# 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:

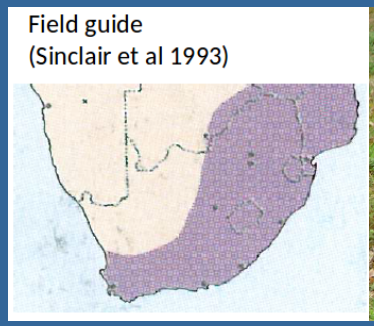
Topic: **Occupancy Models**
Who: Prof Res Altwegg
When: **Thursday 25 August 2016 (1-2pm)**
Where: PD Hahn Lecture Theatre 3,
PD Hahn Building Level 4, UCT

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

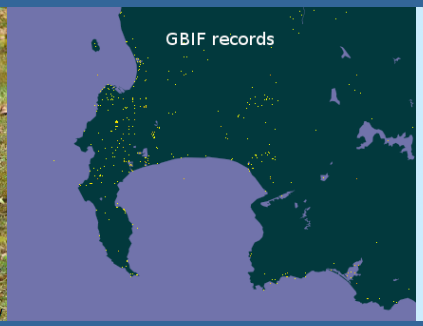
SEEC - Statistics in Ecology, Environment and Conservation

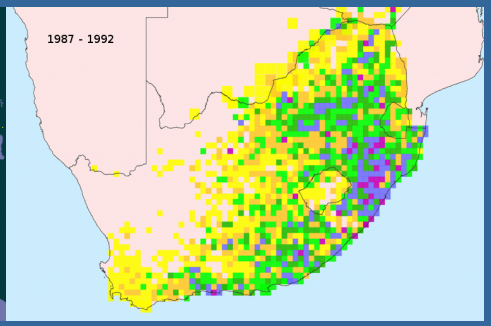
# Occupancy

- 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

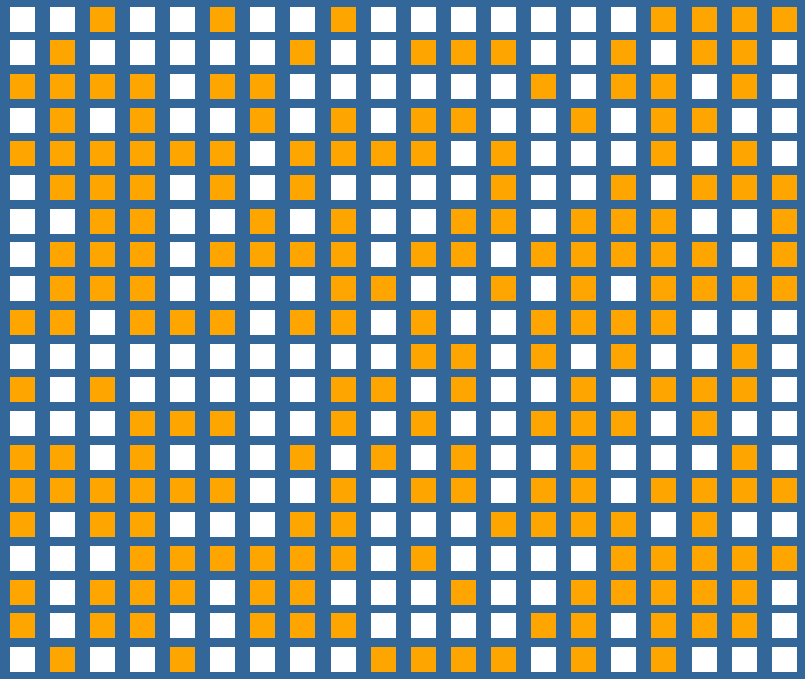




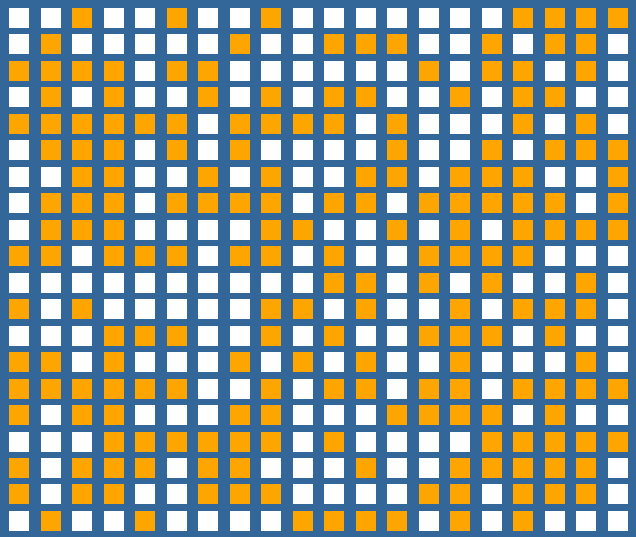


# Occupancy: the proportion of sites occupied by a species



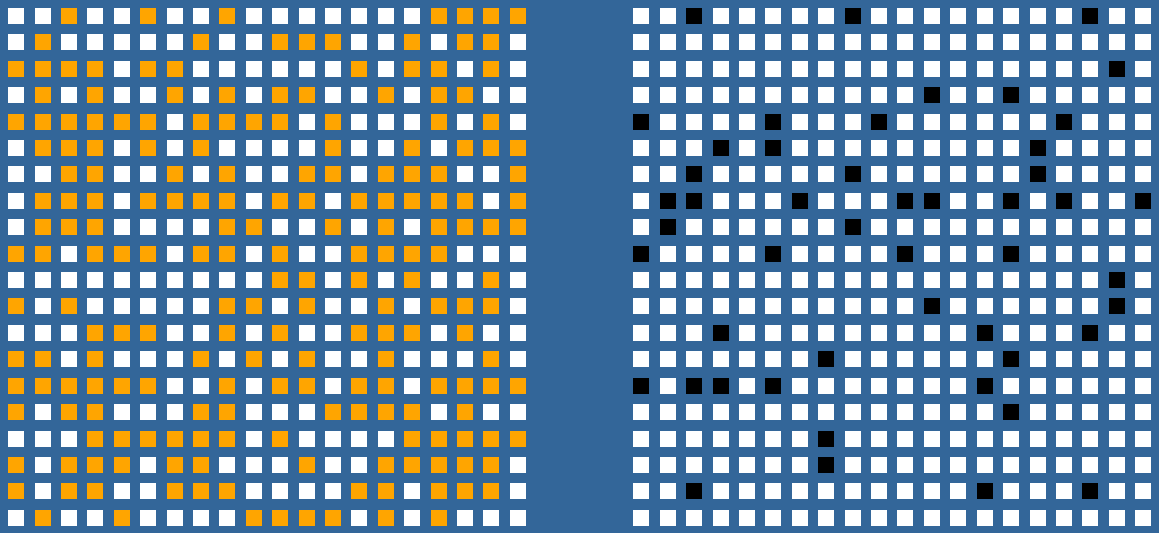
# Occupancy: the proportion of sites occupied by a species



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

# The species is not detected in all occupied cells

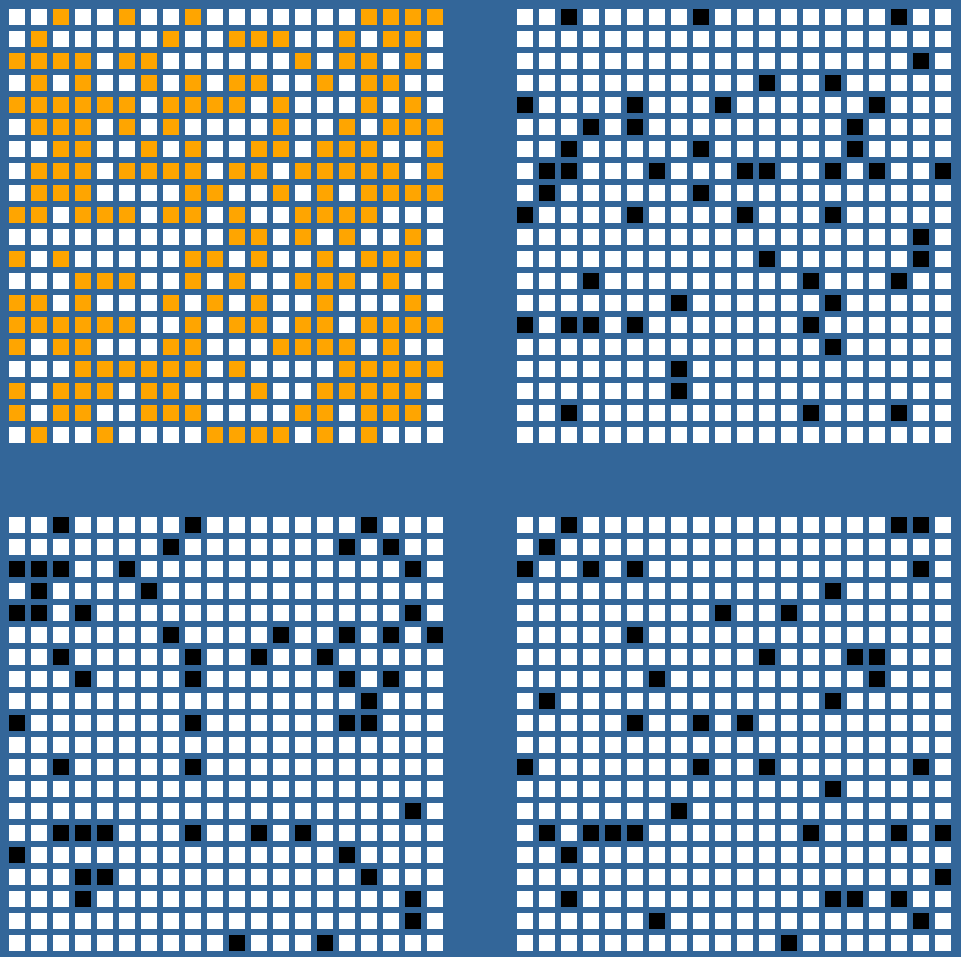
Detection probability $p < 1$



'Naive approach':

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

# The species is not detected in all occupied cells

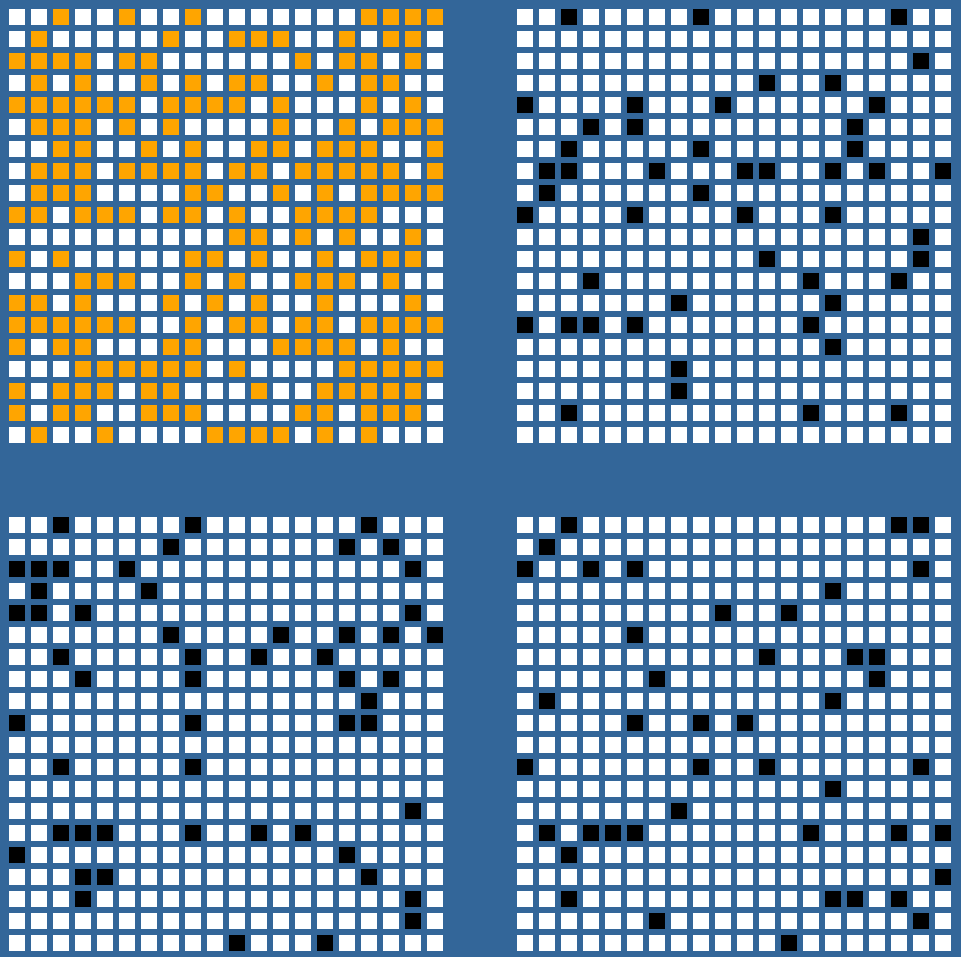


Repeated sampling

**Assumptions:**

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

# The species is not detected in all occupied cells



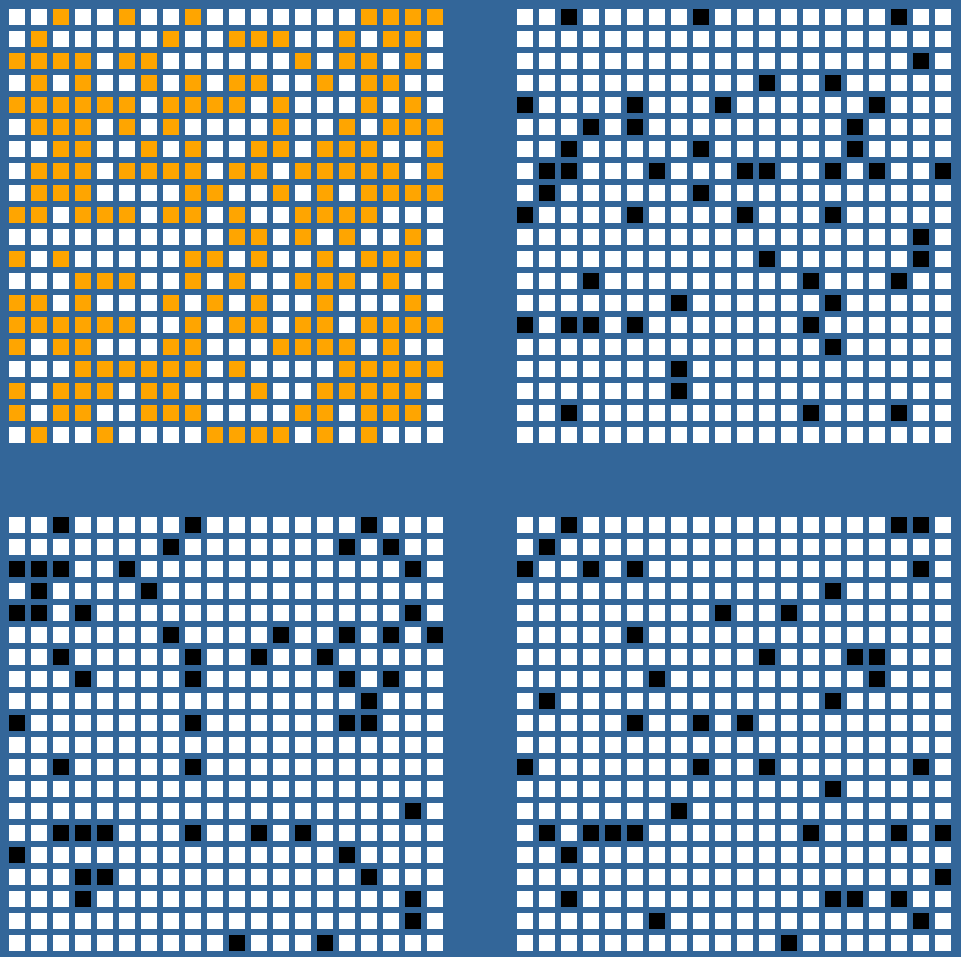
**Survey histories:**

1 = detected

0 = not detected

- (1,1) 000
- (1,3) 111
- (2,2) 001
- (1,6) 000

# The species is not detected in all occupied cells



**Survey histories:**

1 = detected

0 = not detected

- (1,1) 000
- (1,3) 111
- (2,2) 001
- **(1,6) 000**

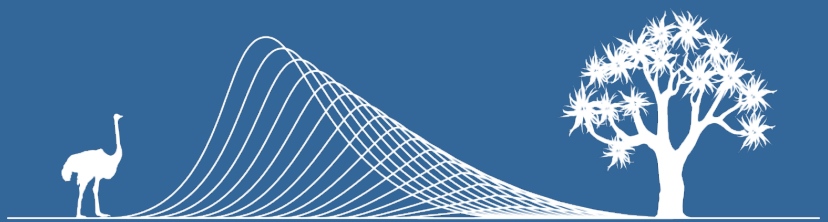
How many occupied cells have no detections?

# A model for the detections

$\Psi$ = 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 $i$

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

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

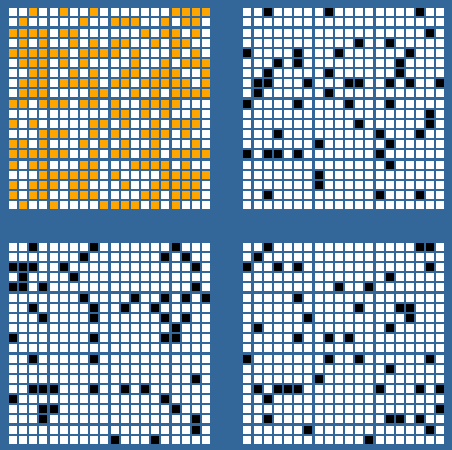




# Preparing the data

$\Psi = 0.5$

$p = 0.3$

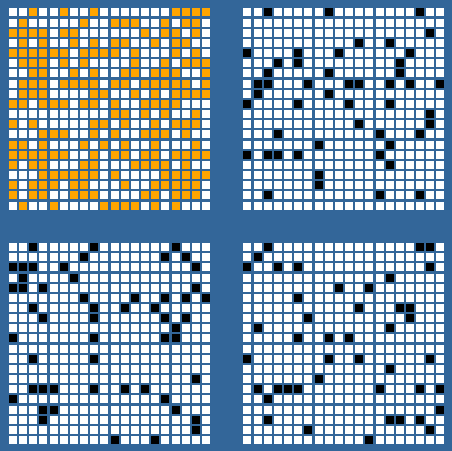


```
> head(y)
  detected1 detected2 detected3
1         0         0         0
2         0         0         0
3         1         1         1
4         0         0         0
5         0         0         0
6         0         0         0
```

# Fitting the model to the data

$\Psi = 0.5$

$p = 0.3$



'Naive' occupancy:

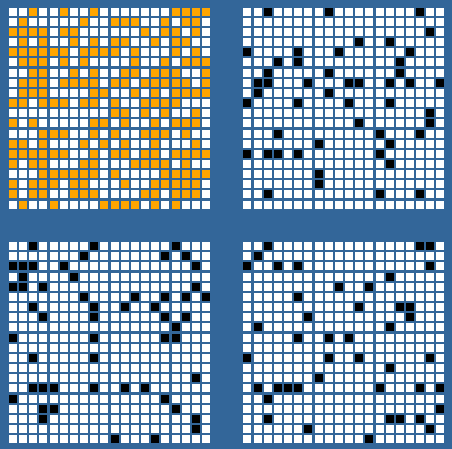
$\frac{128}{400} = 0.32$

```
> library(unmarked)
> ex1 <- unmarkedFrameOccu(y = y)
> summary(ex1)
unmarkedFrame Object

400 sites
Sites with >0 detections: 128
...
Tabulation of y observations:
   0    1 <NA>
1021  179    0
```

# Fitting the model to the data

$\Psi = 0.5$

$p = 0.3$

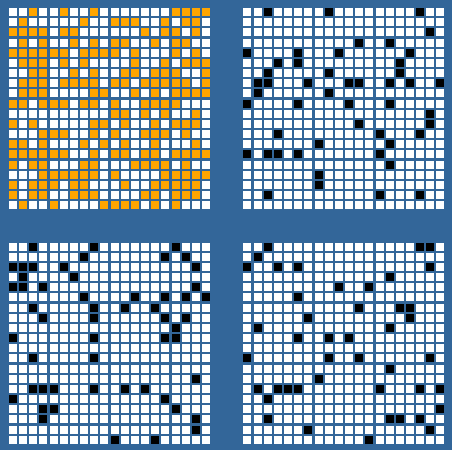


```
> m1 <- occu(~1 ~1, data=ex1)
> summary(m1)

Occupancy (logit-scale):
 Estimate    SE      z P(>|z|)
   -0.129 0.187 -0.688   0.491

Detection (logit-scale):
 Estimate    SE     z  P(>|z|)
   -0.759 0.153 -4.97 6.76e-07

AIC: 974.8914
```

# Fitting the model to the data

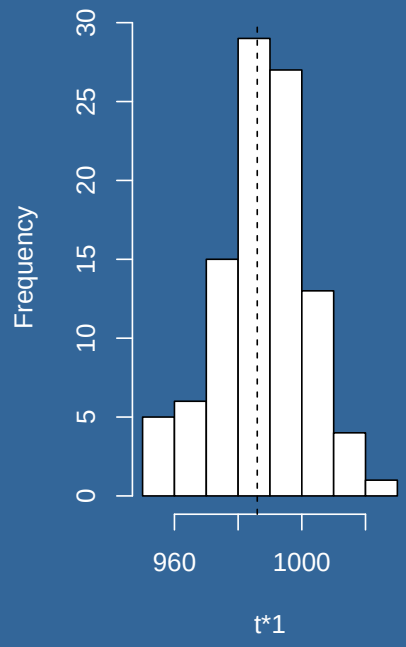
$\Psi = 0.5$

$p = 0.3$



```
> backTransform(m1, type="state")
Backtransformed Occupancy estimate(s)

 Estimate     SE LinComb (Intercept)
    0.468 0.0465  -0.129           1

> backTransform(m1, type="det")
Backtransformed Detection estimate(s)

 Estimate     SE LinComb (Intercept)
    0.319 0.0332  -0.759           1
```

# Goodness-of-fit




```
chisq <- function(fm) { # chi-square
  observed <- getY(fm@data)
  expected <- fitted(fm)
  sum((observed - expected)^2/expected)
  }

pb <- parboot(m1,
        statistic=chisq, nsim=100)

plot(pb, main = "")
```

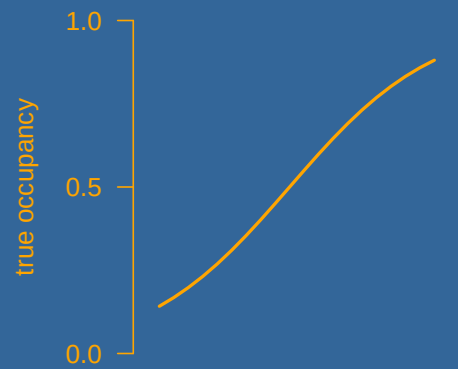
# Covariate modelling

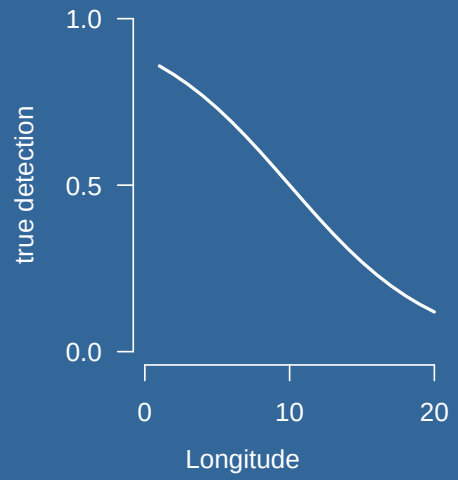
Want to know how occupancy and detection vary among sites, $i$, and visits, $j$.

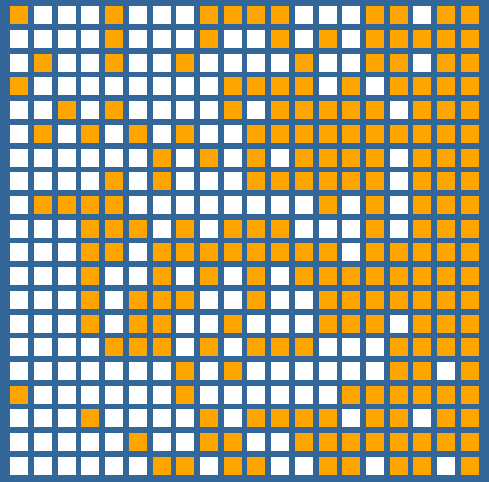
$$logit(\Psi_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_U x_{iU}$$
$$logit(p_{ij}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_U x_{iU} + \beta_{U+1} y_{ij1} + \ldots + \beta_{U+V} y_{ijV}$$

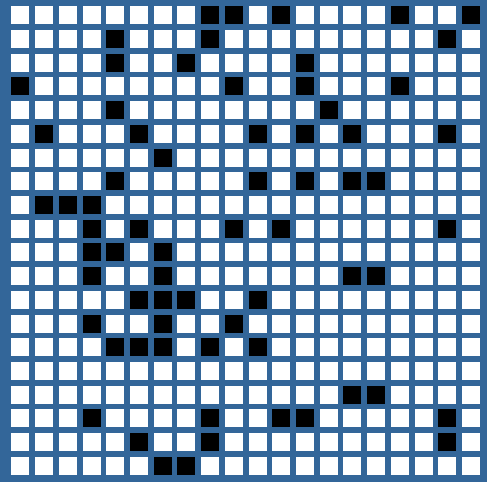
$U$ site-level covariates: $x_{i1}, \ldots, x_{iU}$
$V$ survey-specific covariates: $y_{ij1}, \ldots, y_{ijV}$

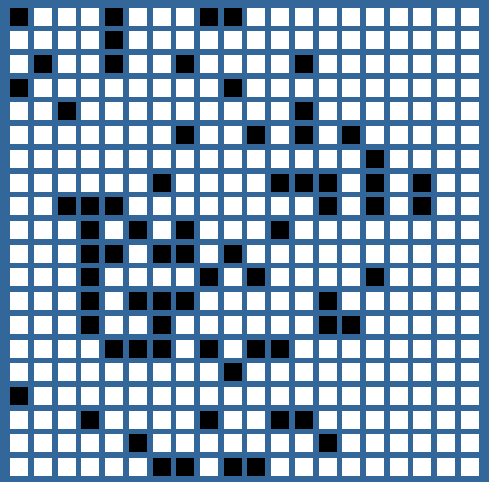
# Occupancy and detection vary in space

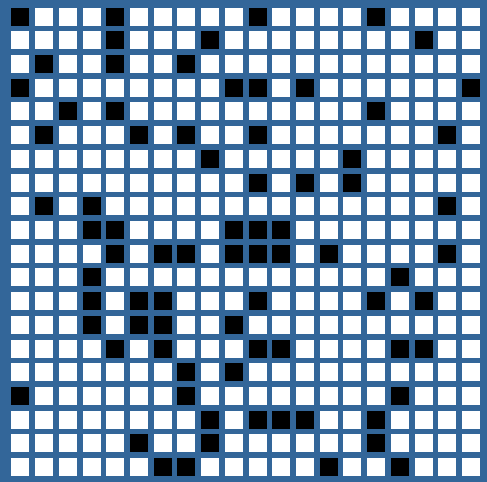


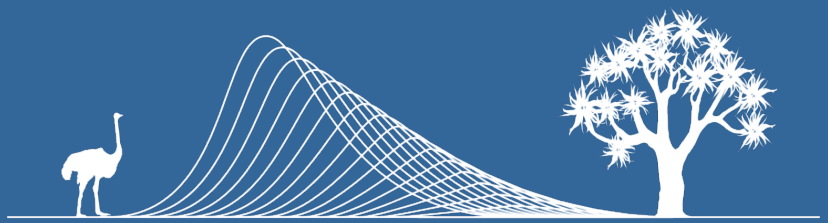





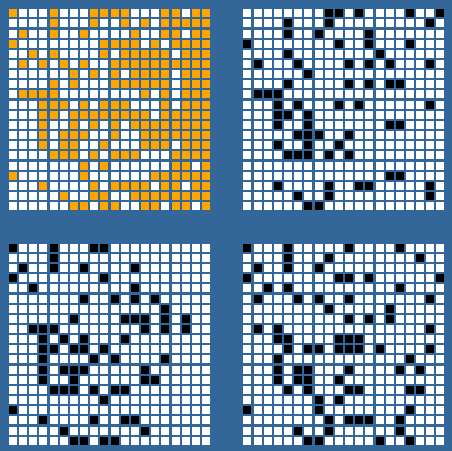






SEEC - Statistics in Ecology, Environment and Conservation

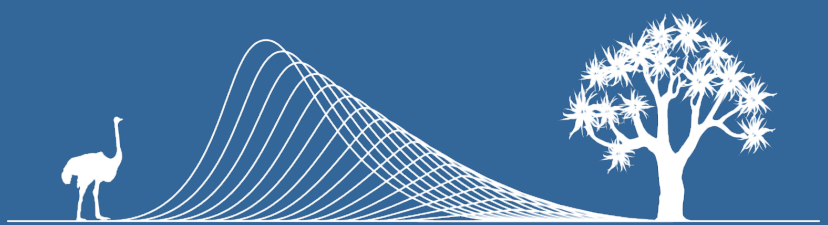
# Occupancy and detection vary in space



```
> siteCovs <- data.frame(long=
                           grid[,"long"])
> ex2 <- unmarkedFrameOccu(y = y,
                         siteCovs = siteCovs)
> summary(ex2)
     unmarkedFrame Object
     ...
     Site-level covariates:
          long
      Min.   : 1.00
      1st Qu.: 5.75
      Median :10.50
      Mean   :10.50
      3rd Qu.:15.25
      Max.   :20.00
```

# Occupancy and detection vary in space

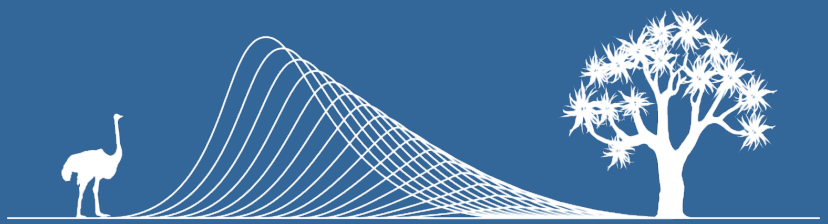
```
> m1 <- occu(~long ~long, data=ex2)
> summary(m1)
...
Occupancy (logit-scale):
             Estimate     SE     z  P(>|z|)
(Intercept)   -1.962 0.3115 -6.30 2.99e-10
long           0.202 0.0377  5.35 8.60e-08

Detection (logit-scale):
             Estimate     SE     z  P(>|z|)
(Intercept)    1.983 0.2853  6.95 3.64e-12
long          -0.195 0.0214 -9.10 9.06e-20
...
```



SEEC - Statistics in Ecology, Environment and Conservation

# Occupancy and detection vary in space

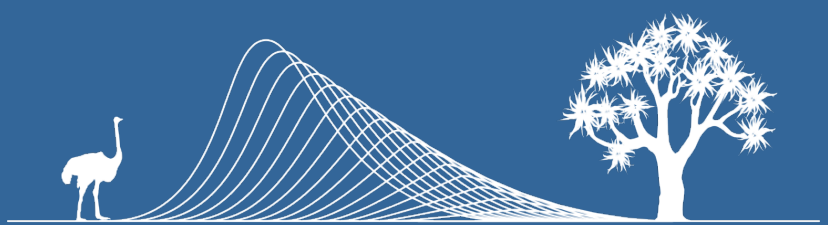
```
> m2 <- occu(~1 ~long, data=ex2)
> summary(m2)
...
Occupancy (logit-scale):
             Estimate     SE     z  P(>|z|)
(Intercept)  -1.0004 0.2586 -3.87 0.000109
long          0.0731 0.0219  3.35 0.000820

Detection (logit-scale):
 Estimate    SE     z P(>|z|)
   -0.201 0.123 -1.64   0.101
...
```



SEEC - Statistics in Ecology, Environment and Conservation

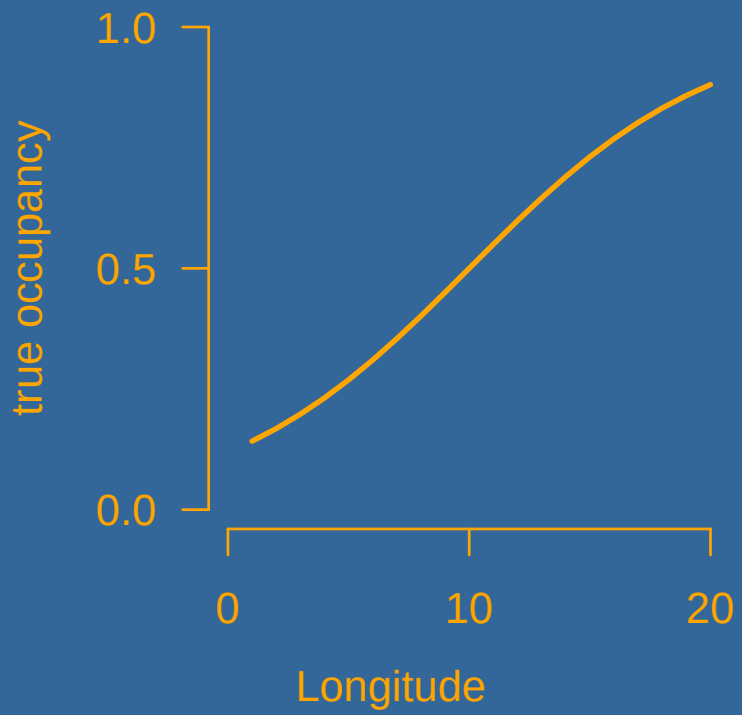
# Model selection

```
> m3 <- occu(~1 ~1, data=ex2)

> ml <- fitList("psi(long)p(long)" = m1,
                "psi(long)p(.)"= m2,
                "psi(.)p(.)" = m3)
> modSel(ml)
                   nPars     AIC delta   AICwt cumltvWt
psi(long)p(long)       4 1019.24  0.00 1.0e+00     1.00
psi(long)p(.)          3 1094.50 75.26 4.5e-17     1.00
psi(.)p(.)             2 1104.50 85.26 3.1e-19     1.00
```

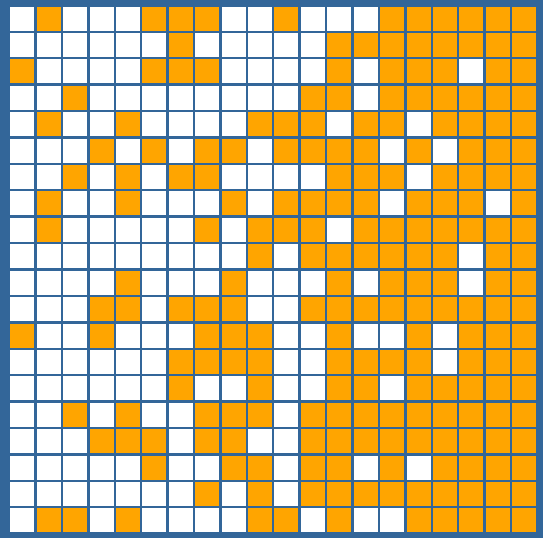




# Estimated occupancy probability

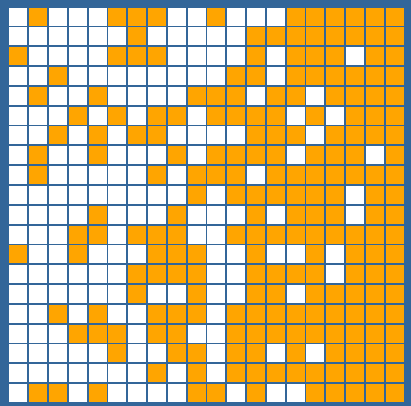



$$logit(\Psi_i) = \beta_0 + \beta_1 long_i$$



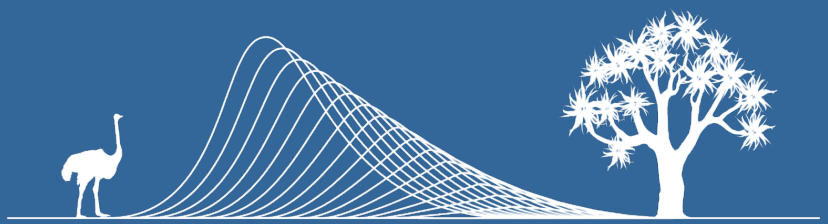
# Estimated occupancy probability

```
> backTransform(linearComb(m1, coefficients = c(1,0),
                    type = "state"))
 Estimate      SE LinComb (Intercept) long
    0.123 0.0337   -1.96             1    0

> backTransform(linearComb(m1, coefficients = c(1,20),
                    type = "state"))
 Estimate      SE LinComb (Intercept) long
    0.888 0.0499    2.07             1   20
```



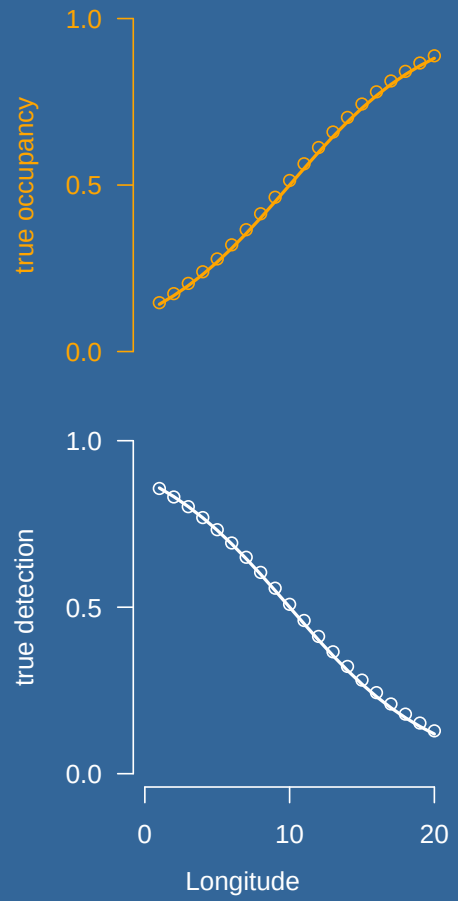
# Estimated occupancy probability

```
> newData <- data.frame(long=1:20)
> predict(m1, type = "state", newdata = newData,
          appendData=TRUE)
    Predicted         SE      lower     upper long
1   0.1467565 0.03493378 0.09053967 0.2290865    1
2   0.1738553 0.03561616 0.11460290 0.2549223    2
3   0.2047571 0.03567048 0.14354172 0.2834390    3
4   0.2395588 0.03519542 0.17744335 0.3150890    4
...
18 0.8414461 0.05756142 0.69495867 0.9251624   18
19 0.8665451 0.05400368 0.72221750 0.9419153   19
20 0.8881989 0.04993117 0.74780857 0.9551262   20
```

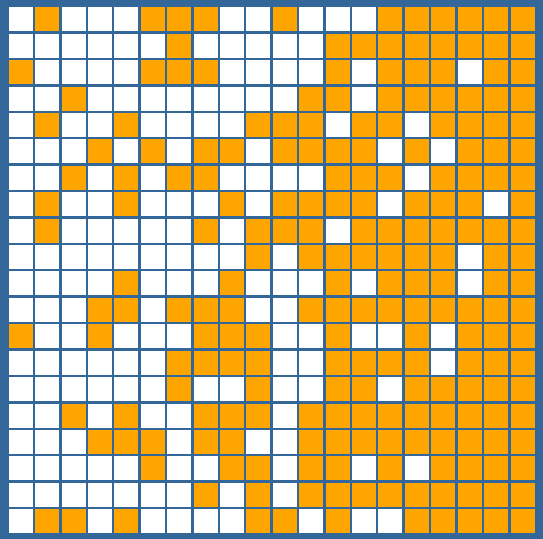


SEEC - Statistics in Ecology, Environment and Conservation

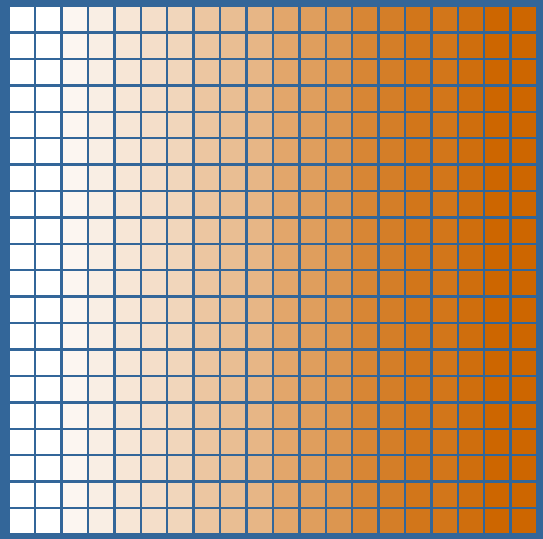
# Estimated occupancy probability

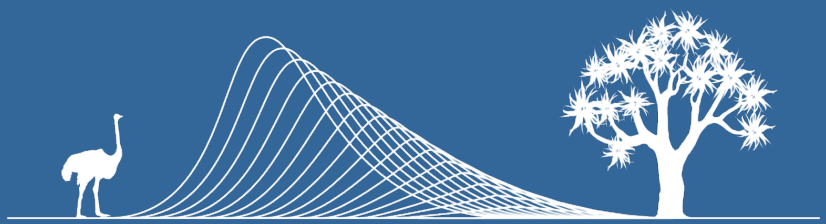



True occupancy



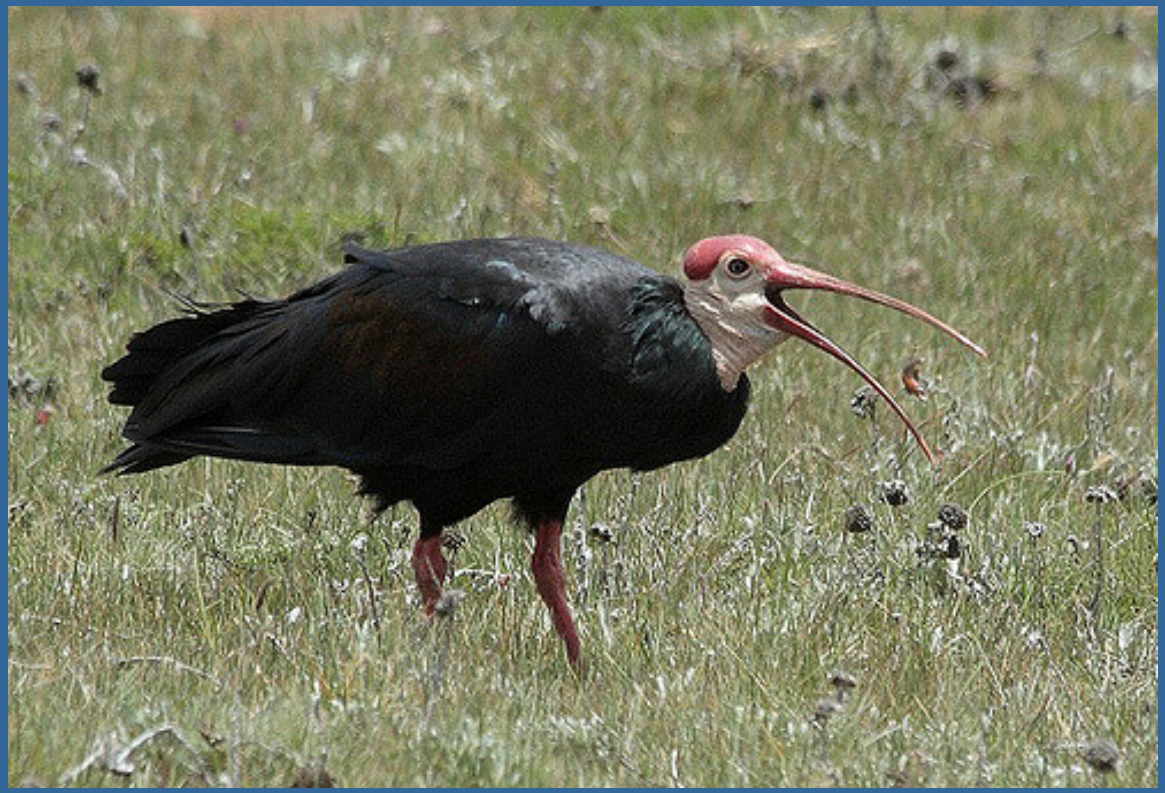
Fitted occupancy





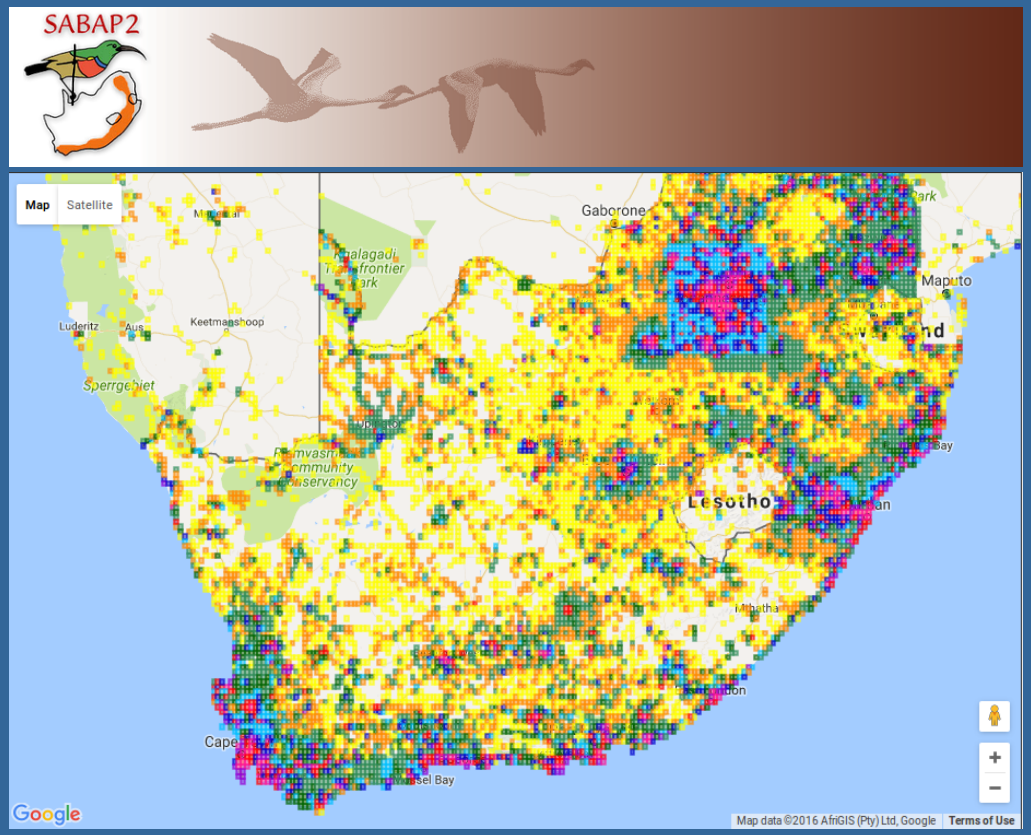
SEEC - Statistics in Ecology, Environment and Conservation

# Southern bals ibis range in South Africa



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

# Second Southern African Bird Atlas Project

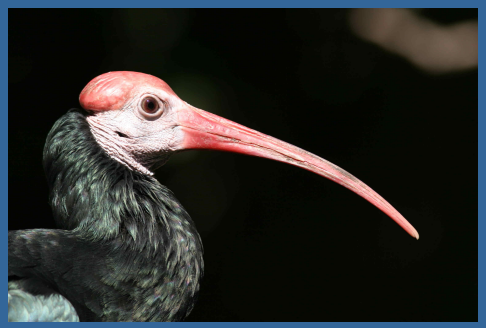



http://sabap2.adu.org.za/

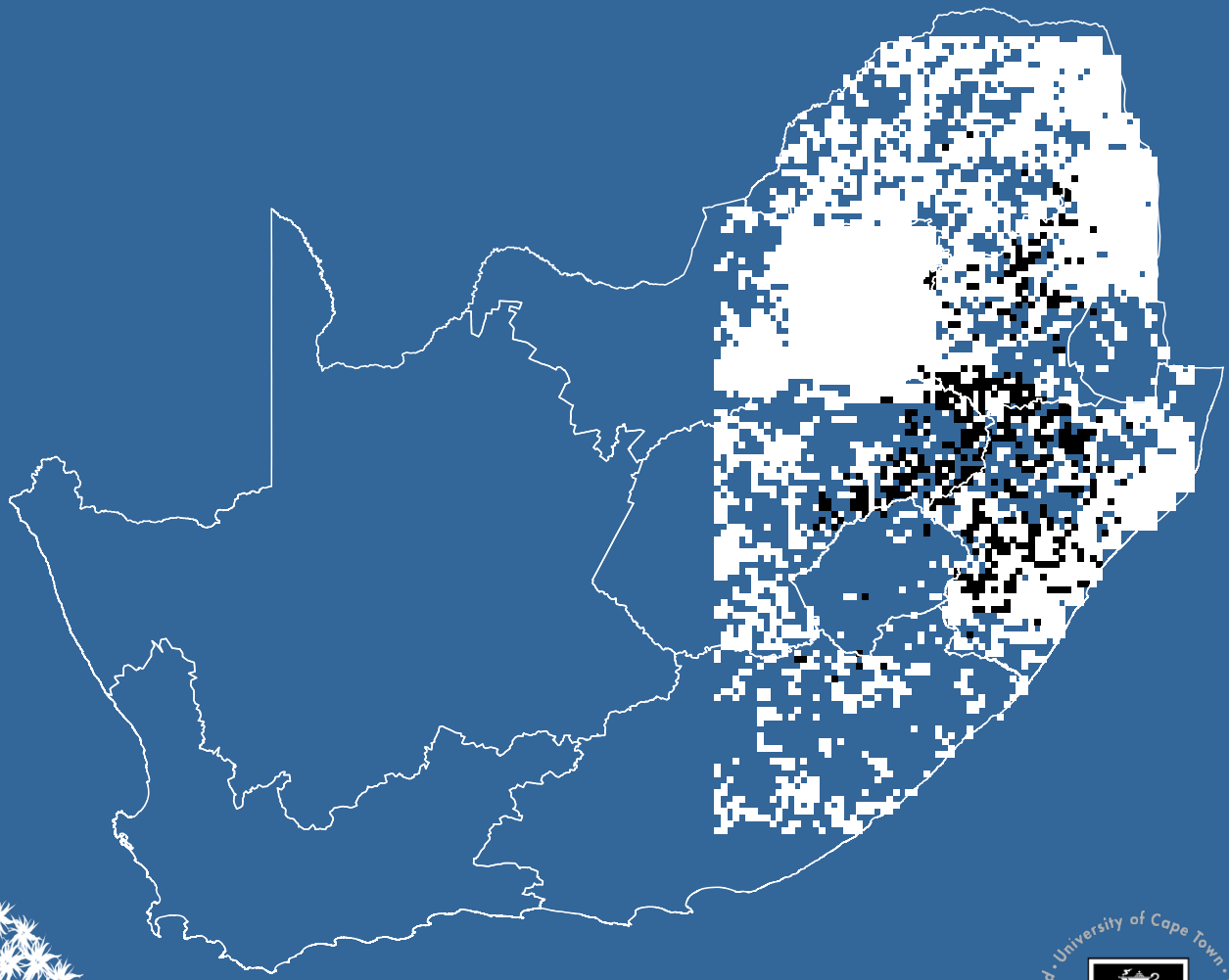
SEEC - Statistics in Ecology, Environment and Conservation

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad

# Southern bals ibis



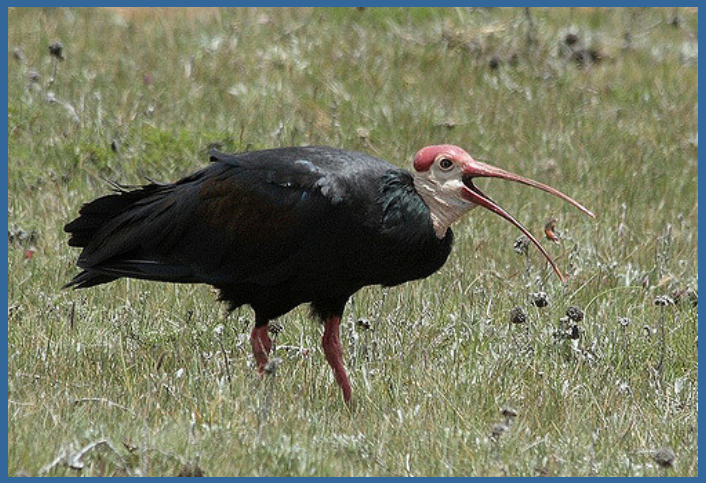
© Peter Ryan



SEEC - Statistics in Ecology, Environment and Conservation

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad

# Southern bals ibis

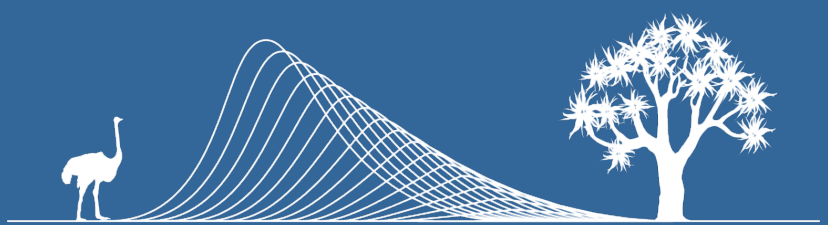


www.flickr.com/photos/12457947@N07/4251701580

- Data: 30 June 2014 to 1 July 2016
- 4191 grid cells $5' \times 5'$
- 37'172 checklists (1 to 680 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

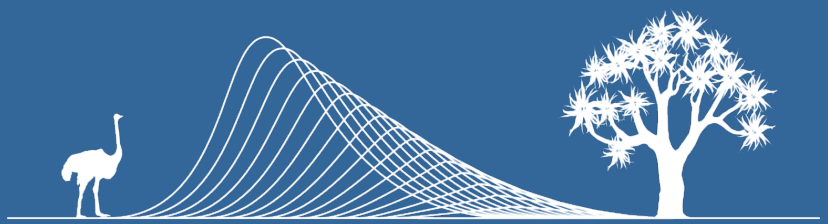
```
> head(bi.m)
    Pentad Start_Date        lat     long Total_hours Spp
2240_2820 2016-05-28 -22.70833 28.37500           4   0
2240_2825 2013-12-10 -22.70833 28.45833           7   0
2240_2820 2015-10-10 -22.70833 28.37500           2   0
2235_2825 2015-10-11 -22.62500 28.45833           2   0
2235_2815 2015-09-25 -22.62500 28.29167           4   0
2240_2815 2015-09-25 -22.70833 28.29167           2   0
```





# Convert to 'unmarked' format

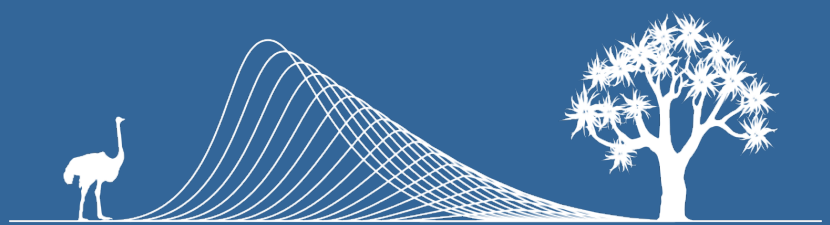
```
> bi.umf <- csvToUMF("bi.csv",long=TRUE,
                     type = "unmarkedFrameOccu")
> summary(bi.umf)
...
Observation-level covariates:
     lhours             JulianDate
 Min.   :0.7        Min.   :   1.0
 1st Qu.:0.7        1st Qu.: 336.0
 Median :0.7        Median : 613.0
 Mean   :1.0        Mean   : 595.2
 3rd Qu.:1.4        3rd Qu.: 872.0
 Max.   :3.9        Max.   :1096.0
 NA's   :2812708    NA's   :2812708
```



SEEC - Statistics in Ecology, Environment and Conservation

# Survey-specific covariates

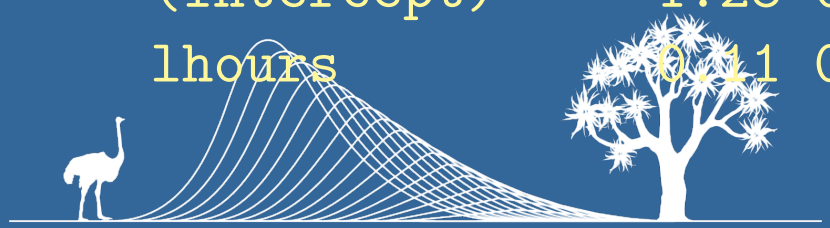
```
> head(obsCovs(bi.umf,matrices=T)$lhours)
          [,1]      [,2]     [,3]     [,4]     [,5]
[1,] 1.6094379        NA       NA       NA       NA
[2,] 1.3862944        NA       NA       NA       NA
[3,] 0.6931472 2.3025851 2.079442 2.079442 1.609438
[4,] 1.0986123        NA       NA       NA       NA
[5,] 1.0986123        NA       NA       NA       NA
[6,] 0.6931472 0.6931472       NA       NA       NA
```





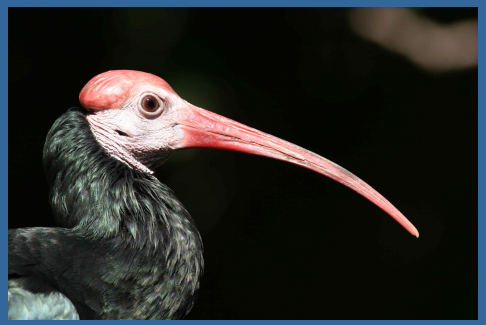
# Fitting the model

```
> m2 <- occu(~ lhours ~ MTCO + MTWA + AET.PET
                + Wet.Intensity, data = bi.umf)
> summary(m2)
Occupancy (logit-scale):
                 Estimate      SE       z  P(>|z|)
(Intercept)         0.115 2.22785  0.0516 9.59e-01
MTCO               -0.171 0.06559 -2.6127 8.98e-03
MTWA               -0.468 0.10183 -4.5946 4.34e-06
AET.PET            22.489 2.66362  8.4429 3.10e-17
Wet.Intensity      -0.064 0.00987 -6.4824 9.03e-11

Detection (logit-scale):
             Estimate     SE      z  P(>|z|)
(Intercept)     -1.28 0.0810 -15.83 1.96e-56
lhours           0.11 0.0662   1.66 9.69e-02
```

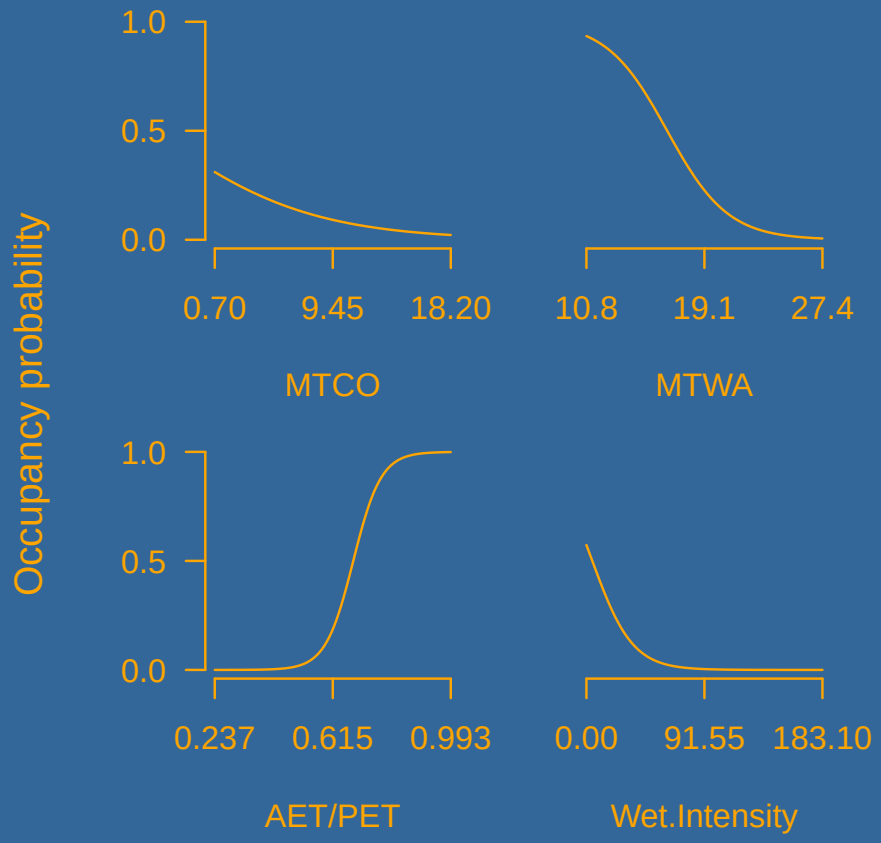


SEEC - Statistics in Ecology, Environment and Conservation

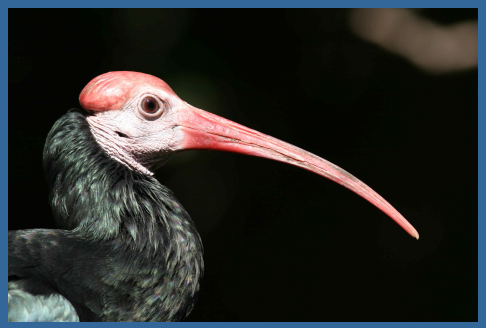
# Southern bals ibis



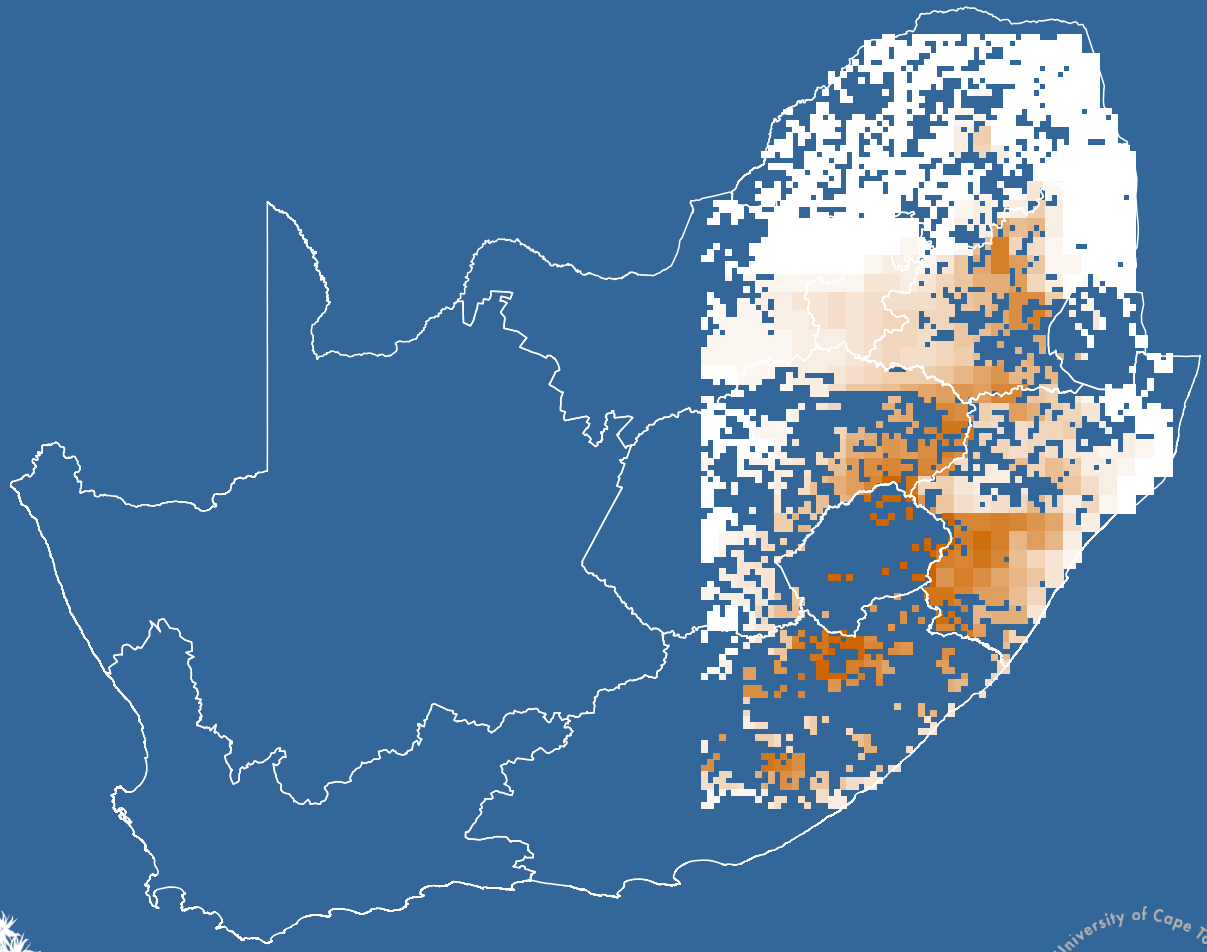
© Peter Ryan



# Southern bals ibis



© Peter Ryan

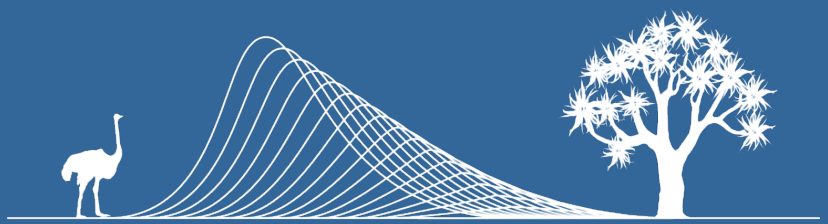


SEEC - Statistics in Ecology, Environment and Conservation

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad

# Single-season occupancy models

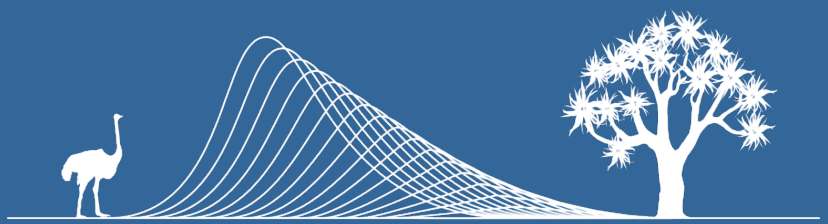
- Repeated detection / non-detection data
- Estimate occupancy and detection process
- **Key assumptions:** closure, no false detections, surveys are independent, sites are independent
- Can be fitted using function `occu()` in R package 'unmarked'
- Other software: PRESENCE, MARK, BUGS





# Extensions (for another day)

- Spatial autocorrelation
- Multi-season occupancy models
- Multi-state occupancy models
- Multi-species occupancy models
- Abundance models





# 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. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Book, Academic Press, Amsterdam.


The 'unmarked' package:


- Fiske, I. J., and R. B. Chandler. 2011. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. Journal of Statistical Software 43: 1-23.
- Fiske, I., and R. B. Chandler. 2012. Overview of Unmarked: An R Package for the Analysis of Data from Unmarked Animals. R Vignette.
- Kéry, M., and J. A. Royle. 2016. Applied Hierarchical Modeling in Ecology. Academic Press.