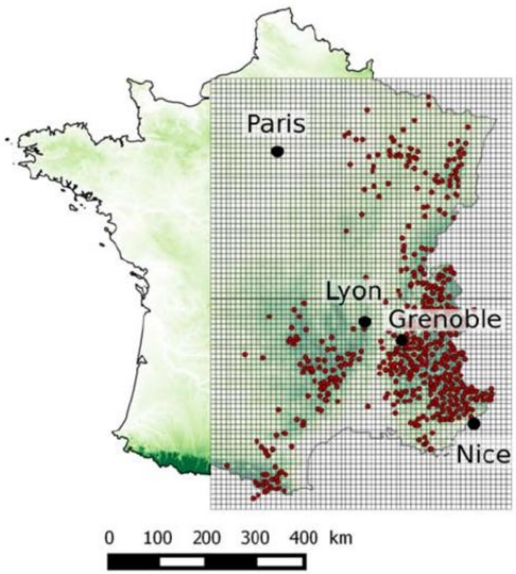


Figure 1. Maps of cumulated species detections (red dots) for the period 1994–2016. Sites were defined as 10 × 10 km cells within a grid covering all detections. Dark green areas represent mountainous areas with an altitude higher than 1500 m.

Ecography
41: 647–660, 2018
doi: 10.1111/ecog.02874

Wolf detection data were made of presence signs sampled all year long from 1992 to 2016 thanks to a network of professional and non-professional observers.

For every presence sign, the date and location of collection were stored in a geo-referenced database.

Covariates

Table 1. Description and expected effects of covariates used to describe the occupancy dynamics of wolf in France.

Covariate	Abbreviation	Parameter	Description	Expected effect	Reference
Forest cover	Forest	Colonisation (γ)	Percentage of mixt, coniferous or deciduous forests cover	+	Oakleaf et al. 2006, Fechter and Storch 2014
Farmland cover	Agr	Colonisation (γ)	Percentage of pasture lands and other farming activities cover	+/–	Glenz et al. 2001
Rock cover	Rock	Colonisation (γ)	Percentage of rock cover	–	
High altitude	Halt	Colonisation (γ)	Proportion of altitude higher than 2500 m	–	Glenz et al. 2001
Altitude	Alt	Colonisation (γ)	Mean altitude	+/–	Llaneza et al. 2012 Falcucci et al. 2013
Distance to the closest barrier	Dbarr	Colonisation (γ)	Minimal distance between a highway or one of the five main rivers in France	–	Falcucci et al. 2013
Short distance occupied neighboring cells	SDAC	Colonisation (γ)	Proportion of observed occupied contiguous cells	+	Bled et al. 2011
Long distance occupied neighboring cells	LDAC	Colonisation (γ)	Proportion of observed occupied cells within a 150 km radius without the contiguous cells	+	
Year (continuous)	Trend-year	Extinction (ε)	Year as a linear effect	–	Marucco 2009
Sampling effort	SEff	Detection (p)	Number of observers per site per year	+	
Road density	Rdens	Detection (p)	Percentage of site covered by roads	+	
Month-survey	survey	Detection (p)	Occasion of survey (categorical)	+/–	Marucco 2009

Dynamic model: example 2

Journal of Applied Ecology

BRITISH ECOLOGICAL SOCIETY

Standard Paper | Free Access

Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models

Arco J. van Strien, Chris A.M. van Swaay, Tim Termaat

- Presence-only (detection) data
- Nondetection data for each species were extracted from the monitoring scheme data and consisted of all visits made without any recorded sightings of the species under consideration,

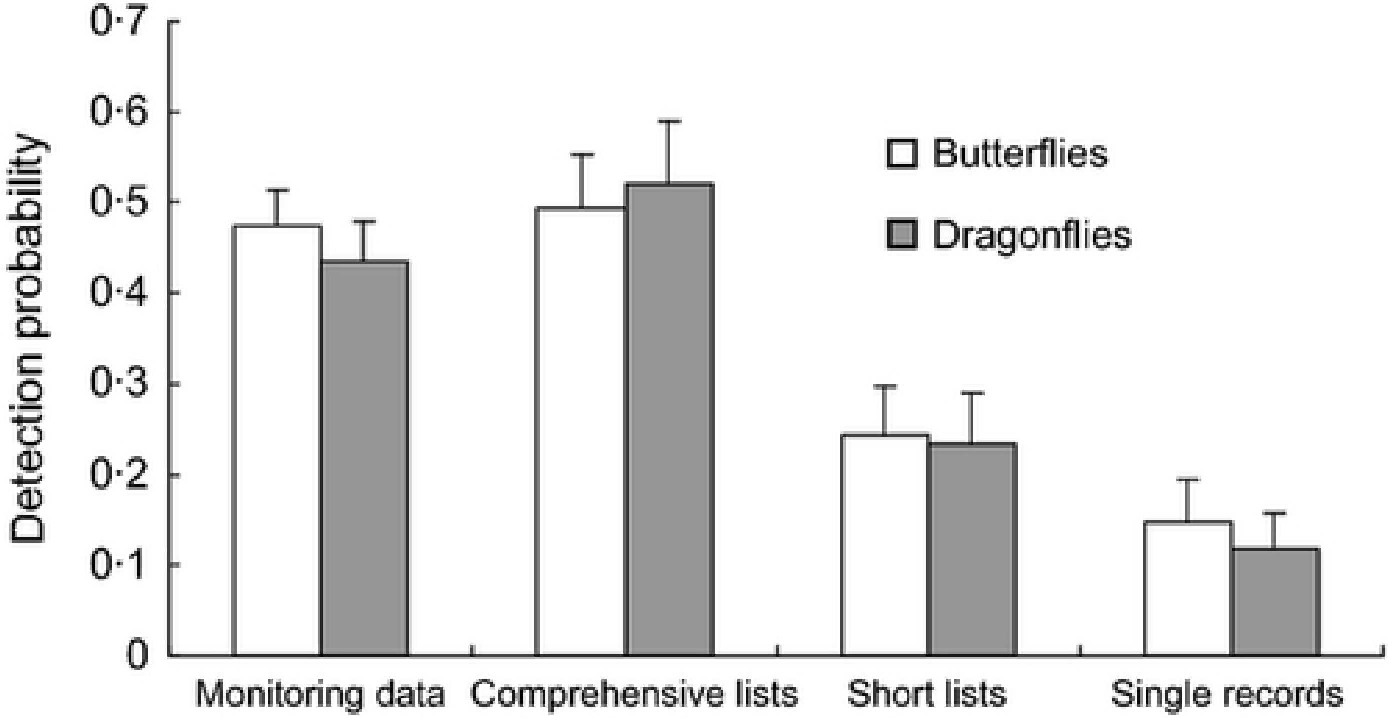
Models

$$\psi_{it} = \psi_{i,t-1}\varphi_{t-1} + (1 - \psi_{i,t-1})\gamma_{t-1}$$

Detection model:

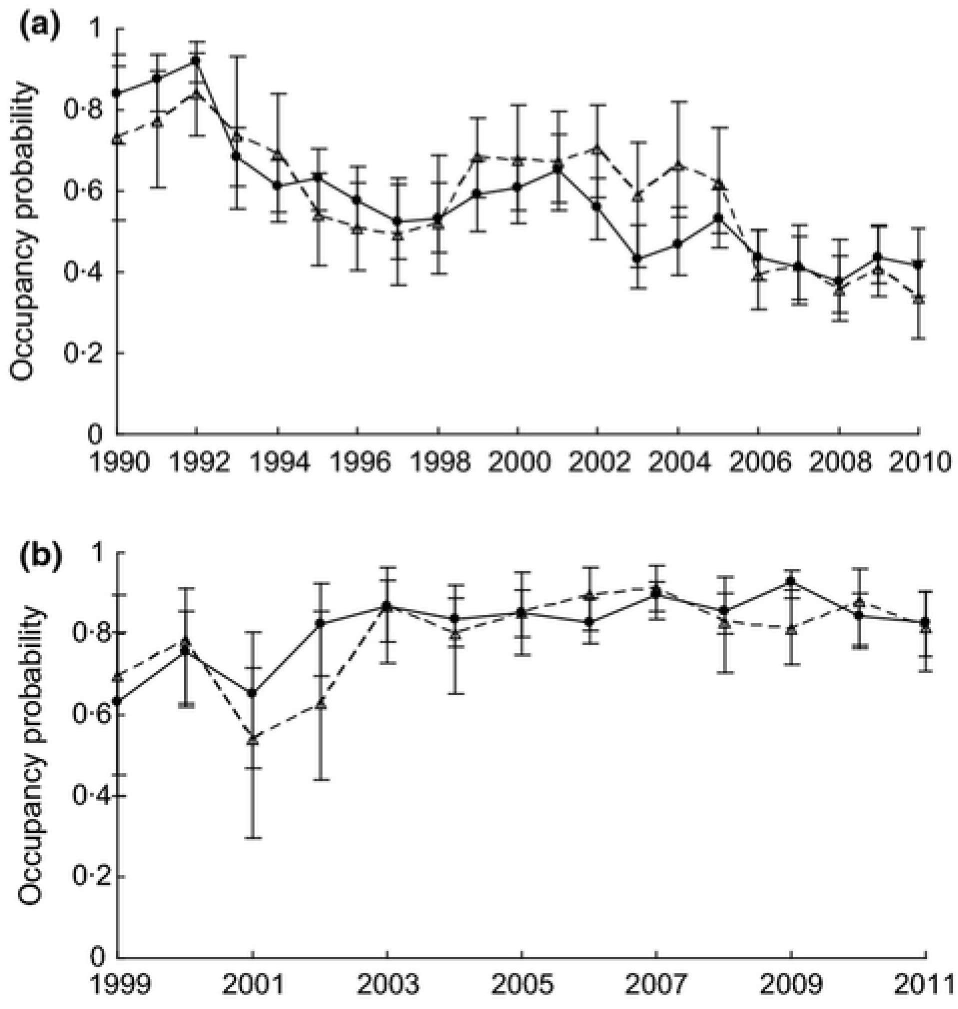
$$\begin{aligned} \text{logit}(p_{ijt}) \; &= \alpha_t + \beta_1 * \text{date}_{ijt} + \beta_2 * \text{date}^2_{ijt} + \delta_1 * (\text{short day-list})_{ijt} + \\ &\delta_2 * (\text{comprehensive day-list})_{ijt} \end{aligned}$$

Results



0·7
0·6
0·5
0·4
0·3
0·2
0·1
0
Detection probability
Butterflies
Dragonflies
Monitoring data
Comprehensive lists
Short lists
Single records

Results



(a)
Occupancy probability
1
0·8
0·6
0·4
0·2
0
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
(b)
Occupancy probability
1
0·8
0·6
0·4
0·2
0
1999
2001
2003
2005
2007
2009
2011

Example 3: Modelling spatio-temporal patterns

Article | Open Access | Published: 05 September 2019

Modeling spatially and temporally complex range dynamics when detection is imperfect

Clark S. Rushing, J. Andrew Royle, David J. Ziolkowski & Keith L. Pardieck

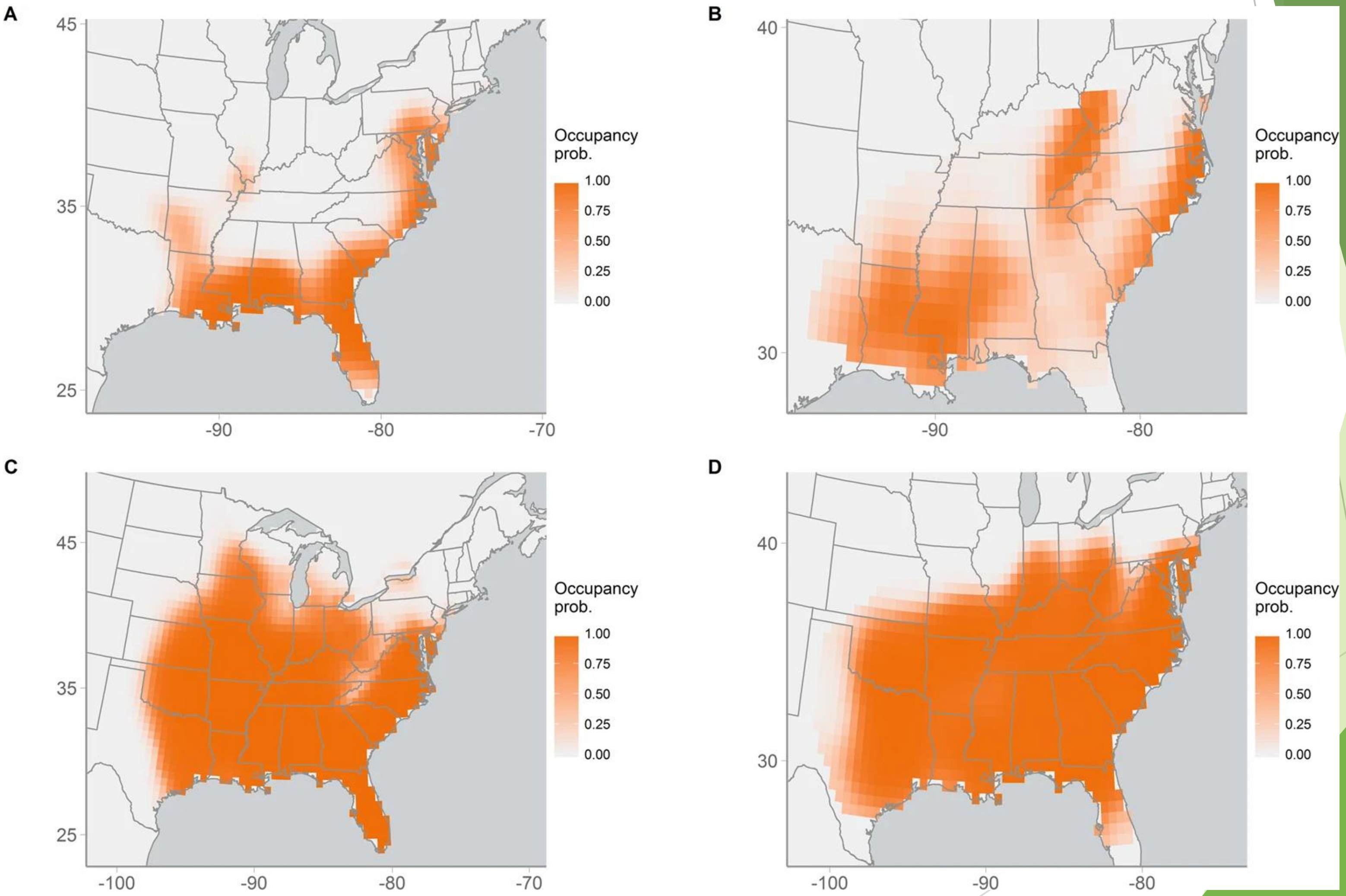
Scientific Reports **9**, Article number: 12805 (2019) | Cite this article

3589 Accesses | **8** Citations | **38** Altmetric | Metrics

- State/ecological model: $logit(\psi_{i,t}) = f_t(lat_i, lon_i) + \omega\beta\mathbf{X}_{i,t}$
- Detection model: $logit(p_{i,t}) = \alpha_0 + \alpha_1 WIND_{i,t} + \alpha_2 \boldsymbol{I}_{i,t} + \alpha_3 \boldsymbol{\kappa}_{i,t} + \eta_{i,t}$

where a_0 is an intercept term, $WIND_{i,t}$ is the wind score, $\boldsymbol{I}_{i,t}$ is a binary dummy variable indicating whether year t was an observer's first year of service, $\boldsymbol{\kappa}_{i,t}$ is a dummy variable indicating the survey protocol used (0 = standard BBS survey, 1 = time-distance protocol), and $\eta_{i,t}$ is a random observer effect

Spatio-temporal predictions



Multi-species occupancy modes

- We have multi-site repeat measurement data available for more than one species

- Option (1) Analyze each species separately and combine results afterwards.
  - Advantage: Easier.
  - Cost: Need to take care to pull uncertainty from species-level models into subsequent models.

- Option(2) Analyze species together in a multi-species model.
  - Advantage: Pool information across species, especially helpful for rarer species. More efficient when number of species is large.
  - Cost: More complex!!

Pooling information across species: species as random effects

$z_{ik} \sim Bernoulli(\psi_{ik})$ # True presence/absence

$\text{logit}(\psi_{ik}) = beta0_k + beta1_k * habitat_i$ # Occupancy probability affected by habitat

$beta0_k \sim Normal(\text{logit}(mean.psi), sig.lpsi^2)$ # Species heterogeneity in the intercept

$beta1_k \sim Normal(mu.beta.lpsi, sig.beta.lpsi^2)$ # Species heterogeneity in the slope

$y_{ijk} \sim Bernoulli(z_{ik} * p_{ijk})$ # Observed detection/nondetection data

$\text{logit}(p_{ijk}) = alpha0_k + alpha1_k * wind_{ij}$ # Detection probability affected by *wind*

$alpha0_k \sim Normal(\text{logit}(mean.p), sig.lp^2)$ # Species heterogeneity in the intercept

$alpha1_k \sim Normal(mu.beta.lp, sig.beta.lp^2)$ # Species heterogeneity in the slope

- Treat species effects as random effects
- Estimate mean species effect (community-level) and inter-specific variation
- Could also include trait effects

Modelling species richness

- Based on species in the observed dataset
  - Extract estimates of z for each species
  - Sum z's across species
  - Based on one species observed at least once
  - Does not account for species that are entirely missed by the survey
    - Underestimate true species richness

Modelling species richness

- Using multi-species occupancy model to predict true species richness
  - Predicts the occurrence of species entirely missed by our survey
  - Based on the common distributions of occurrence and detection that species are assumed to be drawn from
  - Approach called Data Augmentation
  - Alternatives to rarefaction for predicting species richness

1. Superpopulation process : $w_k \sim Bernoulli(\Omega)$
2. State process (occurrence) : $z_{ik} | w_k \sim Bernoulli(w_k \psi_k)$
3. Observation process (detection) : $ysum_{ik} | z_{ik} \sim Binomial(J_i, z_{ik} p_k)$
4. Models of species heterogeneity : $\mathrm{logit}(\psi_k) \sim Normal\left(\mu_{lpsi}, \sigma^2_{lpsi}\right)$

$$\mathrm{logit}(p_k) \sim Normal\left(\mu_{lp}, \sigma^2_{lp}\right)$$

Methods in Ecology and Evolution
BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Multi-species occupancy models as robust estimators of community richness

Morgan W. Tingley, Christopher P. Nadeau, Manette E. Sandor

First published: 13 February 2020 | https://doi.org/10.1111/2041-210X.13378 | Citations: 1

Ecology and Evolution
Open Access

Ecol Evol. 2019 Jan; 9(2): 780–792.
Published online 2019 Feb 5. doi: 10.1002/ece3.4821
PMCID: PMC6362448
PMID: 30766668

Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation

Gurutzeta Guillera-Arroita,[1] Marc Kéry,[2] and José J. Lahoz-Monfort[1]

Author information Article notes Copyright and License information Disclaimer

ECOGRAPHY
A JOURNAL OF S
AND TIME IN ECC

Research | Open Access

Multi-species occupancy models: review, roadmap, and recommendations

Kadambari Devarajan, Toni Lyn Morelli, Simone Tenan

First published: 25 February 2020 | https://doi.org/10.1111/ecog.04957 | Citations: 2

Dealing with false-positive error

- All the occupancy models so far have assumed no false presences
- We may have false-positive errors, i.e., we mistake one species for another. For occurrence it means that we think we detected a species at a site where either it does not occur at all.
- Methods that accommodate false-positive measurement errors in ecological models for abundance, occurrence are still in their infancy
- Various methods have been proposed – usually requiring additional information on the probability to misidentify a species

Occupancy modelling accounting for false positive: one possible approach

Group observations into:

- Certain detections (assume no probability of incorrect identification)
- Uncertain observations (allow some probability of incorrect identification)

Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification

David A. Miller,[1,4] James D. Nichols,[1] Brett T. McClintock,[2] Evan H. Campbell Grant,[1] Larissa L. Bailey,[3] and Linda A. Weir[1]

Ecology, 92(7), 2011, pp. 1422–1428

Table 1. Parameterization for the expected probability of recording the observation state y given the true state of the site z ($P[y\,|\,z]$) for the multiple detection state model with only certain and uncertain detections.

True state	$P(y=0\,|\,z)$	$P(y=1\,|\,z)$	$P(y=2\,|\,z)$
$z = 0$; unoccupied	$1 - p_{10}$	p_{10}	0
$z = 1$; occupied	$1 - p_{11}$	$(1 - b) \times p_{11}$	$b \times p_{11}$

Notes: Possible observations were not detected (0), had uncertain detection (1), or had certain detection (2). Definitions: p_{10}, the probability of (incorrectly) detecting the species at a site given the site is unoccupied; p_{11}, the probability of detecting the species at a site given the site is occupied; and b, the probability that a detection is classified as certain given that the site is occupied and the species was detected.

Methods

- Observations of the ith site on the tth visit, y it, are classified into one of L observation states that differ in the probability of being a false positive detection

- sign to be uncertain (e.g., scat or tracks) and a direct observation to be certain (e.g., visual encounter)

- This kind of model is available

in unmarked

occFP()

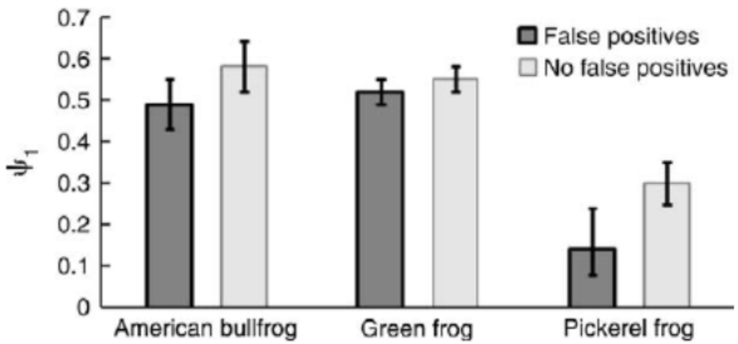


FIG. 3. Estimates of occupancy (proportion of sites in which the species occurs) for three species of frogs at sites in and around the Maryland side of the Chesapeake and Ohio Canal National Historic Park, USA. These estimates were systematically lower when the possibility of false positive detections was included in models than when it was not.

Appendix

Fitting dynamics model using unmarked

- Organising the data

```
umf <- unmarkedMultFrame(
      y = y,
      siteCovs=mydata[,2:3],
          yearlySiteCovs=list(year=years),
           obsCovs=list(date=DATE),
      numPrimary=9) # number of years
```

where y is the binary data in a matrix with nrows equal to number of sites and ncols equal to number of years x number of visits within a year

Fitting dynamics model using unmarked

- fm0 <- colext(~1, ~1, ~1, ~1, data=umf)

- We use the function colext to fit the dynamic model using *unmarked*

- There are 4 models to specify!
  1. *Occupancy probability in year 1*
  2. *Colonization probability*
  3. *Extinction probability*
  4. *Detection probability*

Why do we need the year 1 model?

- The dynamic model explains the occupancy patterns in year 2 onwards as a function of the year before
- For year 1, we don’t have any data on the year before, so we cant use the dynamic model

- We need another model to specify the variation among sites in year 1