Line Transect data in Occupancy Studies: data collection methods and model identifiability

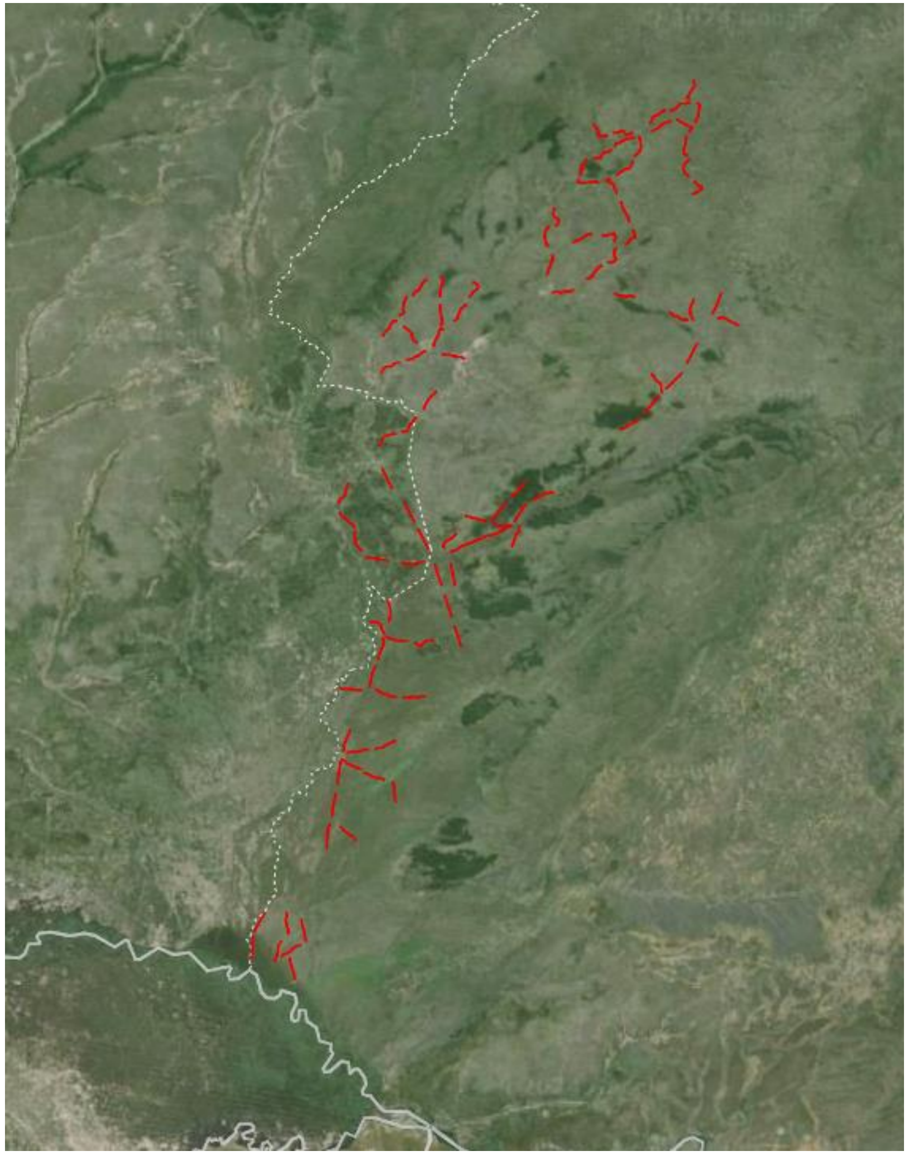
Milly Jones[1,3], Diana Cole[1], Eleni Matechou[1], Nicholas Deere[2], Robin Lines

[1] School of Maths, Statistics, Actuarial Sciences, University of Kent, Canterbury
[2] Durrell Institute of Conservation and Ecology, University of Kent, Canterbury

[3] mj23@kent.ac.uk

Motivating data set

[Lines2019]



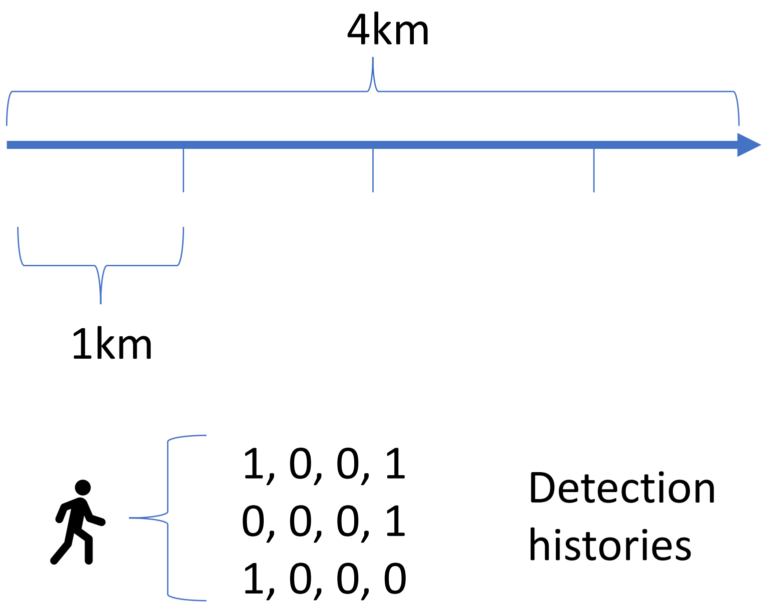
By Lencer (talk) - own work, used:File:Zambezi river basin-de.svg by Eric Gaba (User:Sting) and User:NNWhttp://www.kavangozambezi.org/the_map.php, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=16751696

Motivating data set

2015, large scale spoor detection and species sighting survey
102 line transects, each of length 4 km.

Each transect is visited 3 times over the course of 10 days.
Detection/non-detection histories taken.

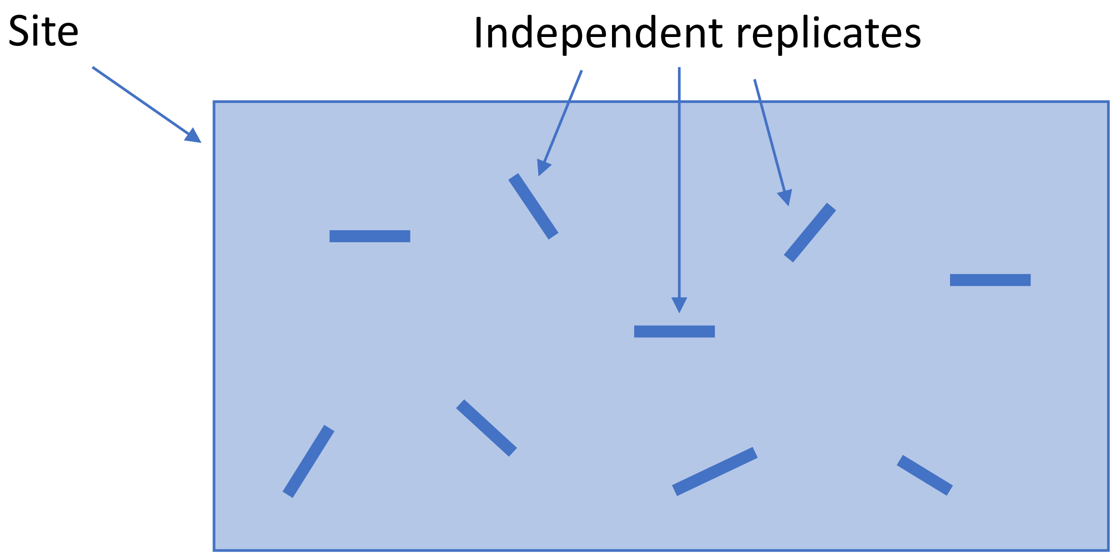
Detections: Leopard (279), Lion (87), Hyena (237), Wild dog (40).



Spatial Replicates

Occupancy studies require replication.

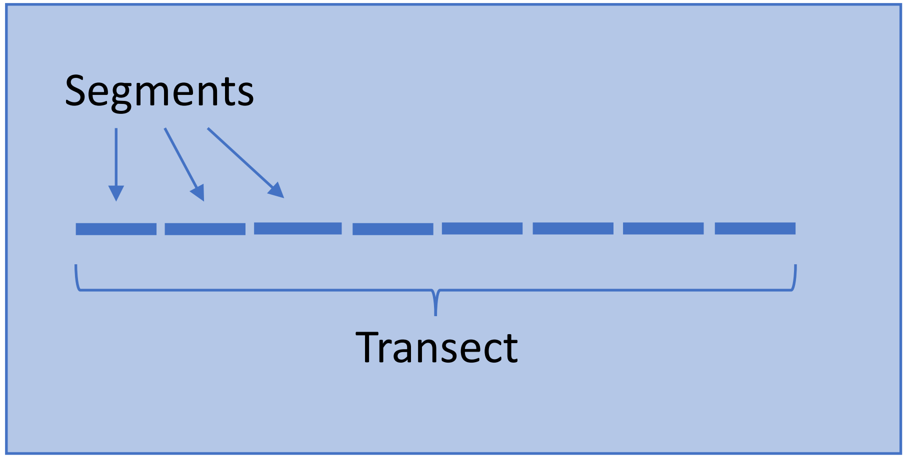
Replication can be spatial as well as temporal.



Random spatial replicates

Established transects at a site are convenient.

Consecutive segments are no longer independent [Kendall2009].



Consecutive segments

Markovian Segment Occupancy

The model: [Hines2010]

1. Site occupancy

$$z_i \sim \text{Bernoulli}(\psi)$$

Notation:

$\psi = \mathbb{P}(\text{site is occupied})$

z_i = occupancy site i

z_{ik} = occupancy segment k, site i

y_{ik} = detection segment k, site i

$\theta_{01} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment unoccupied})$

$\theta_{11} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment occupied})$

$p = \mathbb{P}(\text{detection} \mid \text{segment is occupied})$

Markovian Segment Occupancy

The model: [Hines2010]

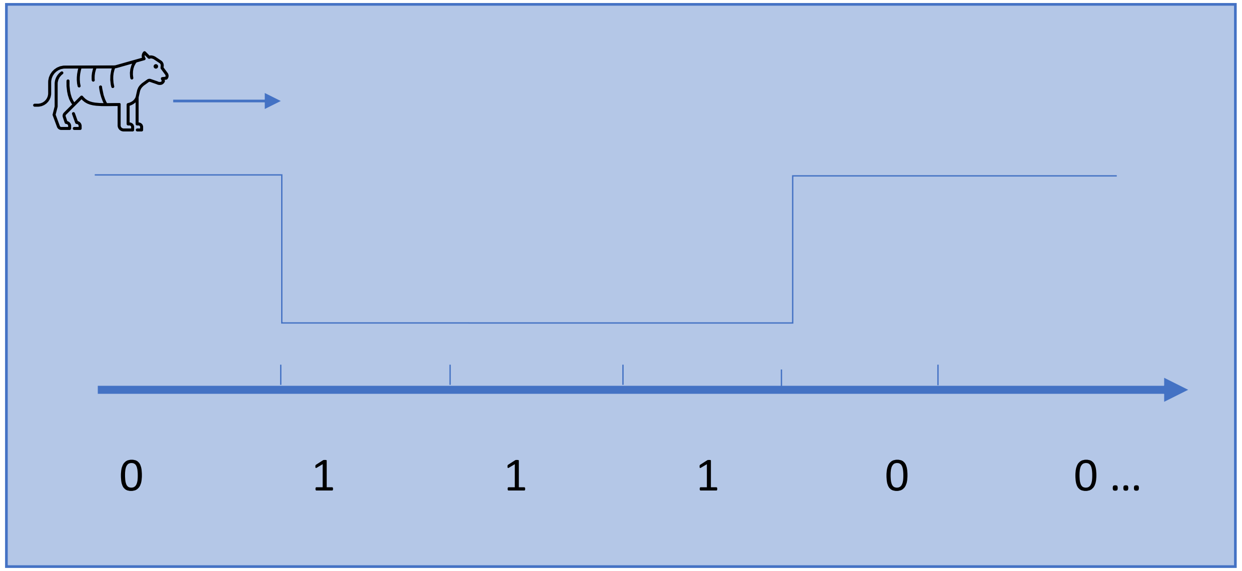
<u>1. Site occupancy</u>

$$z_i \sim \text{Bernoulli}(\psi)$$

<u>2. Segment occupancy</u>

$$z_{ik} | z_i \sim \text{Bernoulli}(z_i \times \theta_{ik})$$

$$\begin{bmatrix} 1-\theta_{01} & \theta_{01} \\ 1-\theta_{11} & \theta_{11} \end{bmatrix}$$



Notation:

z_i = occupancy site i

z_{ik} = occupancy segment k, site i

y_{ik} = detection segment k, site i

$\psi = \mathbb{P}(\text{site is occupied})$

$\theta_{01} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment unoccupied})$

$\theta_{11} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment occupied})$

$p = \mathbb{P}(\text{detection} \mid \text{segment is occupied})$

Markovian Segment Occupancy

The model: [Hines2010]

1. Site occupancy

$$z_i \sim \text{Bernoulli}(\psi)$$

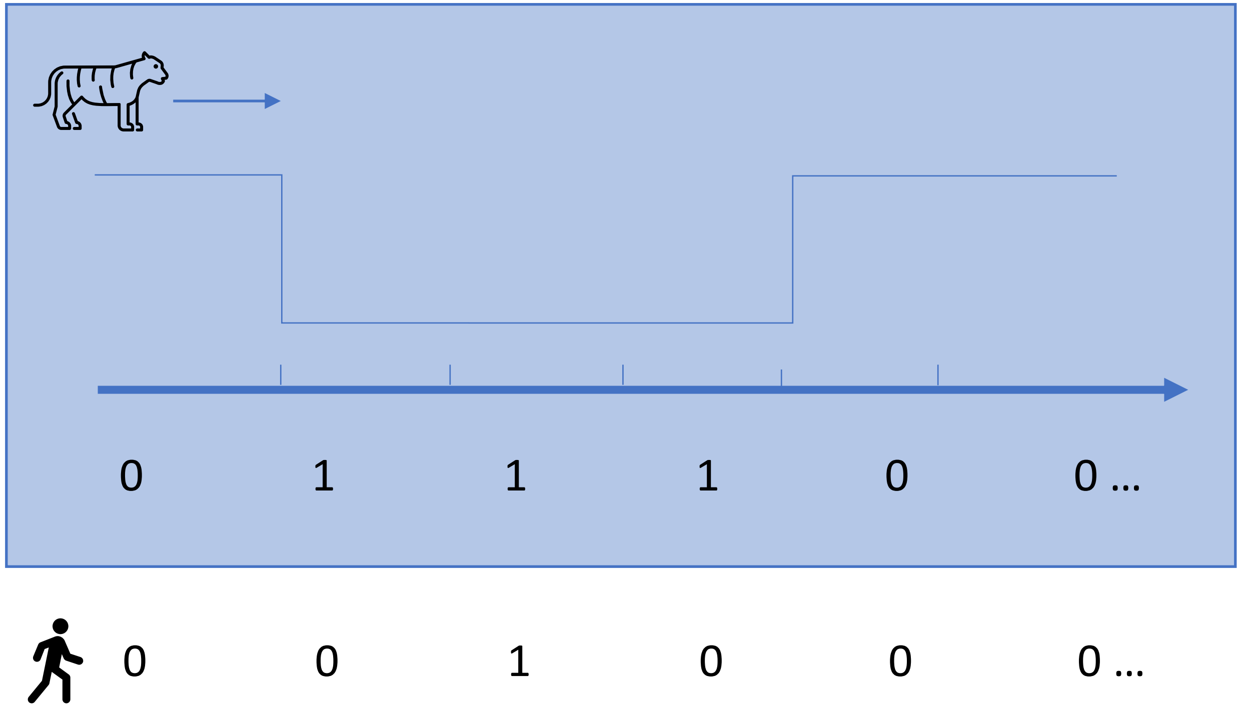
2. Segment occupancy

$$z_{ik} | z_i \sim \text{Bernoulli}(z_i \times \theta_{ik})$$

$$\begin{bmatrix} 1-\theta_{01} & \theta_{01} \\ 1-\theta_{11} & \theta_{11} \end{bmatrix}$$

3. Detection

$$y_{ik} | z_{ik} \sim \text{Bernoulli}(z_{ik} \times p)$$



Notation:

z_i = occupancy site i

z_{ik} = occupancy segment k, site i

y_{ik} = detection segment k, site i

$\psi = \mathbb{P}(\text{site is occupied})$

$\theta_{01} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment unoccupied})$

$\theta_{11} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment occupied})$

$p = \mathbb{P}(\text{detection} \mid \text{segment is occupied})$

Identifiability

Full Model

$$z_i \sim \text{Bernoulli}(\psi)$$
$$z_{ik}|z_i \sim \text{Bernoulli}(z_i \times \theta_{ik})$$
$$y_{ik}|z_{ik} \sim \text{Bernoulli}(z_{ik} \times p)$$

$$\theta_{01} = \theta_{11} = \theta$$

Collapses to

Non-identifiable Model

$$z_i \sim \text{Bernoulli}(\psi)$$
$$y_{ik}|z_{ik} \sim \text{Bernoulli}(z_i \times \theta \times p)$$

Can only identify: $(\psi, \theta p)$

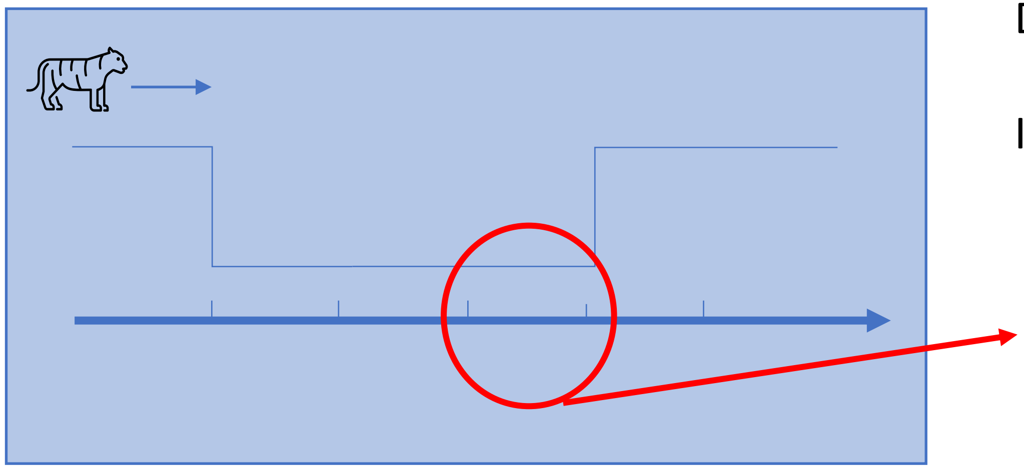
Good news! Can still identify ψ

Notation:

z_i = occupancy site i
z_{ik} = occupancy segment k, site i
y_{ik} = detection segment k, site i

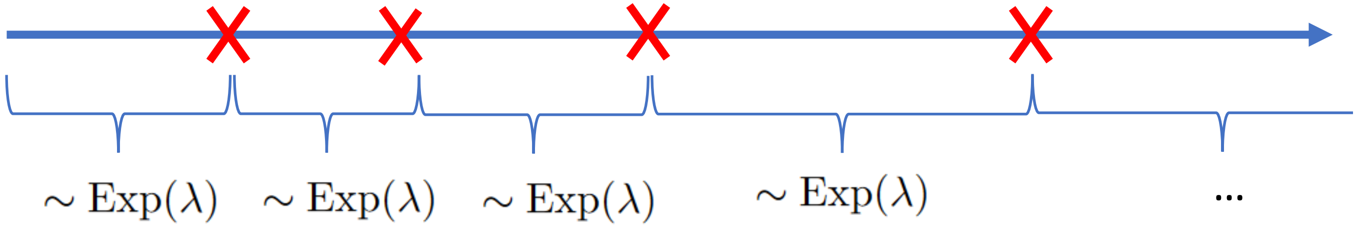
$\psi = \mathbb{P}(\text{site is occupied})$
$\theta_{01} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment unoccupied})$
$\theta_{11} = \mathbb{P}(\text{segment occupied} \mid \text{previous segment occupied})$
$p = \mathbb{P}(\text{detection} \mid \text{segment is occupied})$

How do we record data: [Pautrel2024]

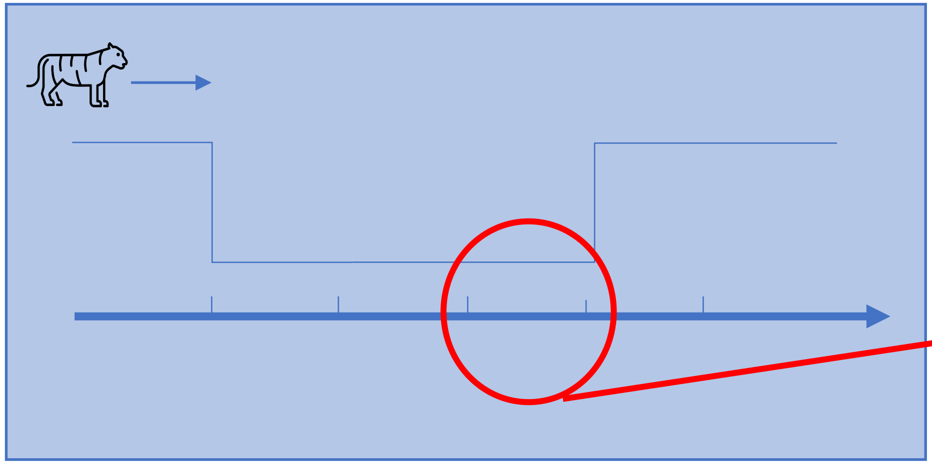


Detections are made at exponential rate λ on occupied segments.

If all segments have length R: $p = 1 - \exp(\lambda R)$

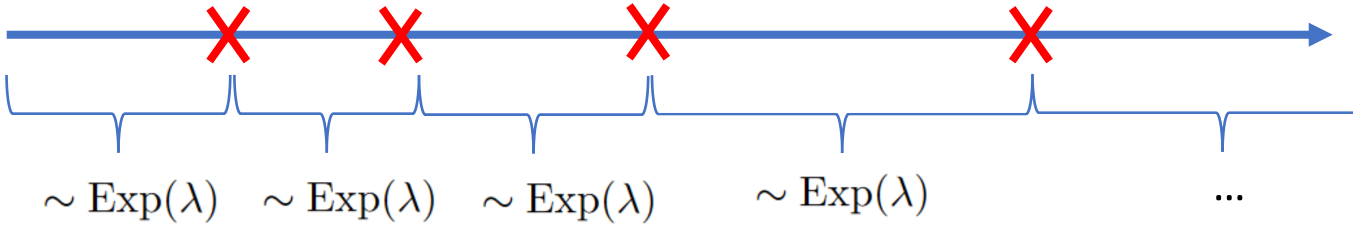


How do we record data: [Pautrel2024]



Detections are made at exponential rate λ on occupied segments.

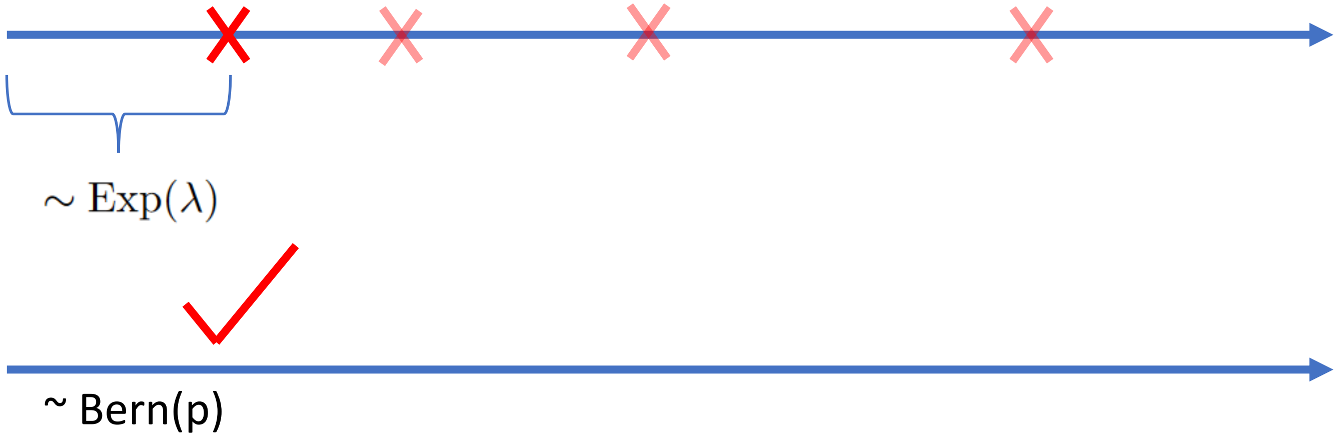
If all segments have length R: $p = 1 - \exp(\lambda R)$



Single Observation – Discrete vs Continuous

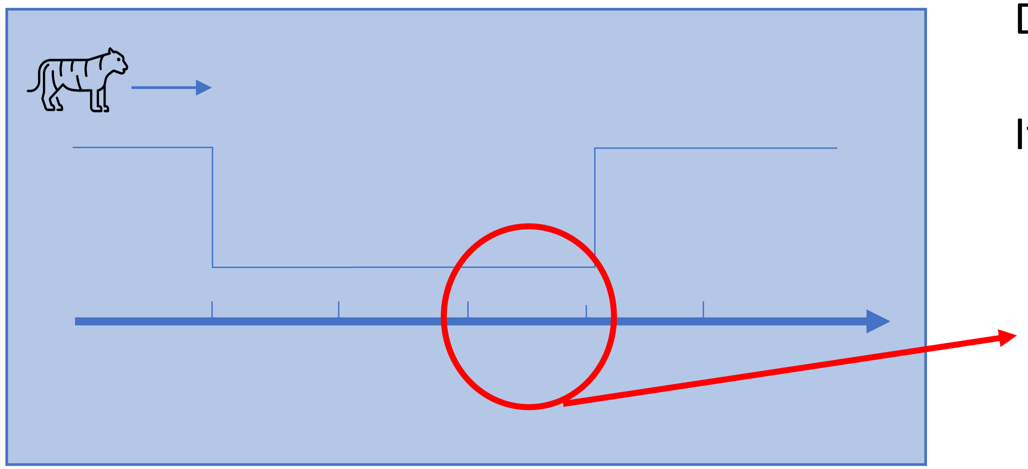
Two Observations

All observations



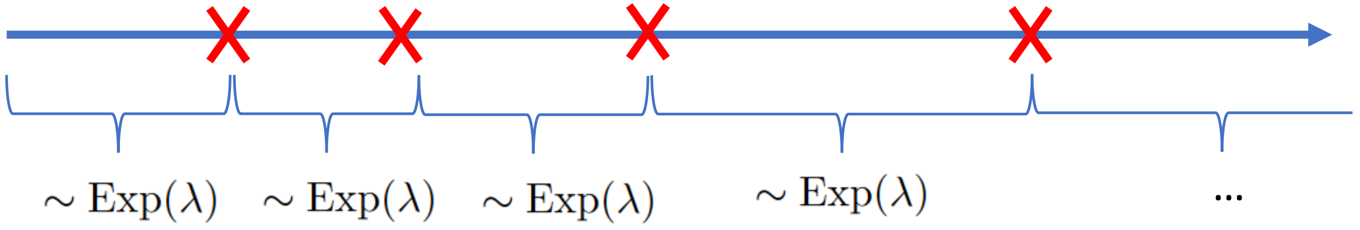
‘Time to first detection’ [Garrard2008]

How do we record data: [Pautrel2024]

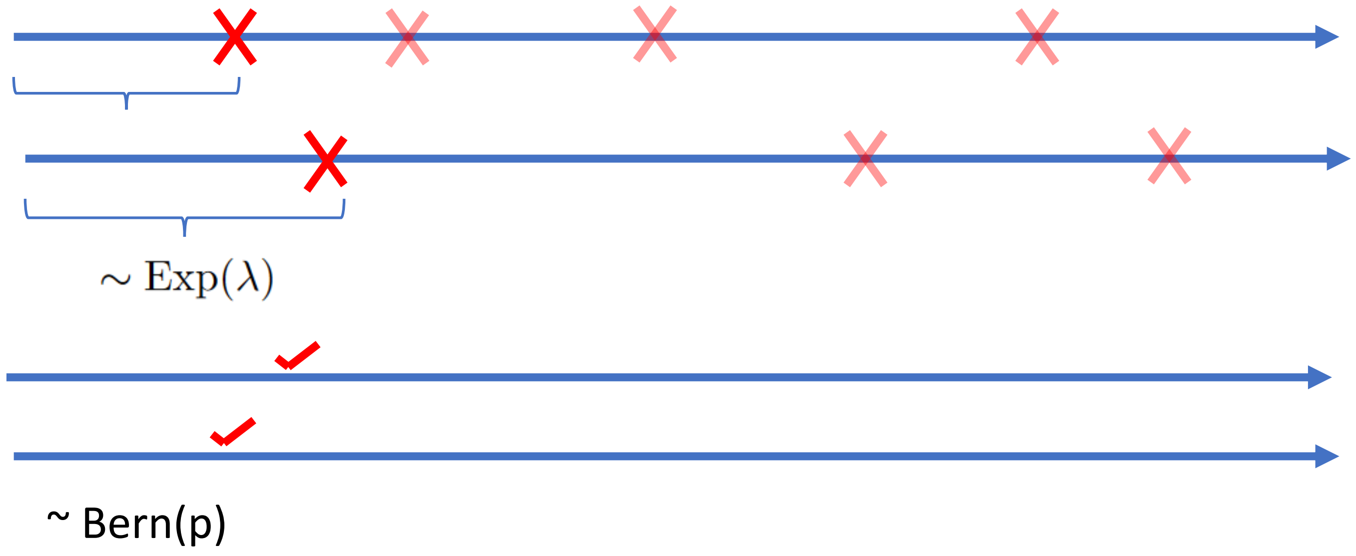


Detections are made at exponential rate λ on occupied segments.

If all segments have length R: $$p = 1 - \exp(\lambda R)$$

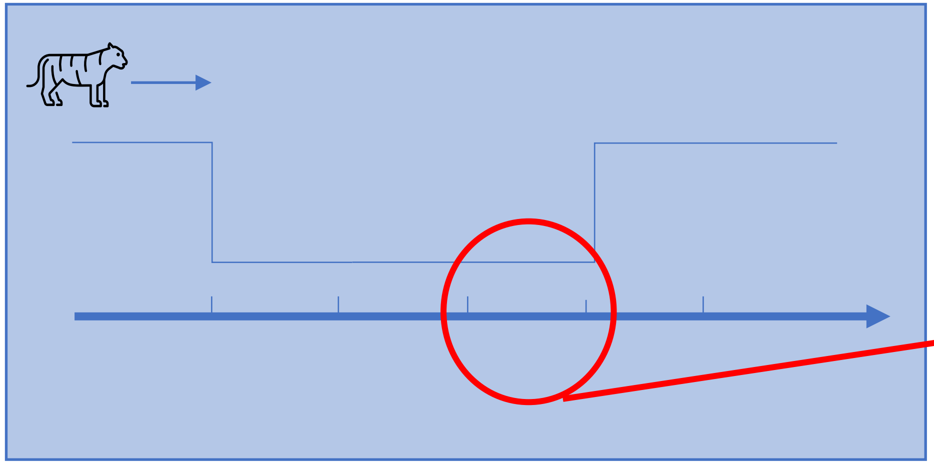


Single Observation | **Two Observations** – Discrete vs Continuous | All observations



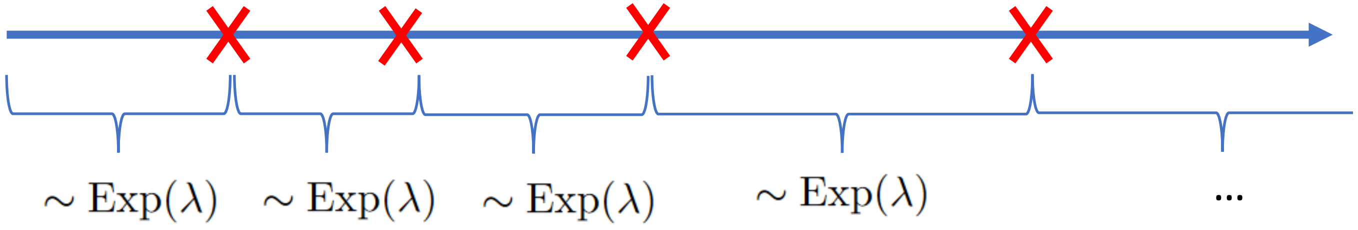
Independent surveys

How do we record data: [Pautrel2024]

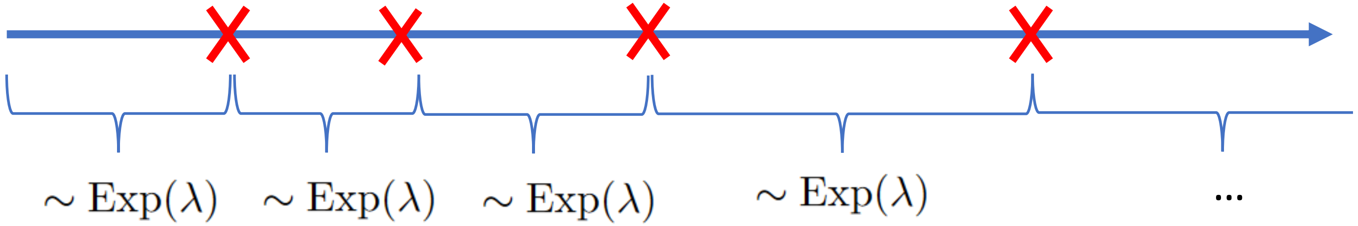


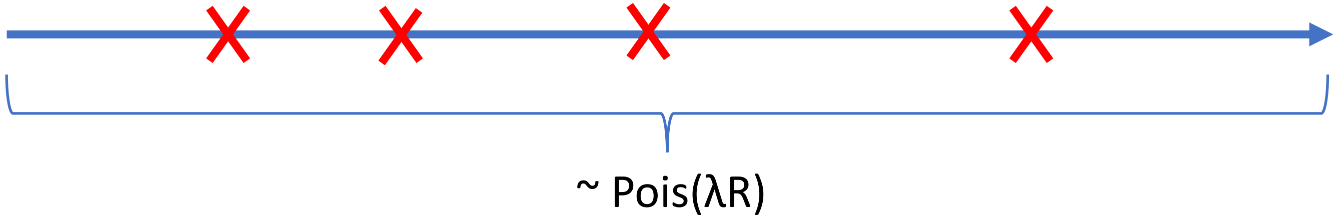
Detections are made at exponential rate λ on occupied segments.

If all segments have length R: $p = 1 - \exp(\lambda R)$



Single Observation | Two Observations | **All observations** – Discrete vs Continuous





Count of observations

Summary:

Single Observation

Discrete

Detection/non-detection

~ Bern(p)

1, 0, 0, 1, 1, …

Continuous

Distance to first detection

~Exp(λ)

x1, -, -, x4, x5, …

≥Two Observations

Discrete

Two independent detection\non-detection

~Bern(p)

1, 0, 0, 1, 1, …
0, 0, 1, 1, 0, …

Continuous

Two independent distance to first detection

~Exp(λ)

x1, -, -, x4, x5, …
-, -, x3, x4, -, …

All Observations

Discrete

Counts

~Pois(λR)

5, 0, 0, 7, 3, …

Continuous

All inter-detection distances

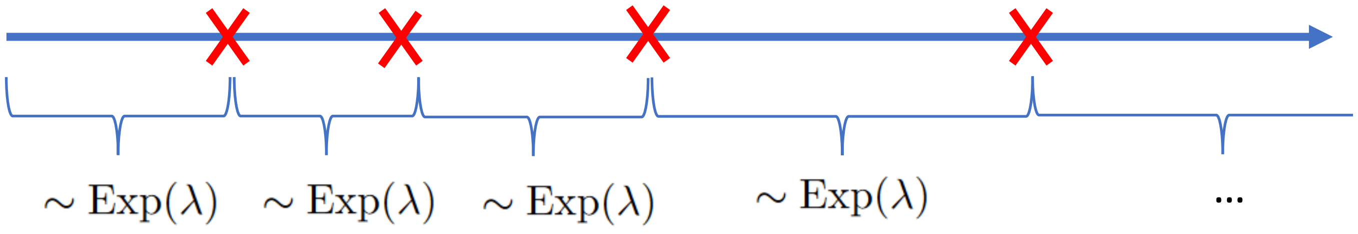
~Exp(λ)

x11, x12, x13, …
x21, x22, x23, …
…

Simulations overview

- We simulate:

1. Site occupancy status
2. Segment occupancy status
3. Continuous detections
4. Discretise as needed



200 sites, 20 segments per transect.

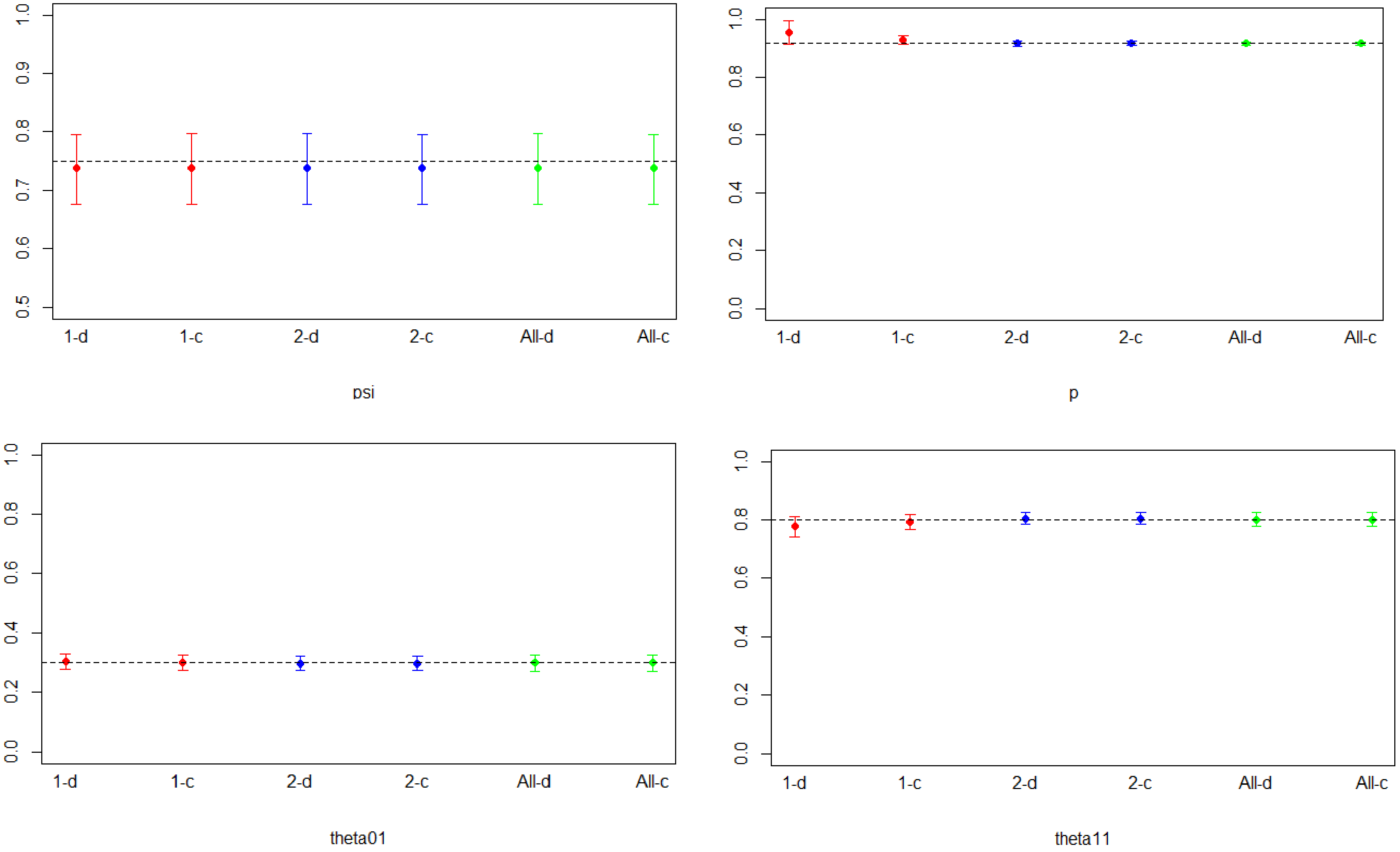
Compare 6 data collection methods.

Recall that: $p = 1 - \exp(\lambda R)$

All models fit in nimble.

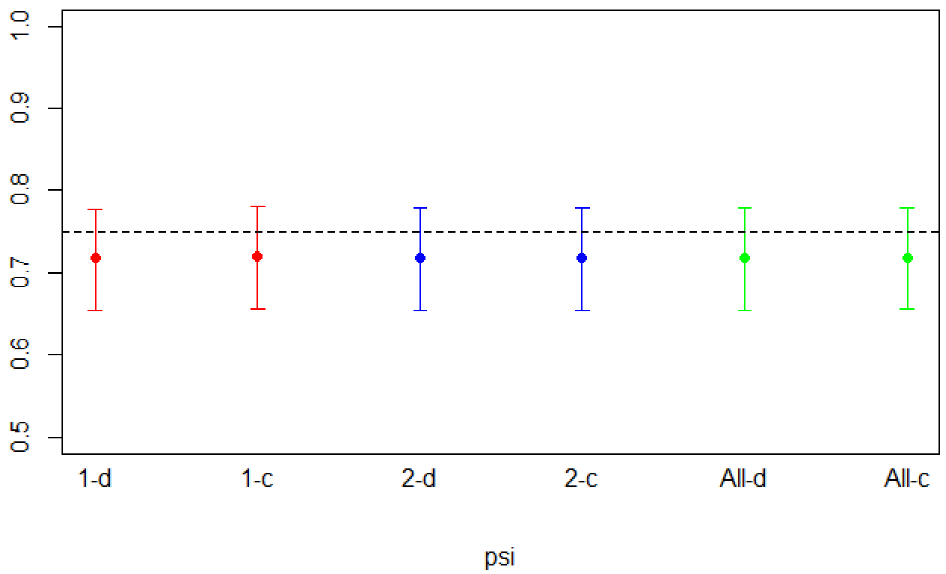
Simulations

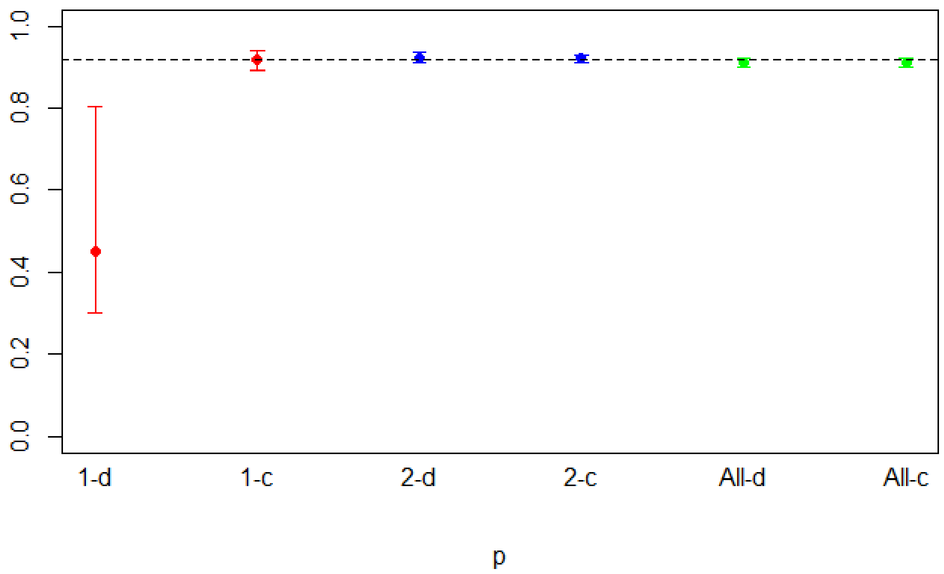
$\psi = 0.75$
$p = 0.918$
$\theta_{01} = 0.3$
$\theta_{11} = 0.8$

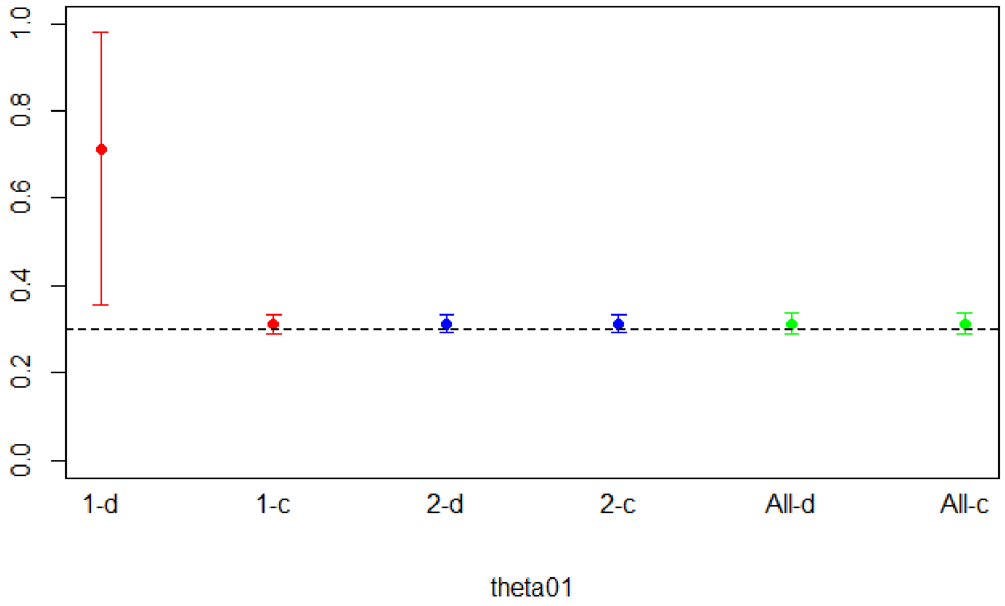


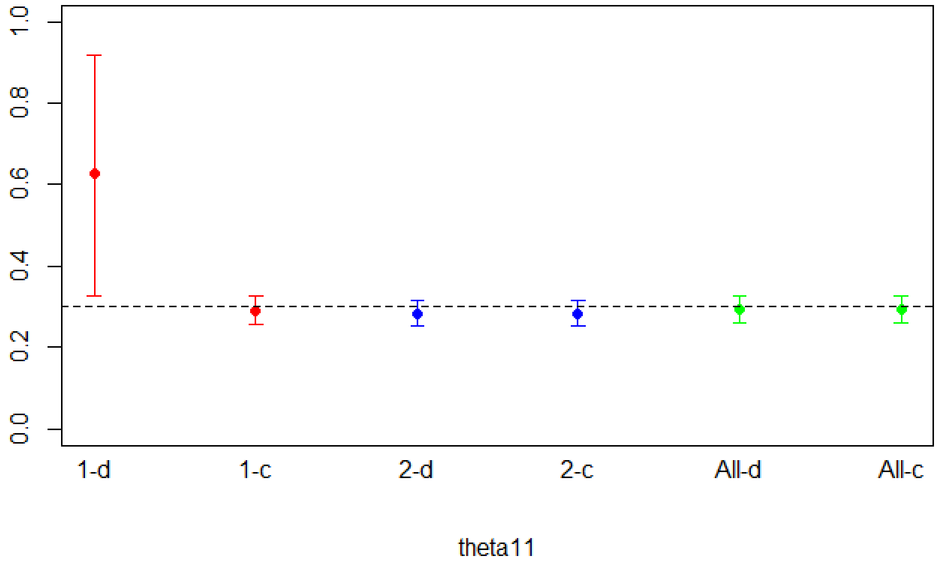
Simulations

$\psi = 0.75$
$p = 0.918$
$\theta_{01} = 0.3$
$\theta_{11} = 0.3$



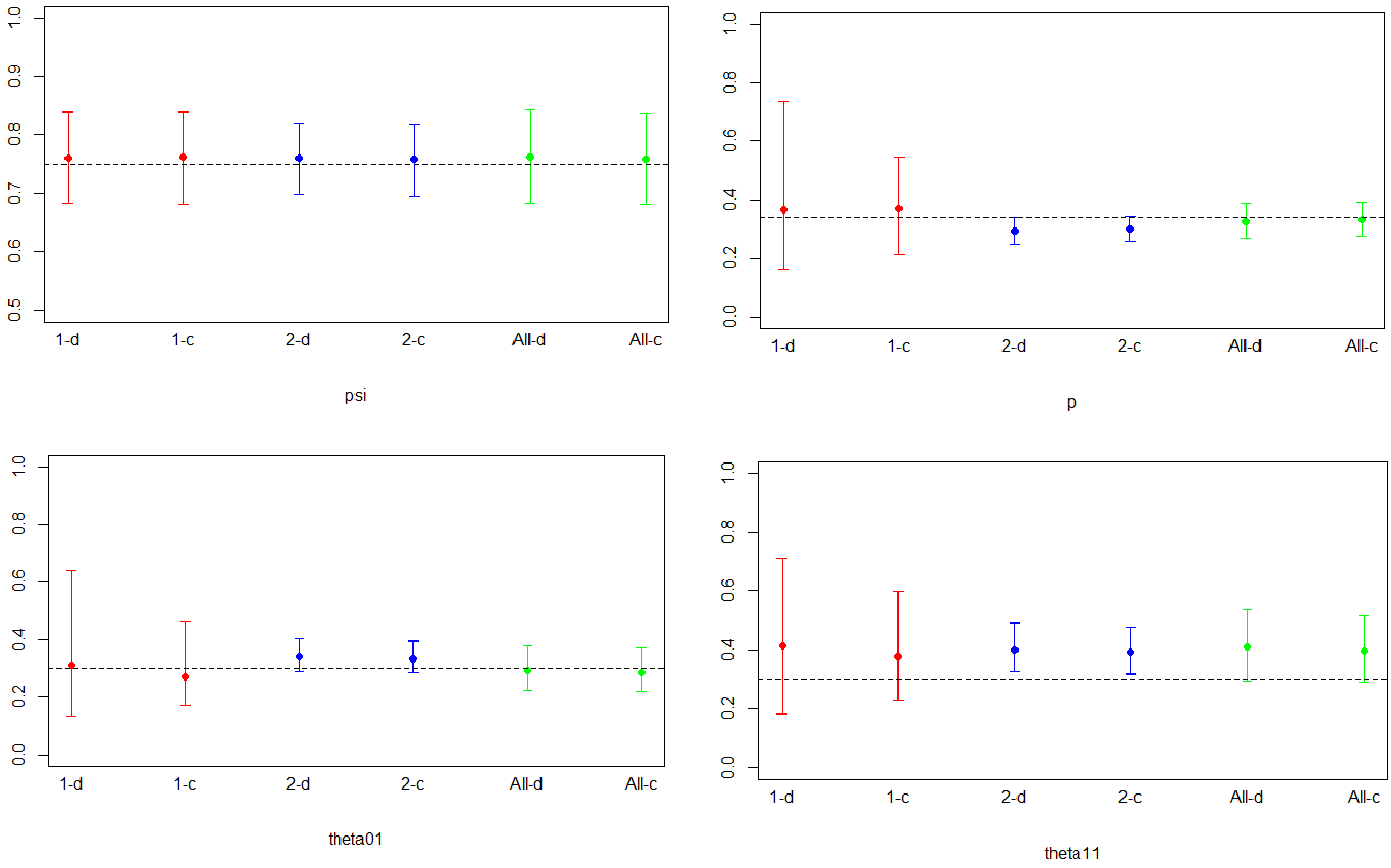


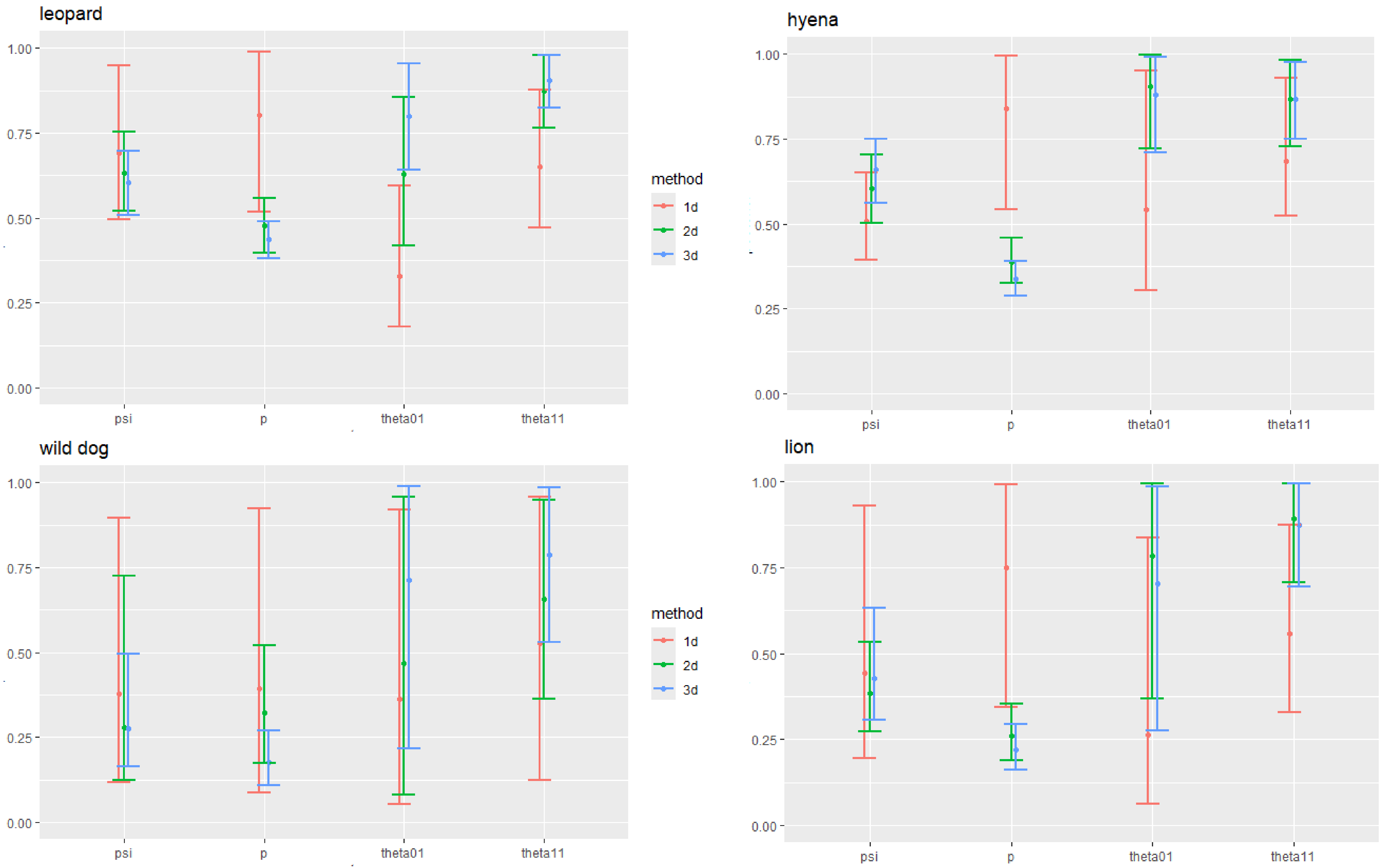




Simulations

$\psi = 0.75$
$p = 0.340$
$\theta_{01} = 0.3$
$\theta_{11} = 0.3$




leopard
hyena
wild dog
lion
1.00
0.75
0.50
0.25
0.00
psi
p
theta01
theta11
method
1d
2d
3d

Conclusion

- Three layer occupancy models using Markovian dependence in segment occupancy and a single detection are not identifiable when spatial correlation does not exist.

- We show that using a single time to detection, two or more independent observations, or counts make the model identifiable even when spatial correlation does not exist.

- Site occupancy probabilities perform well even other parameters struggle.

Where next

Discretising transects into segments is often arbitrary. Continuous models for handling spatial correlation include [Guru2011], involving a 2 state Markov Modulated Poisson Process (2-MMPP).

We have other line transect data collected for occupancy models that include perpendicular distances to observations (distance sampling data).

References

Lines, R. Tzanopoulos, J. MacMillan, D. (2019). Status of terrestrial mammals at the Kafue-Zambezi interface: implications for transboundary connectivity. *Oryx.*

Garrard, G. E. Bekessy, S. A. McCarthy, M. A. Wintle, B. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology.*

Kendall, W. L. White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology.*

Hines, J. E. Nichols, J. D. Royle, J. A. MacKenzie, D. I. Gopalaswamy, A. M. Samba Kumar, N. Karanth, K. U. (2010). Tigers on trails: occupancy modelling for cluster sampling. *Ecological Applications.*

Pautrel, L. Moulherat, S. Gimenez, O. Etienne, M. P. (2024) . Analysing biodiversity observation data collected in continuous time: Should we use discrete or continuous time occupancy models? *Methods in Ecology and Evolution.*

Guillera-Arroita, G. Morgan, B. J. T. Ridout, M. S. Linkie, M. (2011). Species Occupancy Modeling for Detection Data Collected Along a Transect. *Journal of Agriculture.*