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 you to statistical methods that are useful for ecologists, environmental and conservation scientists.

Our next seminar:

Topic: **Species occupancy models**
Who: Res Altwegg
When: Thursday 31 May (1-2pm)
Where: John Day LT3

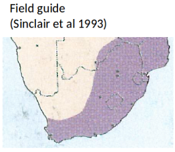
More details: www.seec.uct.ac.za

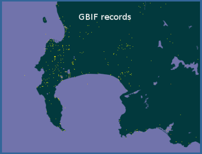
SEEC - Statistics in Ecology, Environment and Conservation

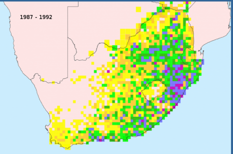
University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

Occupancy

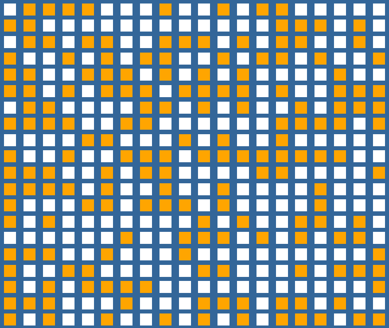
- Where a species occurs; which of a set of suitable patches are occupied; what determines where a species can live...
- (Metapopulation) ecology, conservation, red-listing



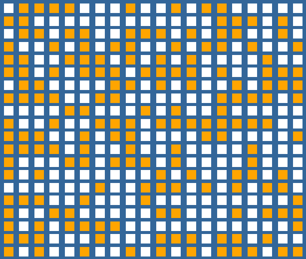




Occupancy: the proportion of sites occupied by a species



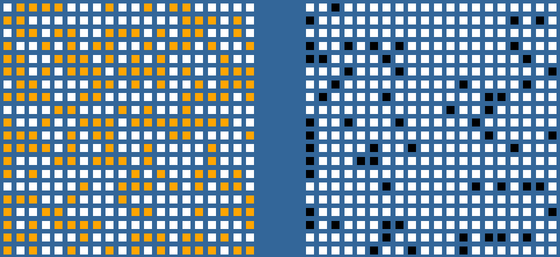
Occupancy: the proportion of sites occupied by a species



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

The species is not detected in all occupied cells

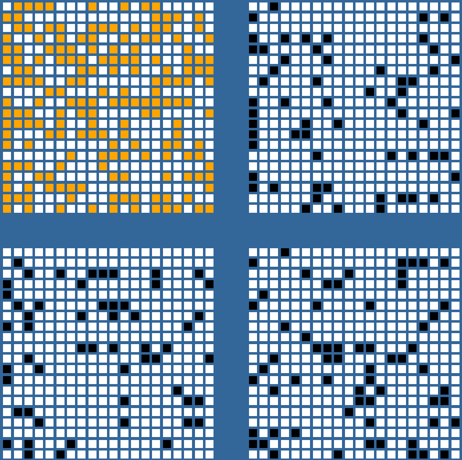
Detection probability $p < 1$



'Naive approach':

- $\Psi \times p = \frac{occupied}{total}$
- $logit(\Psi \times p) = f(covariates)$

The species is not detected in all occupied cells

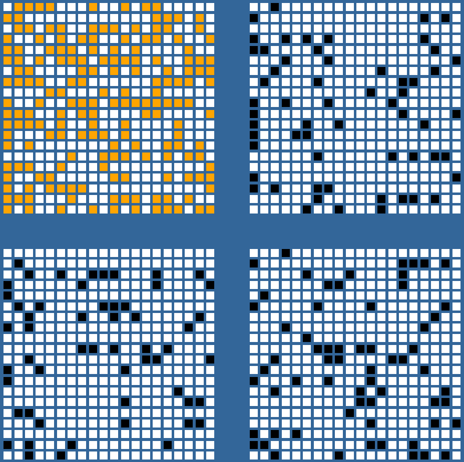


Repeated sampling

Assumptions:

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

The species is not detected in all occupied cells



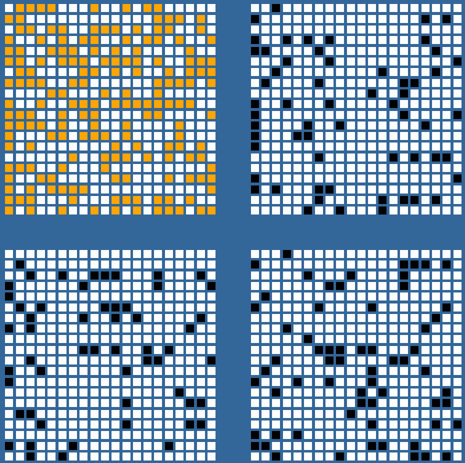
Survey histories:

1 = detected

0 = not detected

- (1,1) 000
- (1,3) 100
- (2,1) 101
- (1,9) 000

The species is not detected in all occupied cells



Survey histories:

1 = detected

0 = not detected

- (1,1) 000
- (1,3) 100
- (2,1) 101
- **(1,9) 000**

How many occupied cells have no detections?

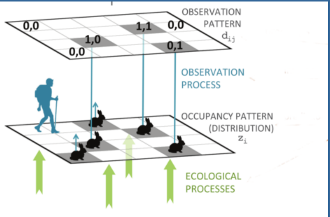
A model for the detections

Ψ = probability of a cell to be occupied
p = probability of detecting the species given that the cell is occupied
K = number of visits to each site
y_i = number of detections at site i

$$Pr(Y = y_i) = \Psi \binom{K}{y_i} p^{y_i} (1-p)^{K-y_i}, y_i > 0$$

$$= \Psi(1-p)^K + (1-\Psi), y_i = 0$$

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

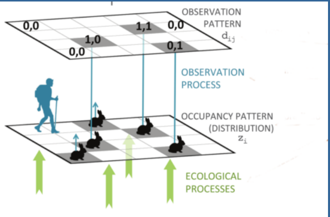
$$z_i|\Psi_i \sim Bernoulli(\Psi_i)$$

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

A hierarchical model

Ecological process:
$z_i | \Psi \sim Bernoulli(\Psi)$

Observation process:
$y_{ij} | z_i, p \sim Bernoulli(z_i p)$

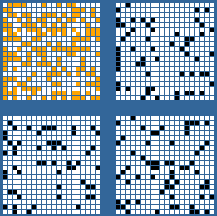



```
model {
for (i in 1:nsites) {
z[i] ~ dbern(psi)

  p.eff[i] <- z[i] * p
  for (j in 1:nvisits) {
    y[i,j] ~ dbern(p.eff[i])
  } #j
} #i

# Priors
psi ~ dunif(0, 1)
p ~ dunif(0, 1)

# Derived quantities
occ.fs <- sum(z[])
}
```

Preparing the data

$\Psi = 0.5$

$p = 0.3$



'Naive' occupancy:

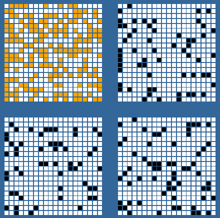
$\frac{126}{400} = 0.32$

```
> head(y)
  detected1 detected2 detected3
1         0         0         0
2         0         0         0
3         0         1         1
4         0         0         0
5         0         0         0
6         1         1         0

> dim(y)
[1] 400   3
```

Preparing inputs for JAGS

$\Psi = 0.5$

$p = 0.3$



'Naive' occupancy:

$\frac{126}{400} = 0.32$

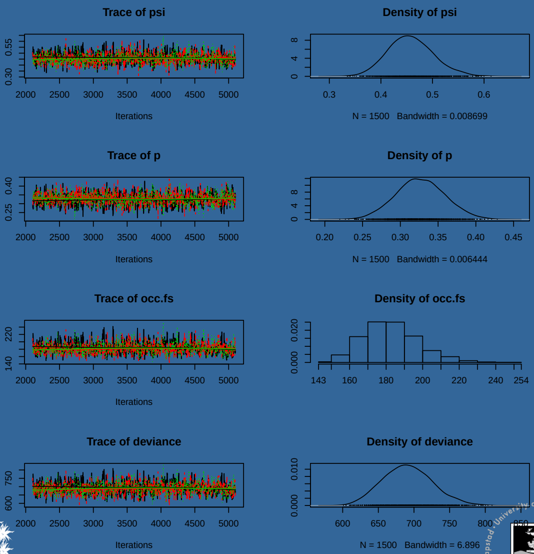
```
library(jagsUI) # requires JAGS

occ.data <- list(y = y, nsites = nrow(y),
                 nvisits = ncol(y))
# Initial values
zst <- apply(y, 1, max)
inits <- function() list(z = zst)

# Parameters monitored
params <- c("psi", "p", "occ.fs")

# MCMC settings
nc <- 3; ni <- 5000; nb <- 2000
```

Fitting the model to the data

```
> out.occ <- jags(occ.data, inits, params, "occ.txt",
      n.chains=nc, n.iter=ni, n.burn = nb)
> out.occ$summary
             mean      sd   Rhat n.eff
psi        0.4578  0.0448 1.0028   752
p          0.3274  0.0331 1.0016  1675
occ.fs   182.8655 15.2283 1.0016  1312
deviance 690.6184 35.3438 1.0016  1276
```

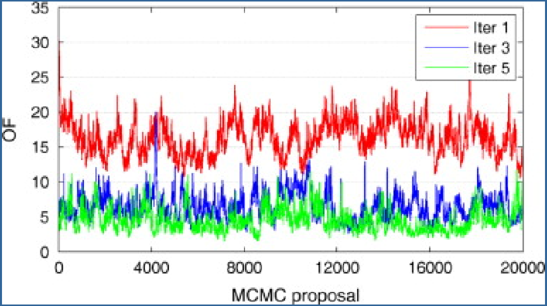
WARNING: MCMC can be dangerous!

```
> plot(out.occ)
```



Bad example

If your chains look like this, don't trust the output!!



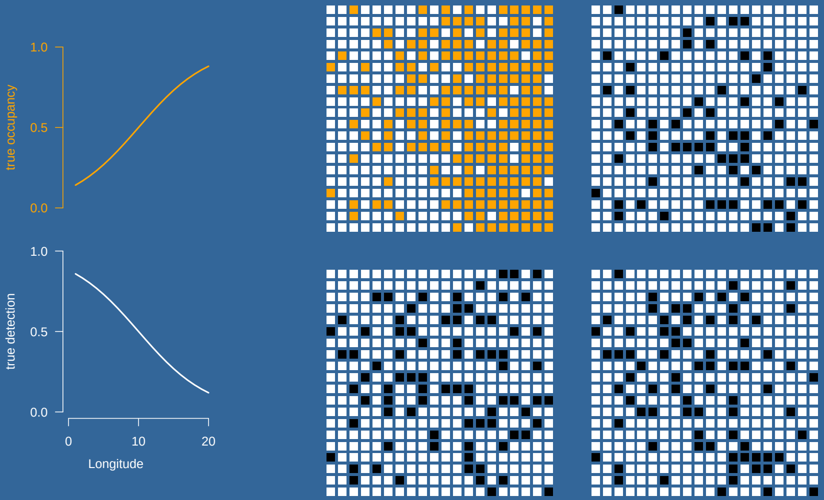
Covariate modelling

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

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

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

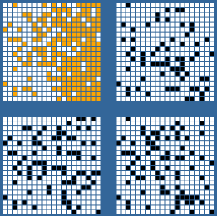
Occupancy and detection vary in space



Occupancy and detection vary in space

$$logit(\Psi_i) = \beta_1^{\Psi} + \beta_2^{\Psi} long_i$$

$$logit(p_{ij}) = \beta_1^{p} + \beta_2^{p} long_i$$



```
model {
# Ecological model
for (i in 1:nsites) {
  z[i] ~ dbern(psi[i])

  # Observation model
  p.eff[i] <- z[i] * p[i]
  for (j in 1:nvisits) {
    y[i,j] ~ dbern(p.eff[i])
  } #j
} #i

# covariates
for (i in 1:nsites){
  logit(psi[i]) = beta.psi[1] + beta.psi[2]*long[i]
  logit(p[i]) = beta.p[1] + beta.p[2] * long[i]
}

# Priors
for (b in 1:2){
  beta.psi[b] ~ dnorm(0, 0.01)
  beta.p[b] ~ dnorm(0, 0.01)
}
}
```

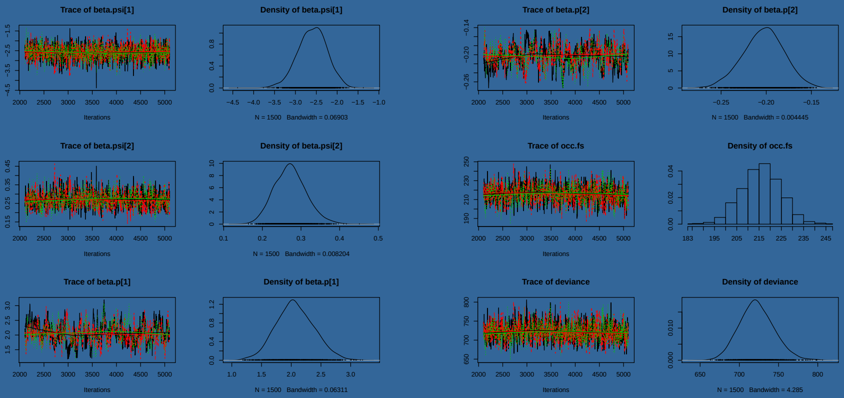
Occupancy and detection vary in space

```
occ.data <- list(y = y, nsites = nrow(y),
                 nvisits = ncol(y), long=long)

inits <- function() list(z = zst, beta.psi=runif(2,-3,3),
                                  beta.p=runif(2,-3,3))

params <- c("beta.psi", "beta.p", "occ.fs")

out.occ.cov <- jags(occ.data,inits,params,"occ_cov.txt",
         n.chains=3, n.iter=5000, n.burn = 2000)
```

Occupancy and detection vary in space

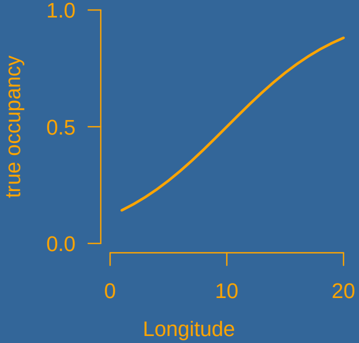
```
> out.occ.cov$summary
                 mean     sd   2.5%   97.5% Rhat n.eff
beta.psi[1]   -2.59  0.350  -3.31   -1.93    1  2674
beta.psi[2]    0.27  0.042   0.20    0.36    1  2755
beta.p[1]      2.07  0.320   1.46    2.72    1  1340
beta.p[2]     -0.20  0.023  -0.25   -0.16    1  1185
deviance     722.23 22.120 681.30  768.56    1  4385
```

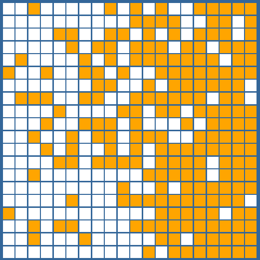
Check convergence!



Estimated occupancy probability

$$logit(\Psi_i) = \beta_0 + \beta_1 long_i$$





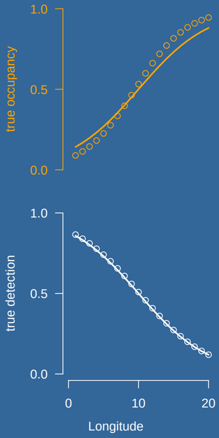
Estimated occupancy probability

$$logit(\Psi_i) = \beta_0 + \beta_1 \times long_i$$

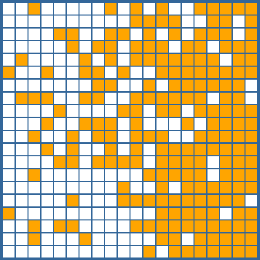
$$\Psi_i = \frac{1}{1+e^{-(\beta_0+\beta_1 \times long_i)}}$$

```
new.long <- 1:20
pr.s <- inv.logit(-2.59 + 0.27 * new.long)
> pr.s
 [1] 0.089 0.114 0.145 0.182 0.226 0.277 0.334 0.397
 [9] 0.464 0.532 0.599 0.662 0.720 0.772 0.816 0.853
 [17] 0.884 0.909 0.930 0.945
```

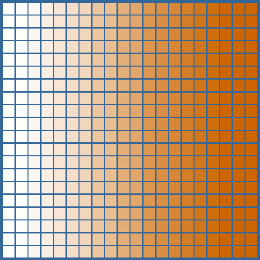
Estimated occupancy probability



True occupancy



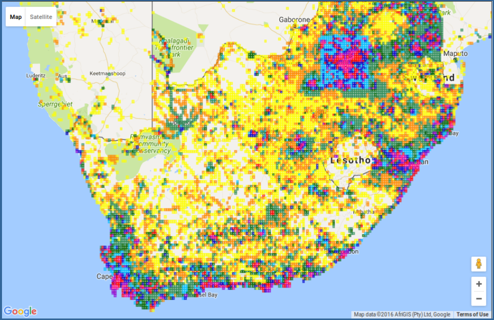
Fitted occupancy



Southern bals ibis range in South Africa

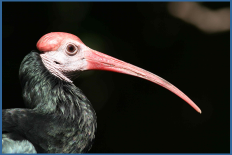
www.flickr.com/photos/12457947@N07/4251701580

Second Southern African Bird Atlas Project

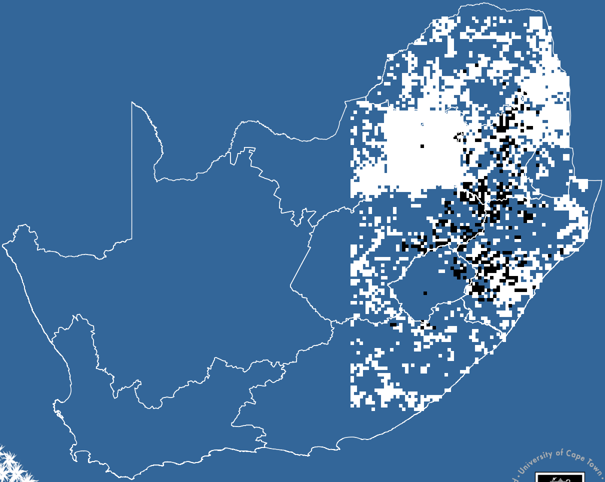


http://sabap2.adu.org.za/

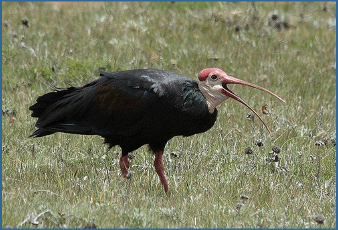
Southern bals ibis



© Peter Ryan



Southern bals ibis



www.flickr.com/photos/12457947@N07/4251701580

- Data: 30 June 2015 to 1 July 2017
- 3220 grid cells $5' \times 5'$
- 26'619 checklists (1 to 719 per cell)
- **Site-level covariates:** mean temp coldest month, mean temp warmest month, ratio actual to potential evapotranspiration, wet season intensity
- **Survey-specific covariates:** log(hours observed)

Preparing the data: long table format

```
> head(bi.m)
     Pentad Start_Date       lat     long Total_hours Spp
2240_2820 2016-05-28 -22.70833 28.37500           4   0
2240_2820 2015-10-10 -22.70833 28.37500           2   0
2235_2825 2015-10-11 -22.62500 28.45833           2   0
2235_2815 2015-09-25 -22.62500 28.29167           4   0
2240_2815 2015-09-25 -22.70833 28.29167           2   0
```

After some data wrangling

```
> y[1:5,1:10]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    0   NA   NA   NA   NA   NA   NA   NA   NA    NA
[2,]    0   NA   NA   NA   NA   NA   NA   NA   NA    NA
[3,]    0    0    0    0    0    0    0    0    0     0
[4,]    0    0   NA   NA   NA   NA   NA   NA   NA    NA
[5,]    0    0   NA   NA   NA   NA   NA   NA   NA    NA

> dim(y)
[1] 3220  719
```

Survey-specific covariates

```
> lhours[1:5,1:10]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 0.69   NA   NA   NA   NA   NA   NA   NA   NA    NA
[2,] 1.39   NA   NA   NA   NA   NA   NA   NA   NA    NA
[3,] 1.61 1.39  1.6  2.1  1.1 0.69  1.8  1.4 0.69   1.4
[4,] 0.69 1.39   NA   NA   NA   NA   NA   NA   NA    NA
[5,] 1.61 0.69   NA   NA   NA   NA   NA   NA   NA    NA

> dim(lhours)
[1] 3220  719
```

Site-specific covariates

```
> head(MTCO)
         [,1]
[1,] 0.981870
[2,] 0.981870
[3,] 1.041489
[4,] 1.071299
[5,] 1.429016
[6,] 1.429016

> dim(MTCO)
[1] 3220    1
```

- One value per grid cell
- Covariates scaled to $\bar{x} = 0, s = 1$

The model

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

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

$$logit(\Psi_i) =$$

$$\beta_1^\Psi + \beta_2^\Psi MTCO_i + \beta_3^\Psi MTWA_i + \beta_4^\Psi AET.PET_i + \beta_5^\Psi Wet.intensity_i$$

$$logit(p_{ij}) = \beta_1^p + \beta_2^p lhours_{ij}$$

The model

New features:

- # visits varies among sites
- multiple covariates
- observation-level covariates

```
model {
for (i in 1:nsites) {
  z[i] ~ dbern(psi[i])

   for (j in 1:ncards[i]) {
    p.eff[i,j] <- z[i] * p[i,j]
    y[i,j] ~ dbern(p.eff[i,j])
  } #j
} #i

# covariates
for (i in 1:nsites){
  logit(psi[i]) = beta.psi[1]
       + beta.psi[2] * MTCO[i]
       + beta.psi[3] * MTWA[i]
       + beta.psi[4] * AET.PET[i]
       + beta.psi[5] * Wet.Intensity[i]

  for (j in 1:ncards[i]){
    logit(p[i,j]) = beta.p[1]
       + beta.p[2] * lhours[i,j]
  } #j
} #i
.
.
```

The model

New features:

- # visits varies among sites
- multiple covariates
- observation-level covariates

```
.
.
.

# Priors
for (b in 1:5){
  beta.psi[b] ~ dnorm(0, 0.01)
}

for (b in 1:2){
  beta.p[b] ~ dnorm(0, 0.01)
}
}
```

The model

Data:

- survey histories `y`
- site-level covariates
  - `MTCO`
  - `MTWA`
  - `AET.PET`
  - `Wet.intensity`
- observation-level covariates: `lhours`
- # sites: `nsites`
- # visits: `ncards`

Parameters:

- coefficients for site-level covariates: `beta.psi`
- coefficients for observation-level covariates: `beta.p`

→ need priors

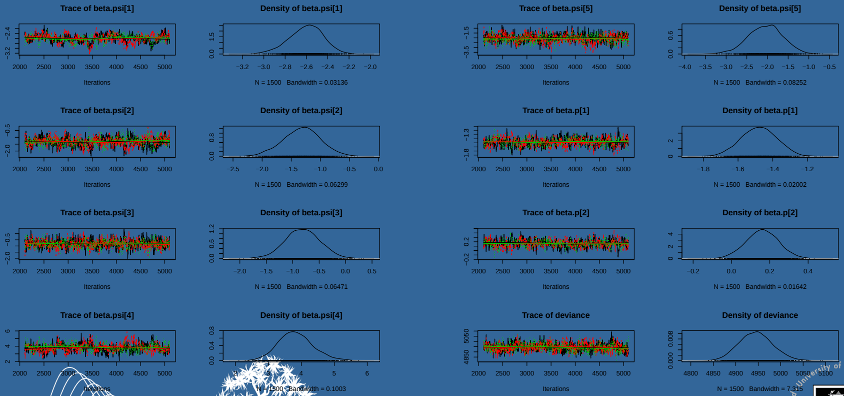
→ need initial values

Unobserved true occupancy state: z_i

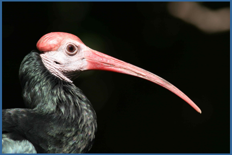
→ need initial values

Fit model and check convergence

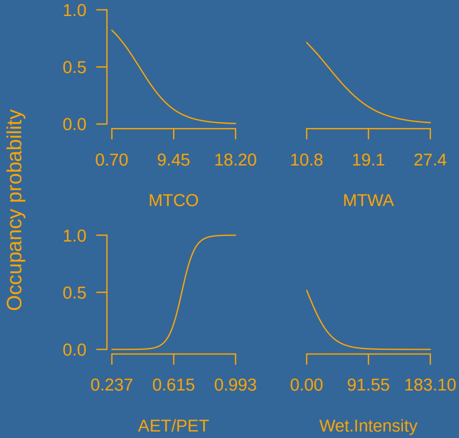
```
out.occ.ib <- jags(occ.data,inits,params,"occ_ibis.txt",
                   n.chains=3, n.iter=5000, n.burn = 2000)
plot(out.occ.ib)
```



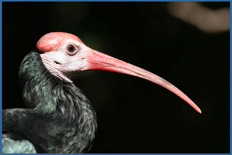
Southern bals ibis



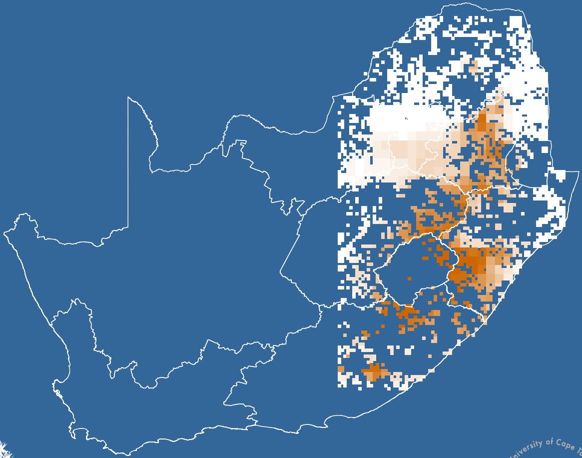
Ⓒ Peter Ryan



Southern bals ibis



© Peter Ryan



Single-season occupancy models

- Repeated detection / non-detection data
- Estimate occupancy and detection process
- **Key assumptions:** closure, no false detections, surveys are independent, sites are independent
- Can be fitted using JAGS via R package 'jagsUI'
- Other software: R package 'unmarked', PRESENCE, MARK

Why go Bayesian?

- more flexible
- easy to add random effects
- spatial autocorrelation
- can use prior information

Key references

Single-season occupancy models:

- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83: 2248-2255.
- 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.
- Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40:281295.

Occupancy models in BUGS:

- Kéry, M., and M. Schaub. 2012. Bayesian Population Analysis using WinBUGS: A hierarchical perspective. Academic Press.
- Kéry, M., and J. A. Royle. 2016. Applied Hierarchical Modeling in Ecology. Academic Press.