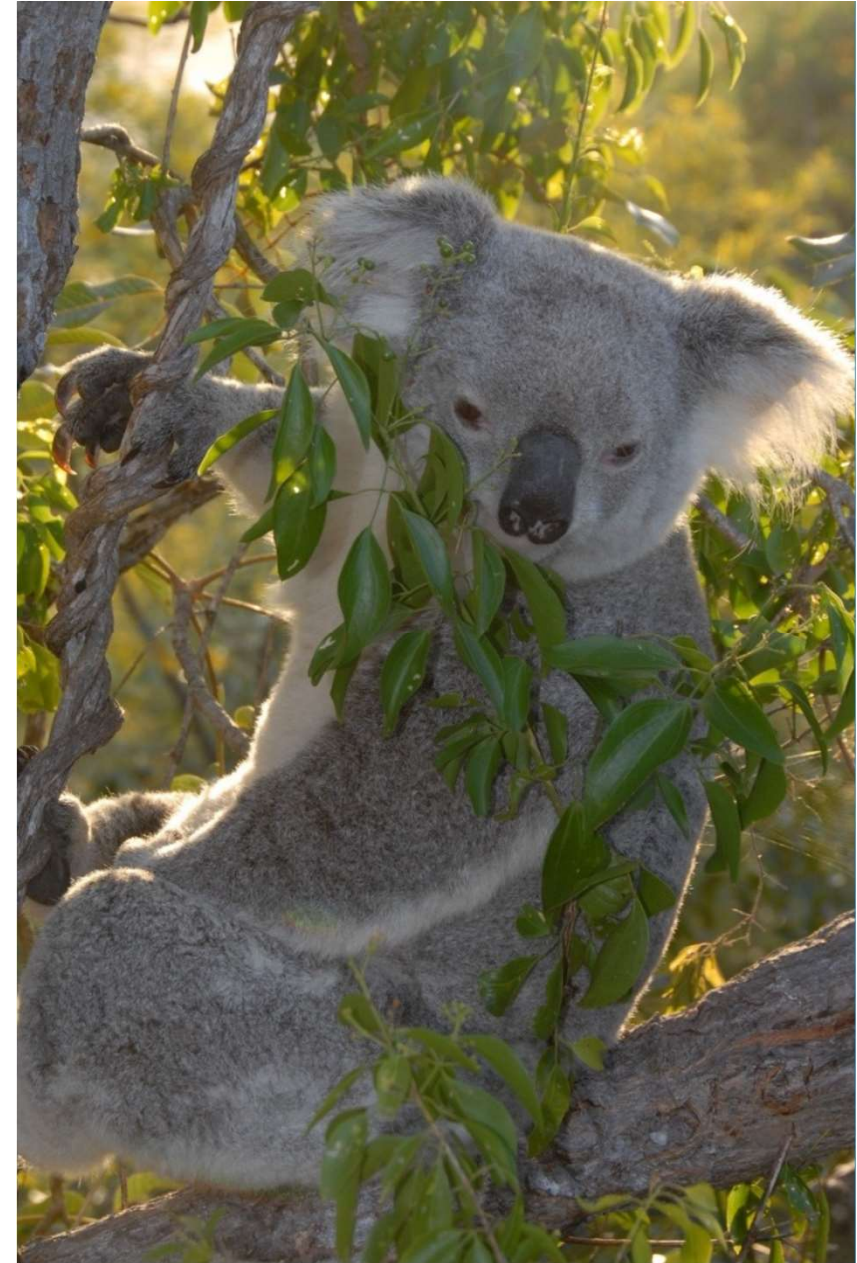




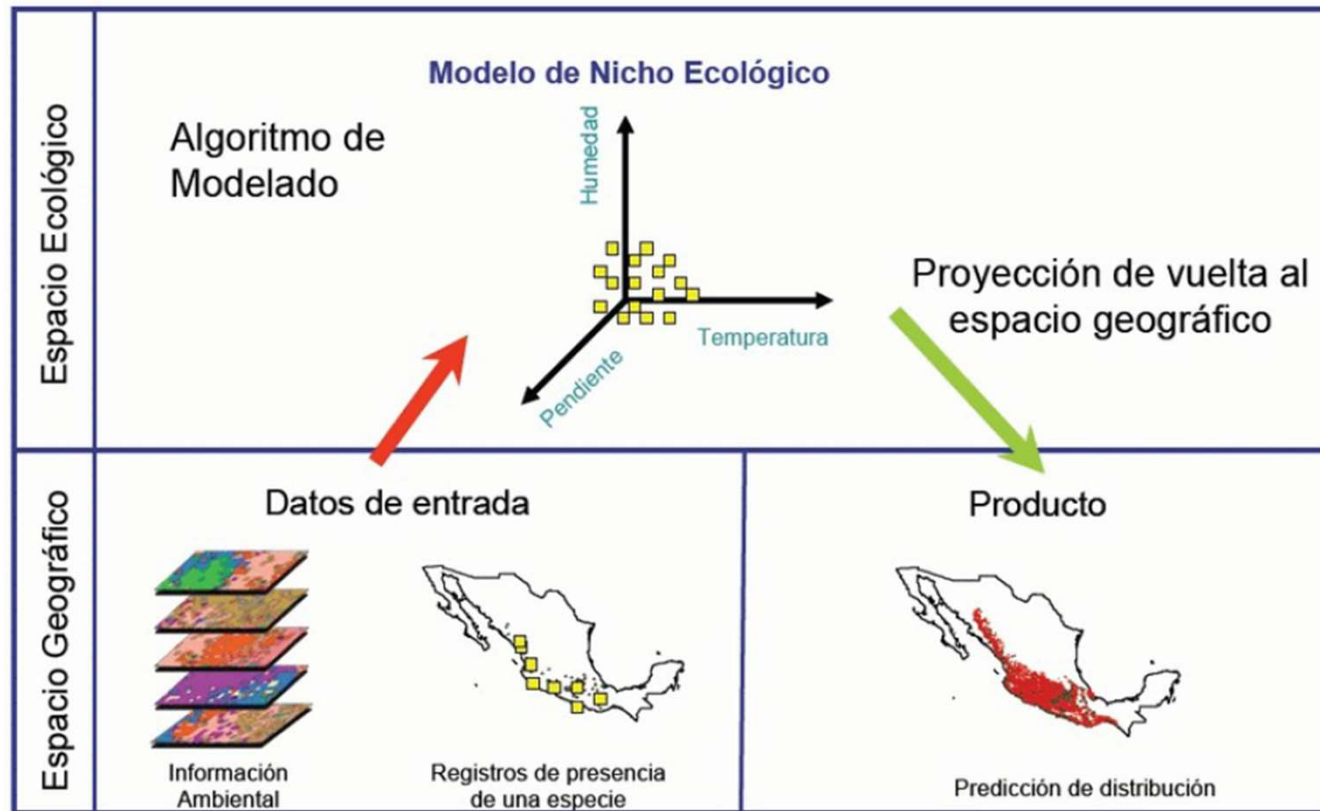
# Modelos de distribución de especies con integración de datos en RISDM (R)

David E. Uribe-Rivera  
[David.Uriberivera@csiro.au](mailto:David.Uriberivera@csiro.au)

Martes 20 de Mayo, 2025  
CSIRO Environment, Brisbane, Australia



# Modelos de distribución de especies (SDM)



Modificado de Peterson et al. 2011?





# Existen diversos tipos de SDMs

February 2022

PREDICTIVE PERFORMANCE OF SDMS

Article e01486; page 11

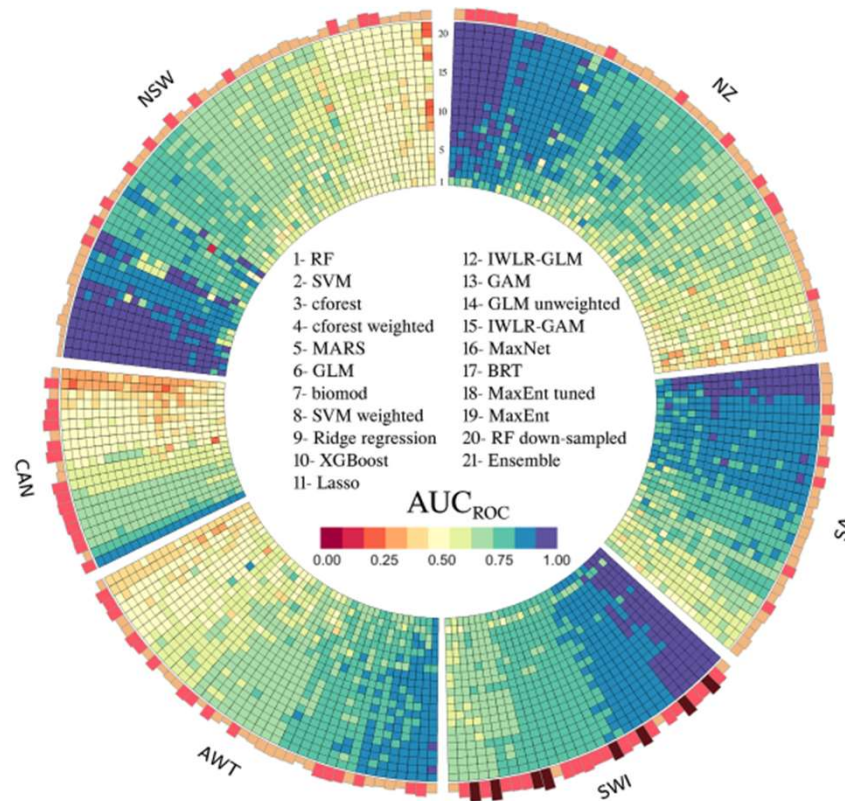
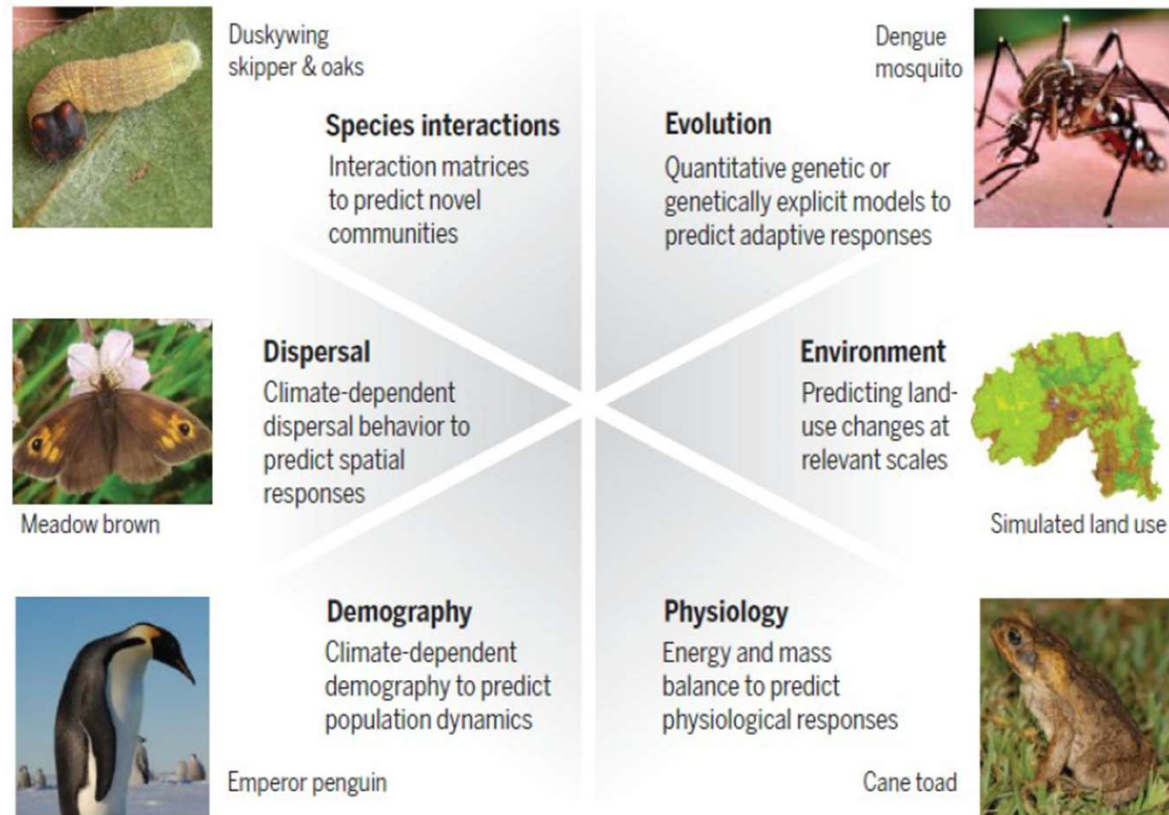


FIG. 2. AUC<sub>ROC</sub> values for all modeling methods across all species. Each circular track is the AUC<sub>ROC</sub> of a model for all species. Models and species are ordered by average AUC<sub>ROC</sub>. The outer text labels indicate the regions. The numbers between NSW and NZ regions label the modeling method as mentioned in middle of the figure. The height of the outer histogram shows the log<sub>10</sub>(number of species presence-only records) in the training data set. Histogram colors follow three categories: light orange, <100; pink, 100–1,000; and dark brown, >1,000 presences. This figure was created in Circos software (Krzywinski et al. 2009).



# Factores ecológicos de la distribución

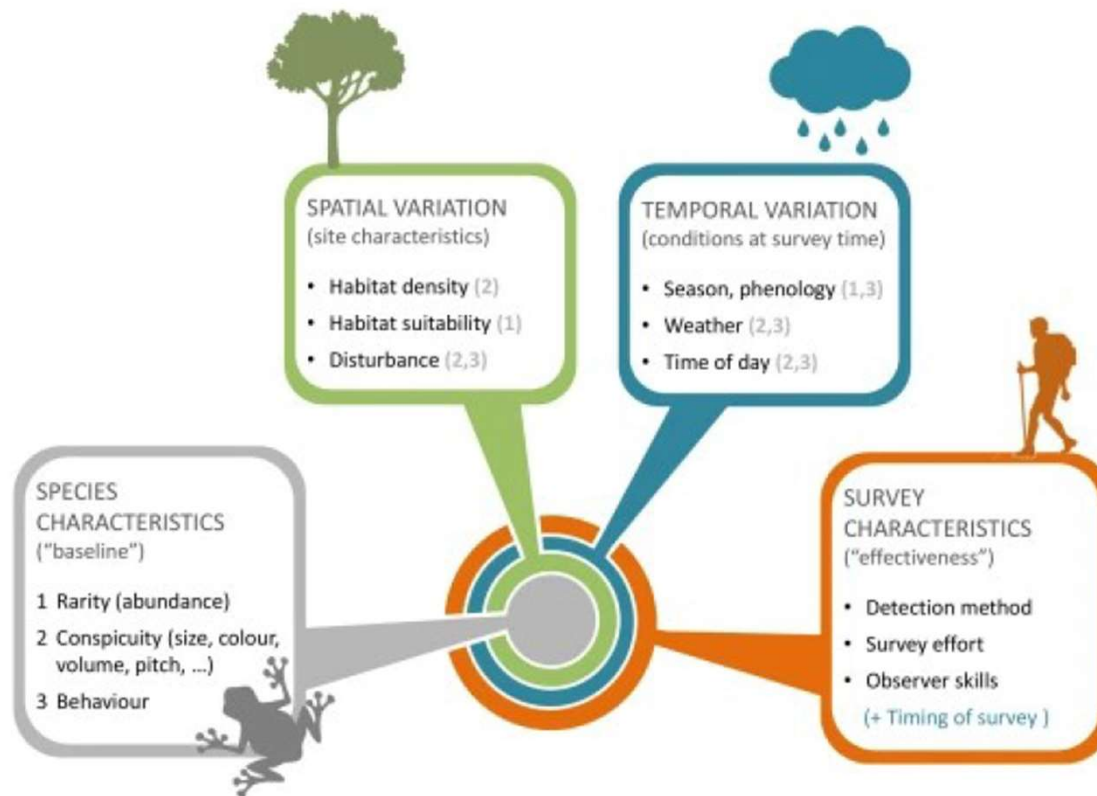


Urban et al 2016. *Science*, 353





# Procesos de observación

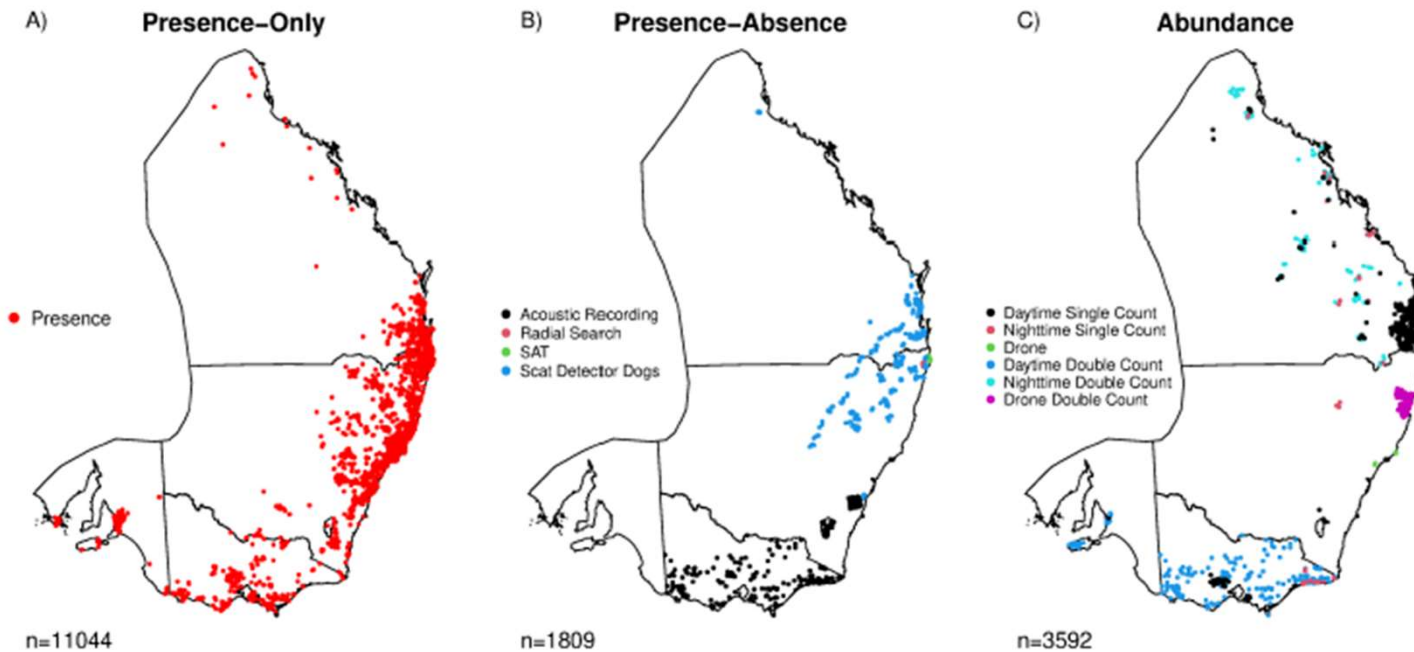


Guillera-Arroita 2017



# Y varios tipos de datos para entrenar SDMs

Mismo patrón de puntos representado de distinta manera

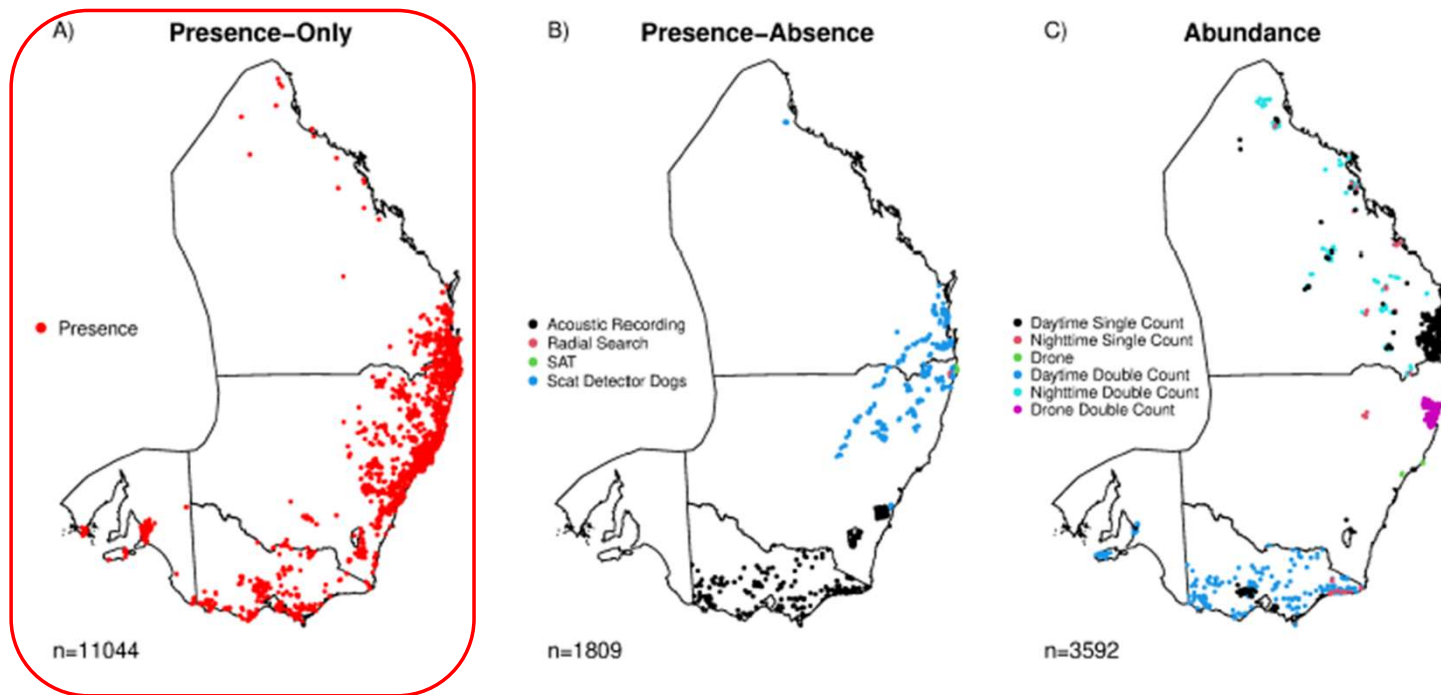


Menor a mayor contenido de información



# Y varios tipos de datos para entrenar SDMs

Mismo patrón de puntos representado de distinta manera



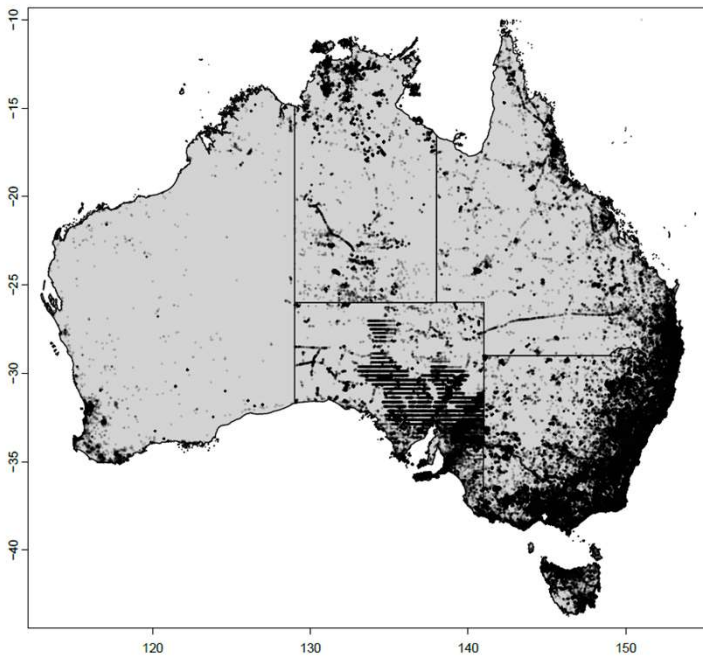
Menor a mayor contenido de información



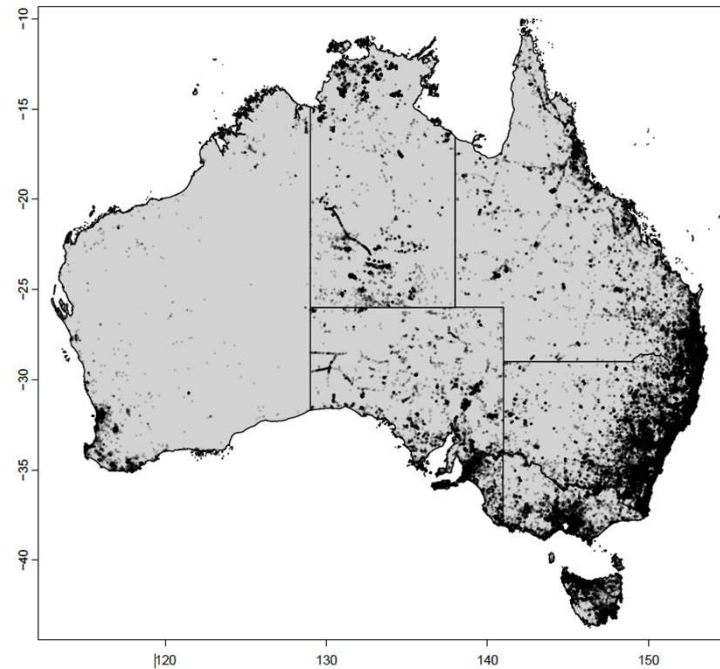


# Integración de datos basada en modelos

Registros de marsupiales en Atlas of Living Australia (2000-2024)



Todos los registros ( $n \sim 2,3M$ )



Solo datos oportunistas ( $n = 585.479$ )



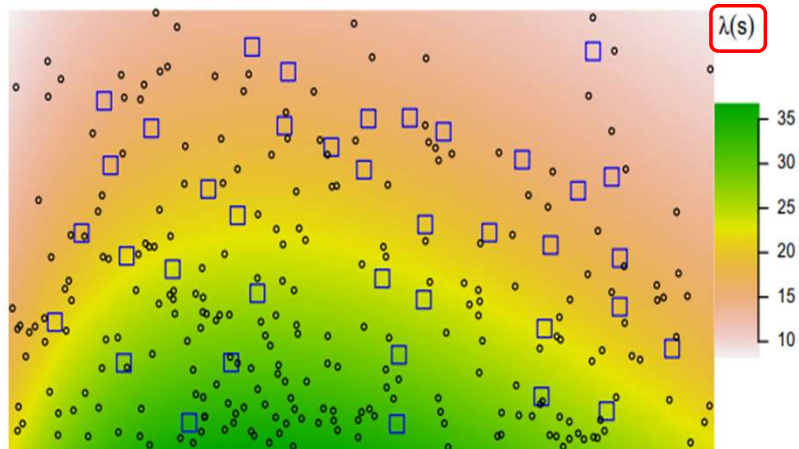


# Integración de datos basada en modelos

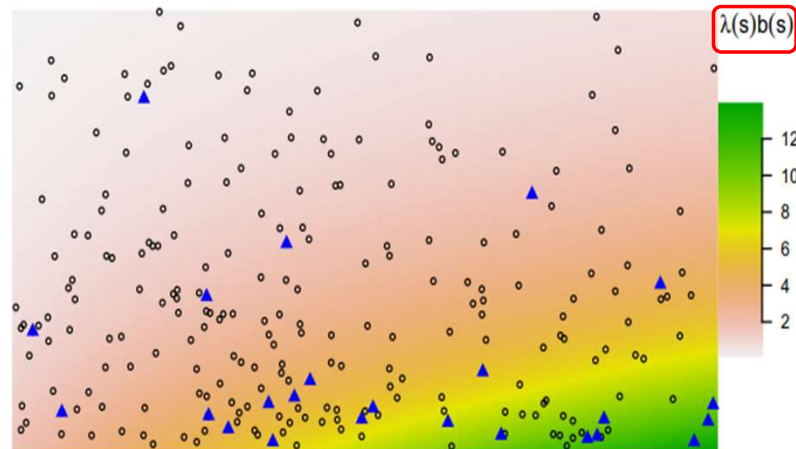
## Bias correction in species distribution models: pooling survey and collection data for multiple species

William Fithian<sup>1\*</sup>, Jane Elith<sup>2</sup>, Trevor Hastie<sup>1</sup> and David A. Keith<sup>3</sup>

Presence-Absence sampling



Biased presence-Only sampling



# Integración de datos disparatados

Sampling artifacts to acknowledge differences across systematic surveys





# El paquete RISDM

## ECOGRAPHY

### *Software note*

#### **‘RISDM’: species distribution modelling from multiple data sources in R**

Scott D. Foster<sup>1</sup>✉, David Peel<sup>1</sup>, Geoffrey R. Hosack<sup>1</sup>, Andrew Hoskins<sup>2</sup>, David J. Mitchell<sup>3</sup>, Kirstin Proft<sup>3</sup>, Wen-Hsi Yang<sup>4</sup>, David E Uribe-Rivera<sup>5</sup> and Jens G. Froese<sup>6</sup>

<sup>1</sup>Data61, CSIRO, Hobart, TAS, Australia

<sup>2</sup>CSIRO Environment, Townsville, QLD, Australia

<sup>3</sup>Australian Bureau of Agricultural and Resource Economics and Sciences, Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia.

<sup>4</sup>Data61, CSIRO, Brisbane, QLD, Australia

<sup>5</sup>CSIRO Environment, Brisbane, QLD, Australia

<sup>6</sup>CSIRO Health and Biosecurity, Brisbane, QLD, Australia

Correspondence: Scott D. Foster ([scott.foster@data61.csiro.au](mailto:scott.foster@data61.csiro.au))



# El paquete RISDM

Es un wrapper para ayudar a implementar modelos geoestadísticos complejos usando la aproximación de Laplace integrada y anidada (INLA)

INLA permite ajustar modelos de manera computacionalmente eficiente, pero es difícil de codificar en R

RISDM depende de R-INLA y terra

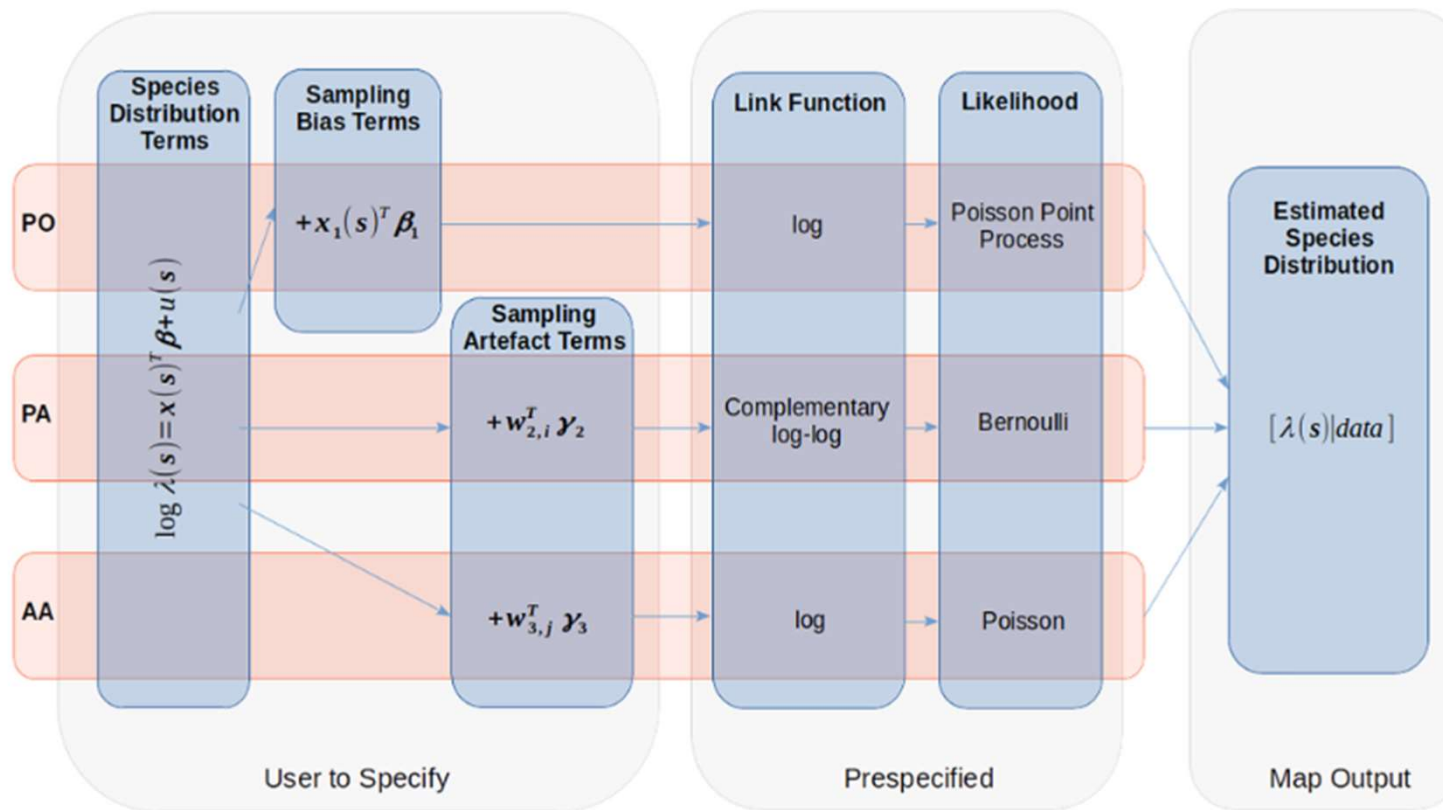
Permite ajustar modelos con efectos aleatorios espaciales y la integración de tres tipos de datos de distribución de especies:

- Solo-presencia (PO)
- Presencia-ausencia (PA)
- Abundancia-ausencia (AA)





# El integrated-SDM en RISDM

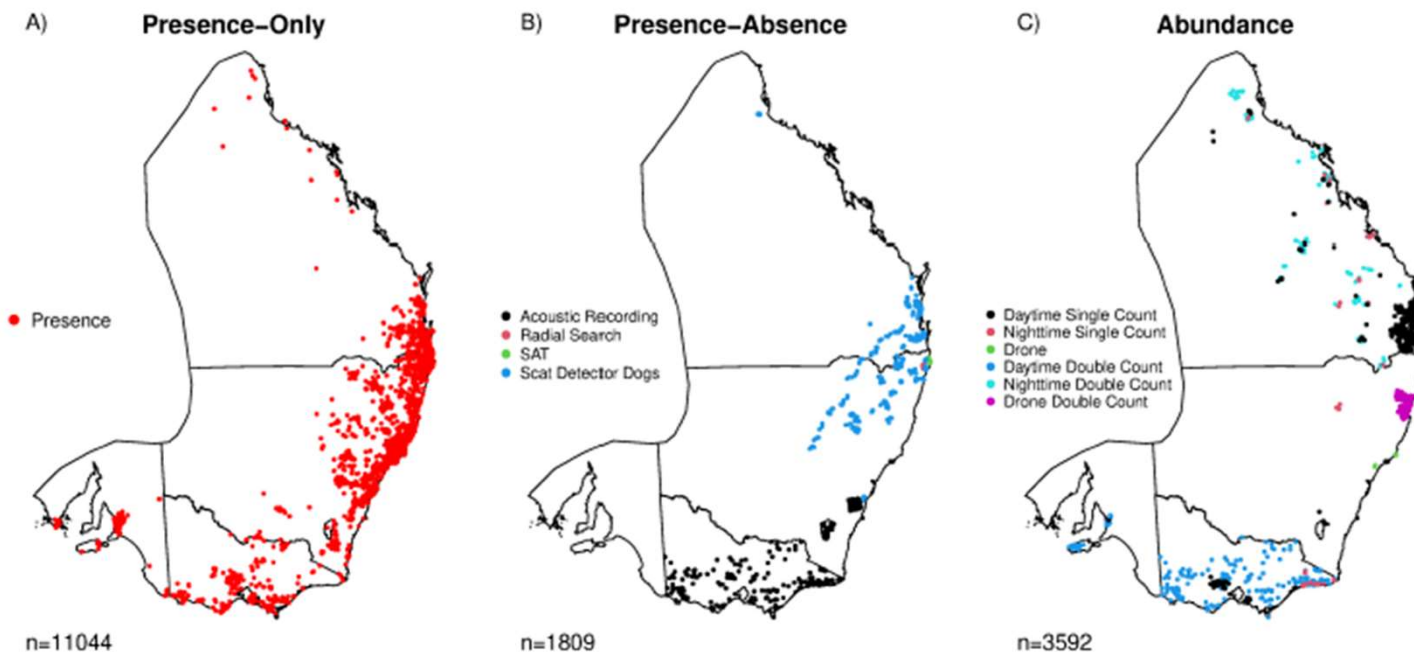


Foster et al (2024) - Ecography



# Integrando varios tipos de datos: ISDMs

Mismo patrón de puntos representado de distinta manera



Menor a mayor contenido de información





# El paquete RISDM

Con este paquete se pueden implementar modelos que logran tomar en cuenta explícitamente las siguientes fuentes **extra** de variabilidad:

1. Sesgos espaciales en el muestreo de los datos PO, incluyendo la posibilidad de agregar covariables que describan la variabilidad del esfuerzo de muestreo de los datos oportunistas (como ciencia ciudadana).
2. Variaciones en la detectabilidad entre distintas campanas de muestreo, ya sea entre datos de PA o de AA. Estos son expresados como diferencias en la intensidad promedio (como interceptos extra).
3. Variabilidad atribuible a otros factores, como por ejemplo a la autocorrelación espacial, legados históricos o simplemente variables no disponibles para la modelación.



# Cuando usar RISDM?

## 1. Cuando quieres corregir el sesgo de muestreo espacial:

Tienes datos de PO abundantes y sospechas están sesgados ya que una gran proporción de ellos vienen de ciencia ciudadana o su distribución espacial se correlaciona con la presencia humana.

Requiere al menos un set de datos de PA o AA

## 2. Cuando tienes datos más de una base de datos sistemáticos (PA y/o AA) que son complementarios en cobertura espacial/ambiental.

Puedes mejorar un modelo de una sola fuente de datos al integrar en un solo modelo varios conjuntos de datos de diferentes muestreos usando modelos de distribución con integración de datos (ISDMs).





# Juguemos con RISDM

Código R para reproducir el tutorial disponible en:

[https://github.com/duriber/tutorial\\_RISDM](https://github.com/duriber/tutorial_RISDM)





# Gracias por escuchar!

David E. Uribe-Rivera  
Postdoctoral research fellow  
CSIRO Environment, Brisbane.  
[David.Uriberivera@csiro.au](mailto:David.Uriberivera@csiro.au)

Australia's National Science Agency

