SEEC STATS TOOLBOX

WANT TO BROADEN YOUR STATS KNOWLEDGE? UNSURE OF WHAT YOU CAN DO WITH YOUR DATA? STILL DEVELOPING YOUR PROPOSAL?

Join us for the monthly SEEC Stats Toolbox seminars where we introduce statistical methods that are useful for ecologists, environmental and conservation scientists.

Our next seminar:

Topic: Dynamic occupancy models

Who: Res Altwegg & David Maphisa

When: Thursday 28 November, 13h00 - 14h00

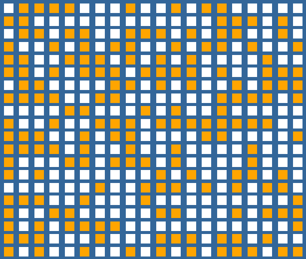
Where: PD Hahn LT4 (Level 6)

SEEC - Statistics in Ecology, Environment and Conservation

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

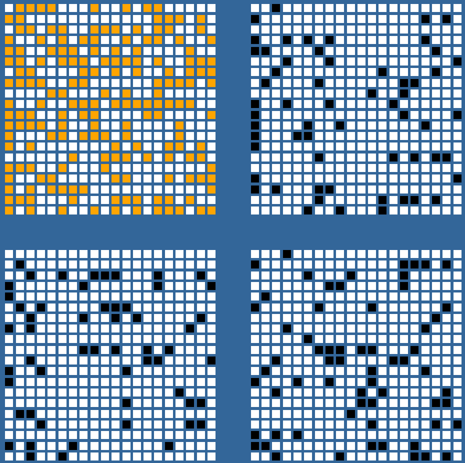
Visit www.seec.uct.ac.za/stats-toolbox-seminars for more information

Occupancy: the proportion of sites occupied by a species



- Occupancy: $\Psi = \frac{occupied}{total}$
- $logit(\Psi) = f(covariates)$

The species is not detected in all occupied cells

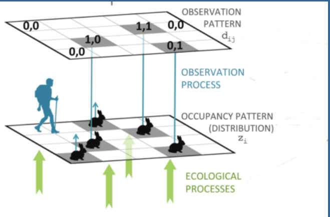


Repeated sampling

Assumptions:

- Closure (no colonisation or exctinction)
- No false detections

A hierarchical model



Guillera-Arroita 2017

Ψ_i = probability that site i is occupied

z_i = true occupancy at site i: 1 = occupied, 0 = not occupied

p_{ij} = prob of detecting the species at site i during survey j

y_{ij} = detection (1) or non-detection (0) at site i during survey j

$$z_i|\Psi_i \sim Bernoulli(\Psi_i)$$

$$y_{i,j}|z_i, p_{i,j} \sim Bernoulli(z_i p_{i,j})$$

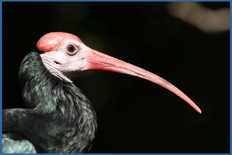
Covariate modelling

Want to know how occupancy and detection vary among sites, i, and survey, j.

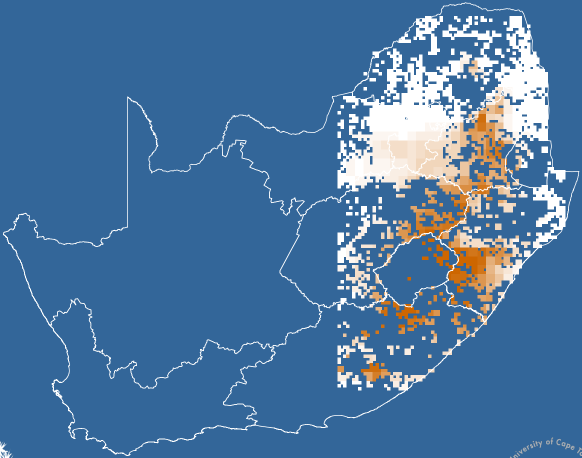
$$logit(\Psi_i) = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_U x_{iU}$$
$$logit(p_{ij}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_U x_{iU} + \beta_{U+1} y_{ij1} + \ldots + \beta_{U+V} y_{ijV}$$

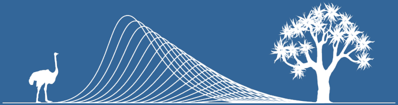
U site-level covariates: $x_{i1}, \ldots, x_{iU}$
V survey-specific covariates: $y_{ij1}, \ldots, y_{ijV}$

Southern bald ibis



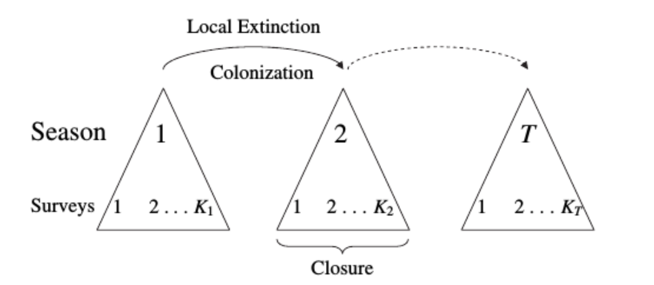
© Peter Ryan





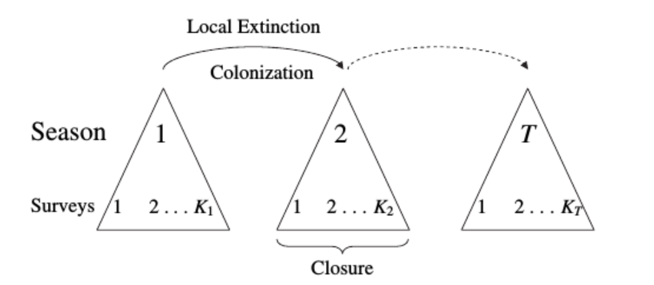
SEEC - Statistics in Ecology, Environment and Conservation

Dynamic occupancy models



$$z_{i,t} | \Psi_{i,t} \sim Bern(\Psi_{i,t})$$
$$y_{i,t,j} | z_{i,t}, p_{i,t,j} \sim Bernoulli(z_i \times p_{i,t,j})$$

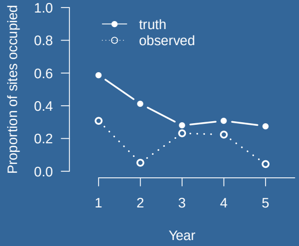
Dynamic occupancy models



$$logit(\Psi_{i,1}) = \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_U x_{iU}$$

$$\Psi_{i,t} | z_{i,t-1} = z_{i,t-1} \times (1 - \epsilon_{i,t-1}) + (1 - z_{i,t-1}) \times \gamma_{i,t-1}, \text{ for } t > 1$$

Preparing the data




```
> head(yy)
     [,1] [,2] [,3] [,4]       [,9] [,10]
[1,]    0    1    1    1  ...     1     0
[2,]    0    0    0    0  ...     0     0
[3,]    0    0    0    0  ...     0     0
[4,]    0    0    0    0  ...     0     0
[5,]    0    0    0    0  ...     0     0
[6,]    0    0    0    0  ...     0     0

> dim(yy)
[1] 250  10
```

- 250 sites
- 5 seasons
- 2 surveys / season
- $\Psi_1 = 0.6$

Preparing the data

```
> library(unmarked)
> simUMF <- unmarkedMultFrame( y = yy,
+              yearlySiteCovs = list(year = year),
+              numPrimary=5)
> summary(simUMF)
unmarkedFrame Object

250 sites
Maximum number of observations per site: 10
Mean number of observations per site: 10
Number of primary survey periods: 5
Number of secondary survey periods: 2
Sites with at least one detection: 136
```

Preparing the data

```
Tabulation of y observations:
   0    1
2242  258

Yearly-site-level covariates:
  year
  01:250
  02:250
  03:250
  04:250
  05:250
```

Fitting the model to the data

```
> m0 <- colext(psiformula= ~1,
+              gammaformula = ~ 1,
+              epsilonformula = ~ 1,
+              pformula = ~ 1,
+              data = simUMF)
```

Looking at the output

```
> summary(m0)
Initial (logit-scale):
  Estimate    SE        z P(>|z|)
  -0.00337 0.236 -0.0143   0.989
Colonization (logit-scale):
  Estimate    SE     z  P(>|z|)
      -1.7 0.195 -8.71 3.01e-18
Extinction (logit-scale):
  Estimate    SE     z P(>|z|)
     0.214 0.318 0.672   0.502
Detection (logit-scale):
  Estimate   SE     z  P(>|z|)
    -0.654 0.17 -3.86 0.000115
```

Parameter estimates

- $logit(\Psi_1) = -0.0034$
- What is Ψ_1? (Should be 0.6?)

Parameter estimates

- $logit(\Psi_1) = -0.0034$
- What is Ψ_1? (Should be 0.6?)
- $\Psi = \frac{e^{-0.0034}}{1+e^{-0.0034}}$
-
```
> exp(-0.00337)/(1+exp(-0.00337))
[1] 0.4991575
```

Parameter estimates

```
> names(m0)
[1] "psi" "col" "ext" "det"

> backTransform(m0, type="psi")

 Estimate     SE
    0.499 0.0591

> confint(backTransform(m0, type="psi"))
     0.025     0.975
 0.3854436 0.6129562
```

Derived parameters

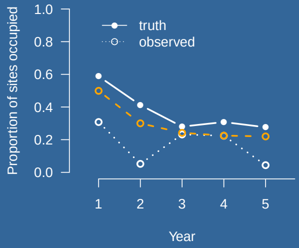
$$\Psi_t = \Psi_{t-1} \times (1 - \epsilon_{t-1}) + (1 - \Psi_{t-1}) \times \gamma_{t-1}$$

```
> backTransform(m0, type="ext")
 Estimate      SE
    0.553 0.0786

> backTransform(m0, type="col")
 Estimate      SE
    0.154 0.0255

> m0@projected.mean
                   1         2         3         4        5
unoccupied 0.5008437 0.6997875 0.7580149 0.7750571 0.780045
occupied   0.4991563 0.3002125 0.2419851 0.2249429 0.219955
```

Will occupancy decrease forever?

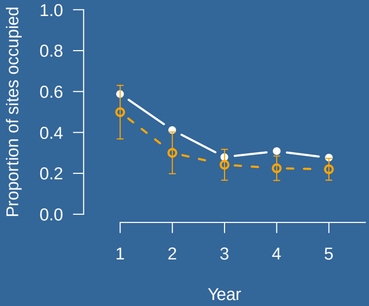


With constant γ and ϵ occupancy heading for an equilibrium:

$$\Psi_E = \frac{\gamma}{\gamma + \epsilon}$$

$$\hat{\Psi}_E = \frac{0.154}{0.154 + 0.553} = 0.218$$

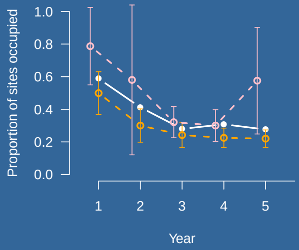
Confidence intervals for derived parameters



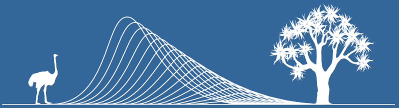

```
> m0 <- nonparboot(m0, B = 1000)

> cbind(projected=projected(m0)[2,],
      SE=m0@projected.mean.bsse[2,])
  projected          SE
1 0.4991563 0.06678017
2 0.3002125 0.05202317
3 0.2419851 0.03846406
4 0.2249429 0.03041457
5 0.2199550 0.02719051
```

Yearly site covariates




```
> m1 <- colext(psiformula= ~1,
+              gammaformula = ~ year,
+              epsilonformula = ~ year,
+              pformula = ~ year,
+              data = simUMF)
```



Model selection

```
> AIC(m0,m1)
Error in UseMethod("logLik") :
  no applicable method for 'logLik' applied to an object
  of class "c('unmarkedFitColExt', 'unmarkedFit')"

> models <- fitList(' psi(.)gam(.)eps(.)p(.)' = m0,
+                   ' psi(.)gam(Y)eps(Y)p(Y)' = m1)
> (ms <- modSel(models))
                        nPars     AIC  delta   AICwt cumltvWt
 psi(.)gam(Y)eps(Y)p(Y)    14 1444.53   0.00 1.0e+00     1.00
 psi(.)gam(.)eps(.)p(.)     4 1556.16 111.63 5.8e-25     1.00
```

Key references

- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:22002207. **This paper introduced dynamic occupancy models.**
- Weir, L. A., I. J. Fiske, and J. A. Royle. 2009. Trends in Anuran Occupancy from Northeastern States of the North American Amphibian Monitoring Program. Herpetological Conservation and Biology 4:389402. **Early application with detailed description.**
- Ian Fiske, Richard Chandler (2011). unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. Journal of Statistical Software, 43(10), 1-23. URL http://www.jstatsoft.org/v43/i10/. **This paper introduced unmarked package; see also package vignettes.**
- Bailey, L. L., D. I. MacKenzie, and J. D. Nichols. 2014. Advances and applications of occupancy models. Methods in Ecology and Evolution 5:12691279. **Concise overview of occupancy models.**
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2017. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Academic Press. **The bible of occupancy modelling.**