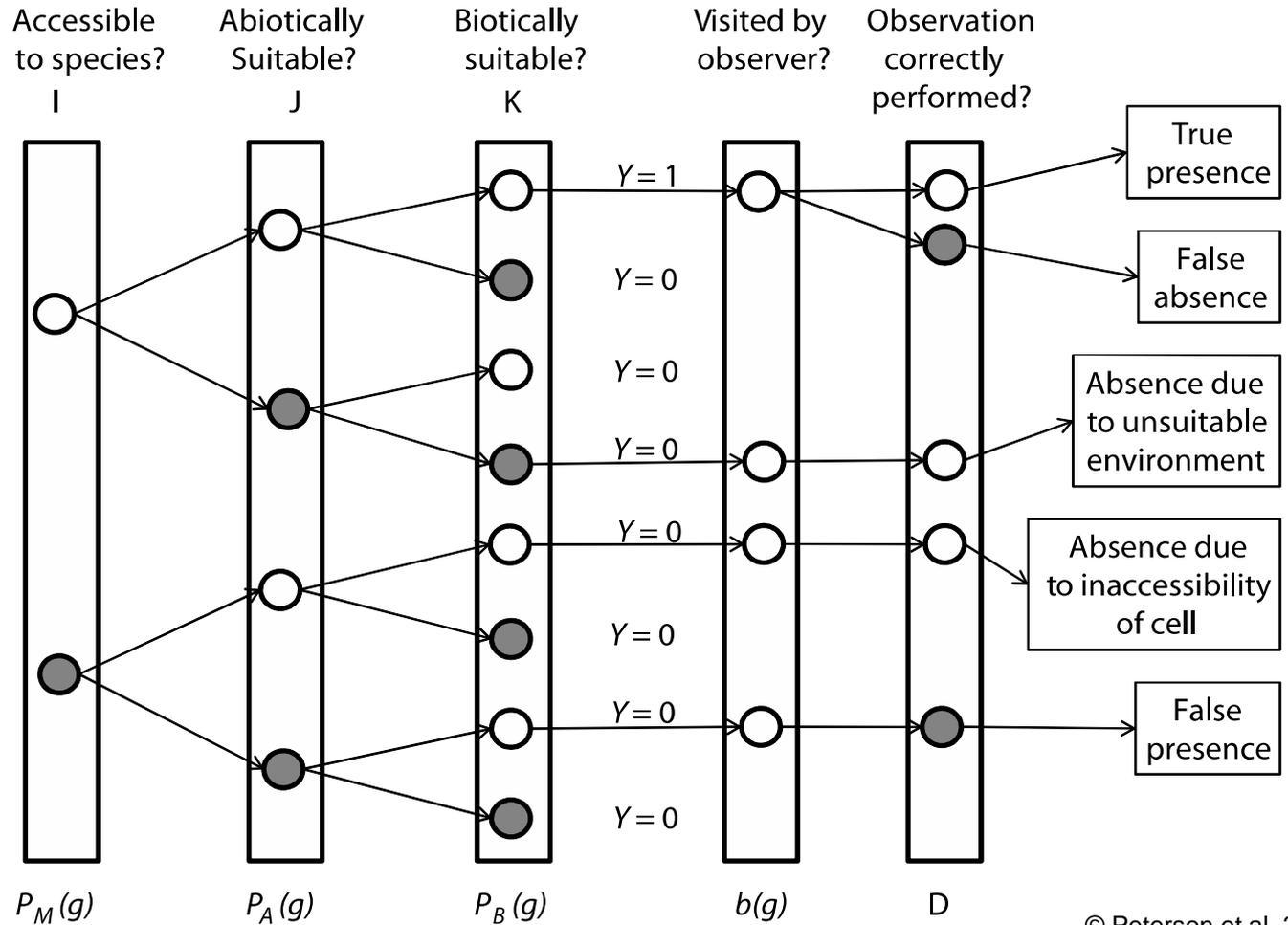
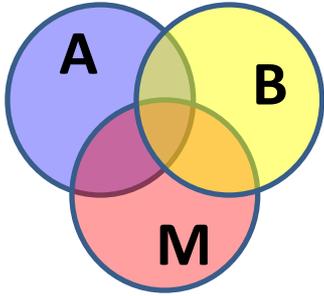


Andrés Lira-Noriega

Universidad de Kansas

# Los datos biológicos y sus sesgos



# Bases del funcionamiento de algoritmos mecánicos y orientados a procesos

# Organización

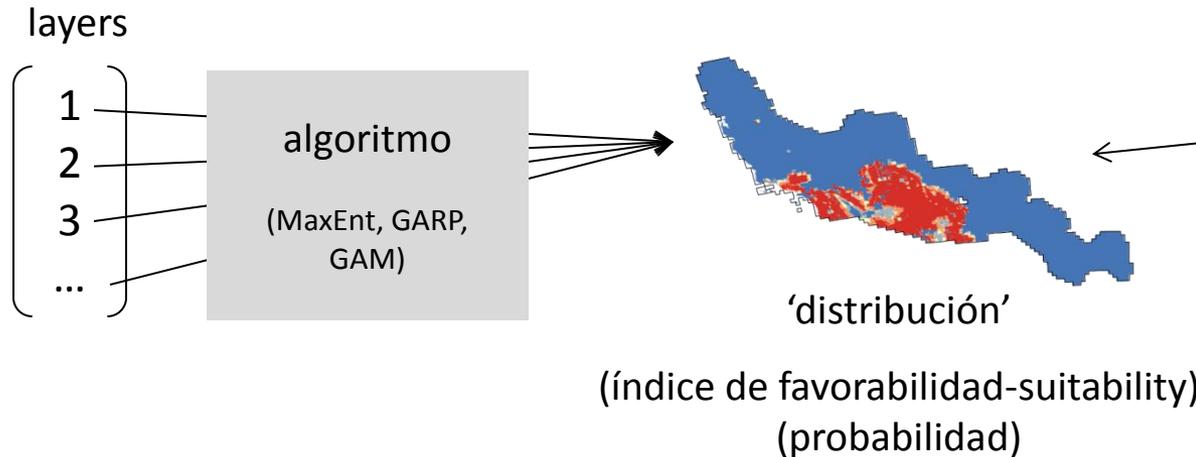
- ¿Por qué usar modelos mecanísticos?
- Diferencia entre modelos correlativos, mecanísticos y orientados a procesos
- Modelos mecanísticos
- Modelos orientado a proceso
  - Marco conceptual: Nicho abiótico y dispersión del muérdago del desierto
- Ventajas y desventajas de modelos correlativos y mecanísticos

# ¿Por qué usar modelos mecanísticos?

- Explicar la abundancia y distribución más que predecirlas (correlativo).
- Incorporar **procesos o mecanismos explícitamente** que limitan el rango de distribución.
- Entender si la respuesta a una variable es debido al **efecto directo causal** de una variable (o interacción), o una respuesta directa a una variable colineal ausente del modelo.
- Mejor aproximación al **nicho fundamental ( $N_F$ )**.
- Zonas de recambio en respuesta al ambiente. Invasiones, traslocaciones, cambio climático, cambios evolutivos.
- Integración: SDM/ENM basados en principios fisiológicos.
- ¡Necesario conocer la **historia natural** y la **fisiología** de la especie!

# Modelos correlativos y mecanísticos

## Correlativos



## Mecanísticos y orientados a procesos

$$p_j = \frac{c_j}{c_j + e_j}$$

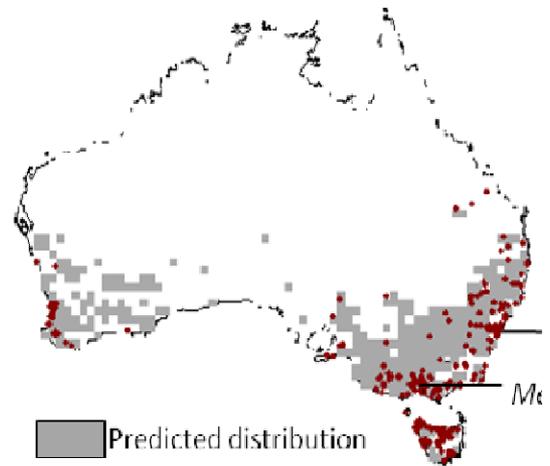
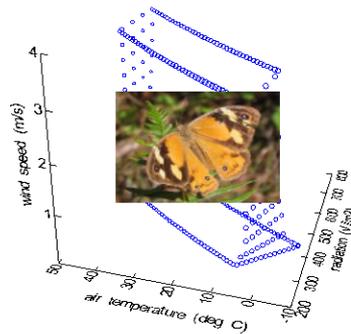
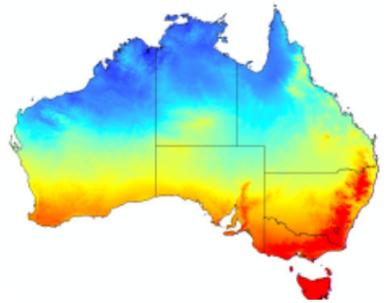
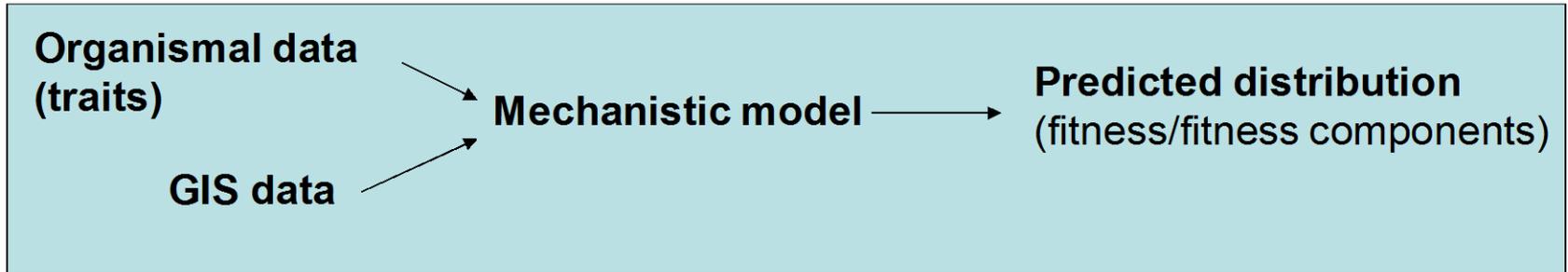
e.g., metapoblaciones

Hanski (1999)

- Asumen la influencia de factores o interacciones entre especies mediadas por correlaciones ambientales PERO no establecen explícitamente una hipótesis del proceso o mecanismo.

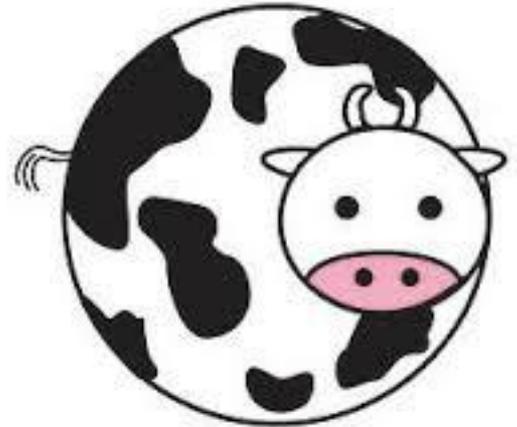
- Permiten hacer hipótesis explícitas sobre la interacción de la especie con factores ambientales u otras especies.

# Modelo mecanístico



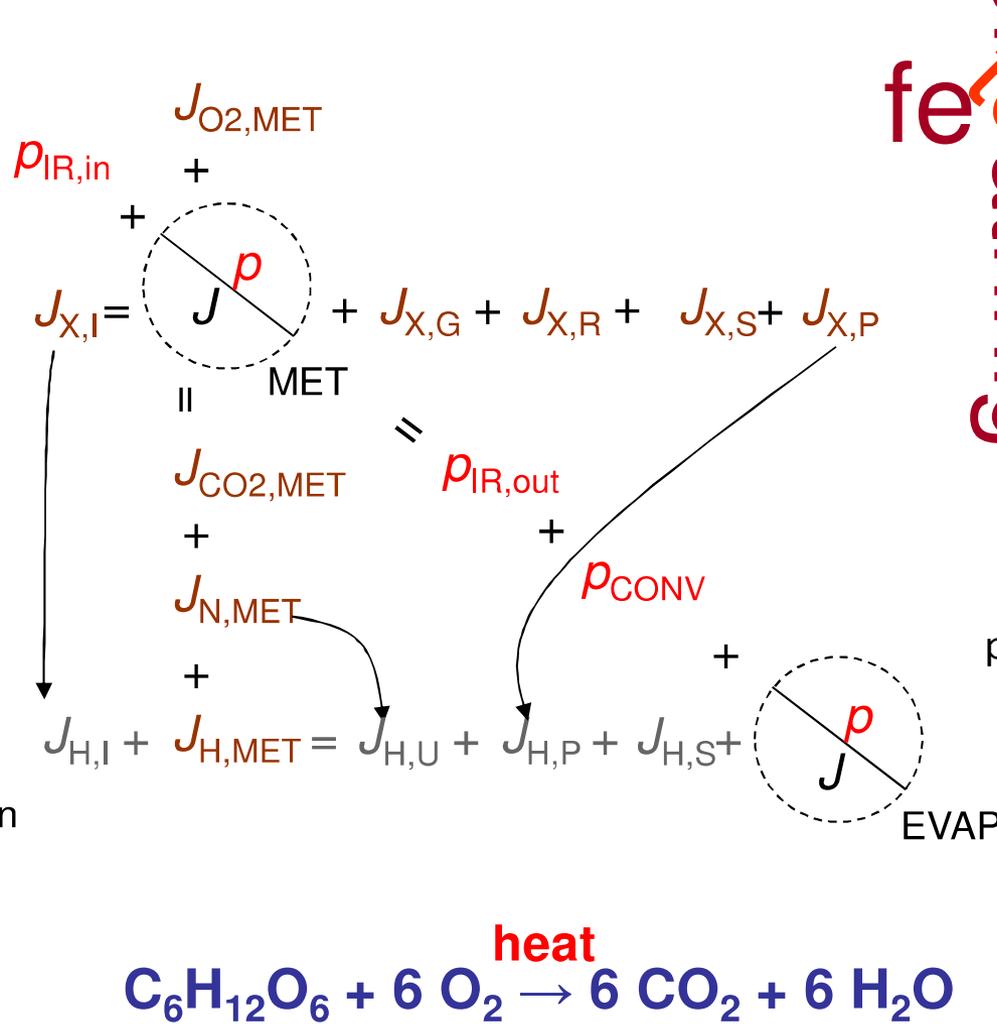
# “Primeros principios” – Ecología fisiológica

- Limitantes fisiológicos de los organismos.
- Procesos de transporte:
  - transferencia de calor
  - transferencia de masa
  - dinámica de fluidos
- Ecuaciones de balance – reglas termodinámicas.
- Balances de energía – intercambio de calor: conducción, convección, radiación, evaporación.
- Balances de masa:
  - intercambio de gases (superficies de respiración)
  - alimentos (materia seca, por el tracto intestinal)
  - agua (intestinos, superficies de respiración y órganos excretores)



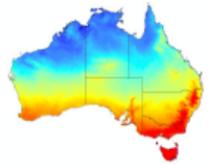
# Restricciones termodinámicas

- $p$  = heat flux
- $J$  = mass flux
- $X$  = food
- $H$  = water
- $I$  = ingested
- $P$  = product (faeces)
- $U$  = urinated
- $G$  = growth
- $R$  = reproduction
- $S$  = stored
- $O_2$  = oxygen
- $CO_2$  = carbon dioxide
- $N$  = nitrogenous waste
- $MET$  = 'metabolism'
- $EVAP$  = evaporation
- $SOLAR$  = solar radiation
- $IR$  = infrared radiation
- $CONV$  = convection
- $COND$  = conduction

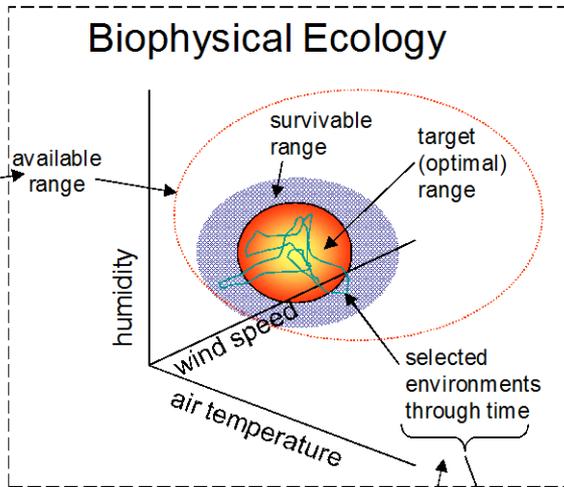


feeding  
 bathing  
 temperature  
 water

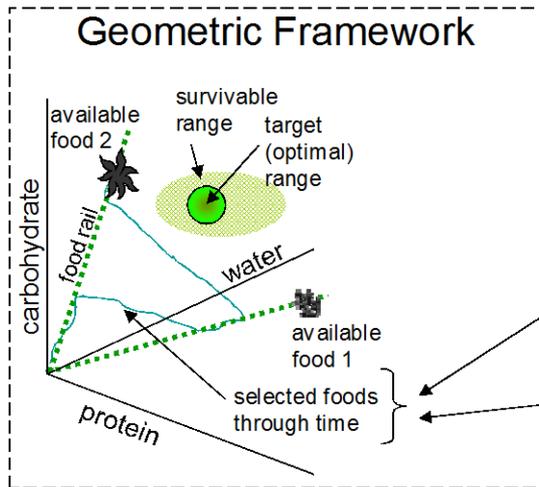
ecología biofísica + presupuestos de energía



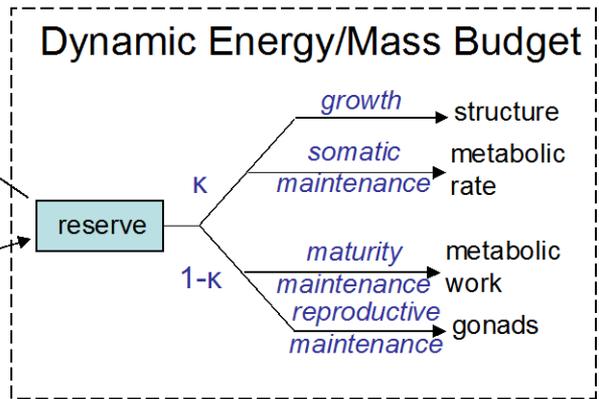
environmental gradients (GIS data)



Kearney and Porter TREE 2006  
Kearney et al. PTRS 2010



body temp/metabolic rate

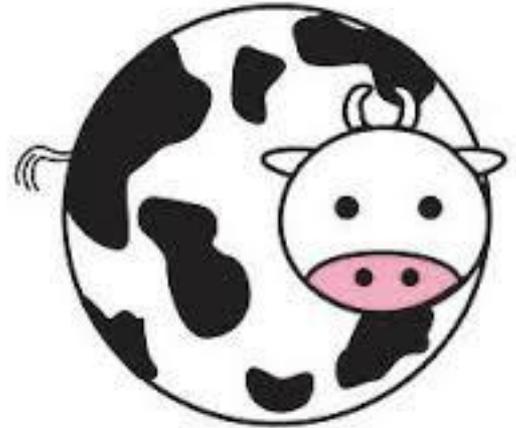


water balance

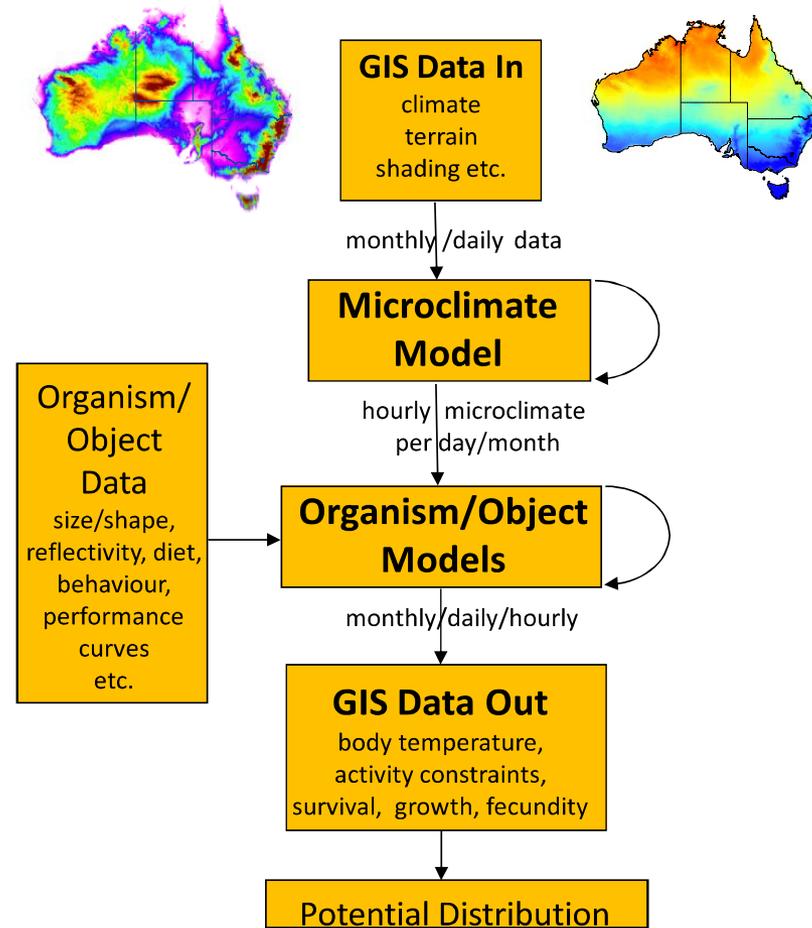
food ingested

# Modelo mecanístico en la práctica

- Los datos de entrada son **caracteres o atributos, no puntos**:
  - comportamiento
  - morfología
  - fisiología
- El reto: relacionar rasgos fisiológicos a los datos espaciales de habitat a través de un modelo “mecanístico” que capture el proceso
  - → si se basa en componentes de adecuación estaría representando un modelo mecanístico del **nicho fundamental ( $N_F$ )** de la especie
  - → inferir la distribución potencial



# Modelo – restricciones termodinámicas



Kearney & Porter 2009

# Variables bioclimáticas vs. variables derivada de modelos mecánicos

## Bioclim

### BIOCLIM

Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g., BIOCLIM, GARP). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

They are coded as follows:

BIO1 = Annual Mean Temperature  
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))  
BIO3 = Isothermality (BIO2/BIO7) (\* 100)  
BIO4 = Temperature Seasonality (standard deviation \*100)  
BIO5 = Max Temperature of Warmest Month  
BIO6 = Min Temperature of Coldest Month  
BIO7 = Temperature Annual Range (BIO5-BIO6)  
BIO8 = Mean Temperature of Wettest Quarter  
BIO9 = Mean Temperature of Driest Quarter  
BIO10 = Mean Temperature of Warmest Quarter  
BIO11 = Mean Temperature of Coldest Quarter  
BIO12 = Annual Precipitation  
BIO13 = Precipitation of Wettest Month  
BIO14 = Precipitation of Driest Month  
BIO15 = Precipitation Seasonality (Coefficient of Variation)  
BIO16 = Precipitation of Wettest Quarter  
BIO17 = Precipitation of Driest Quarter  
BIO18 = Precipitation of Warmest Quarter  
BIO19 = Precipitation of Coldest Quarter

GCMs – interpolaciones (número, distribución y calidad de datos en estaciones meteorológicas)

# Variables bioclimáticas vs. variables derivada de modelos mecanísticos

## Bioclim

### BIOCLIM

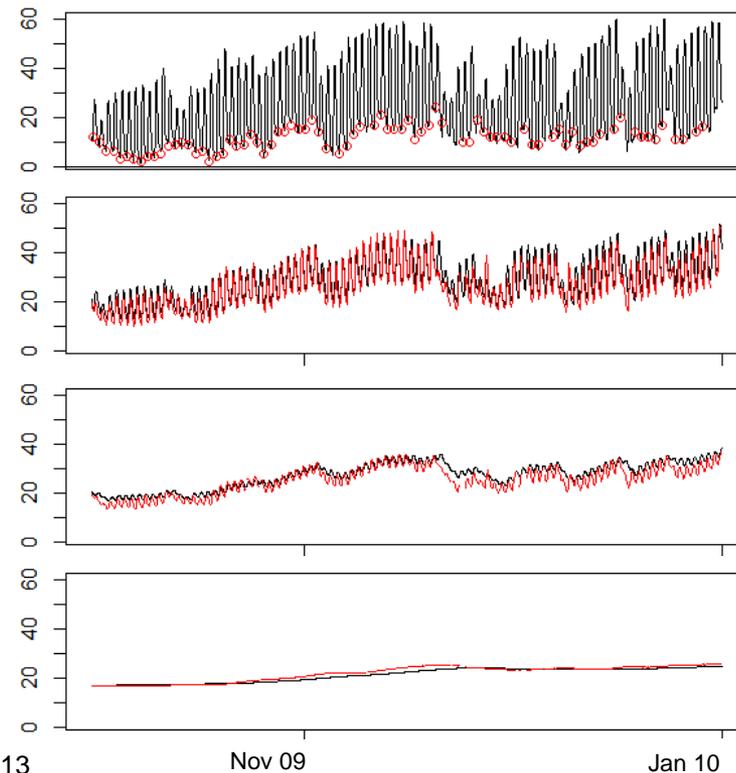
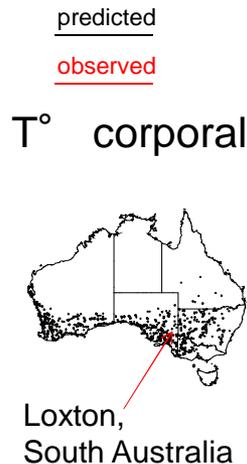
Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modeling (e.g., BIOCLIM, GARP). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4 of the year).

They are coded as follows:

- BIO1 = Annual Mean Temperature
- BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- BIO3 = Isothermality (BIO2/BIO7) (\* 100)
- BIO4 = Temperature Seasonality (standard deviation \*100)
- BIO5 = Max Temperature of Warmest Month
- BIO6 = Min Temperature of Coldest Month
- BIO7 = Temperature Annual Range (BIO5-BIO6)
- BIO8 = Mean Temperature of Wettest Quarter
- BIO9 = Mean Temperature of Driest Quarter
- BIO10 = Mean Temperature of Warmest Quarter
- BIO11 = Mean Temperature of Coldest Quarter
- BIO12 = Annual Precipitation
- BIO13 = Precipitation of Wettest Month
- BIO14 = Precipitation of Driest Month
- BIO15 = Precipitation Seasonality (Coefficient of Variation)
- BIO16 = Precipitation of Wettest Quarter
- BIO17 = Precipitation of Driest Quarter
- BIO18 = Precipitation of Warmest Quarter
- BIO19 = Precipitation of Coldest Quarter

GCMs – interpolaciones (número, distribución y calidad de datos en estaciones meteorológicas)

## Capas microclimáticas y relacionadas con comportamiento y procesos fisiológicos – resolución espacial y temporal



# Niche Mapper / Niche MapR

3 modelos mecanísticos:

- microclima (genera capas relevantes para la ecología térmica e hídrica)
- ectotérmico
- endotérmico (balances de masa y flujos de energía dependiendo de la morfología, fisiología y comportamiento)  
→ traducido en funciones de actividad, dispersión, sobrevivencia, crecimiento, etc.



Warren P. Porter (UW-Madison)



Michael Kearney (Melbourne, Aus.)

Ejemplo de un modelo orientado a proceso:

distribución del muérdago en el Desierto Sonorense

- Se especifica un mecanismo
- No necesariamente biofísico
- Pero más que en un modelo correlativo

# Process-based and correlative modeling of desert mistletoe distribution: a multiscale approach

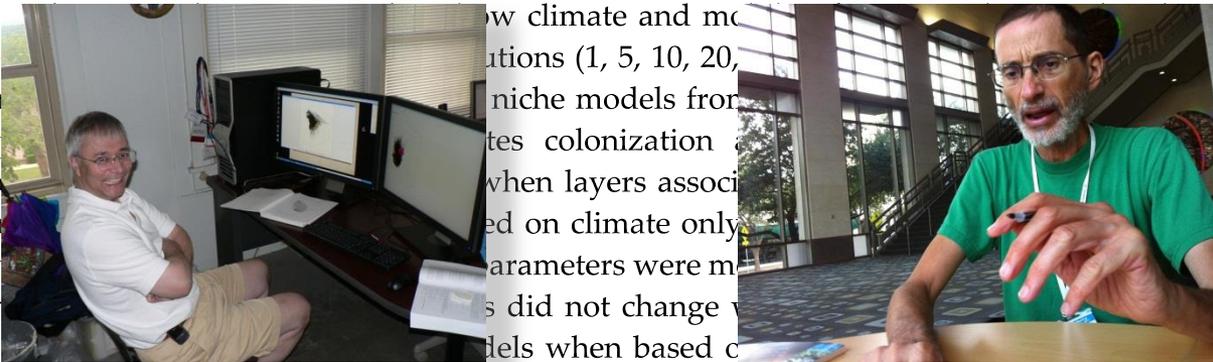
ANDRÉS LIRA-NORIEGA,<sup>1,†</sup> JORGE SOBERÓN,<sup>1</sup> AND CURTIS P. MILLER<sup>1,2</sup>

<sup>1</sup>Biodiversity Institute, University of Kansas, Lawrence, Kansas 66045 USA

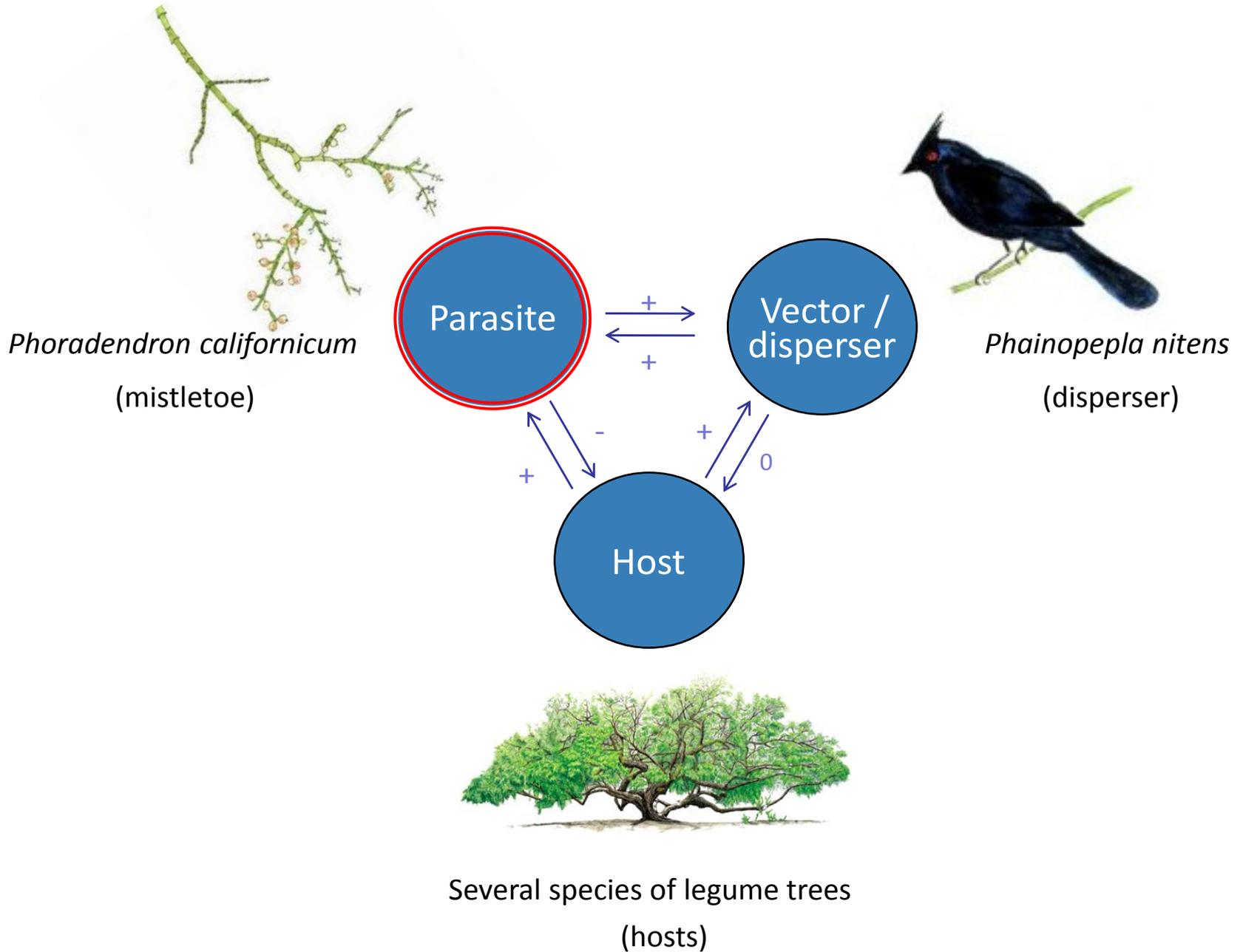
<sup>2</sup>College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001 USA

**Citation:** Lira-Noriega, A., J. Soberón, and C. P. Miller. 2013. Process-based and correlative modeling of desert mistletoe distribution: a multiscale approach. *Ecosphere* 4(8):99. <http://dx.doi.org/10.1890/ES13-00155.1>

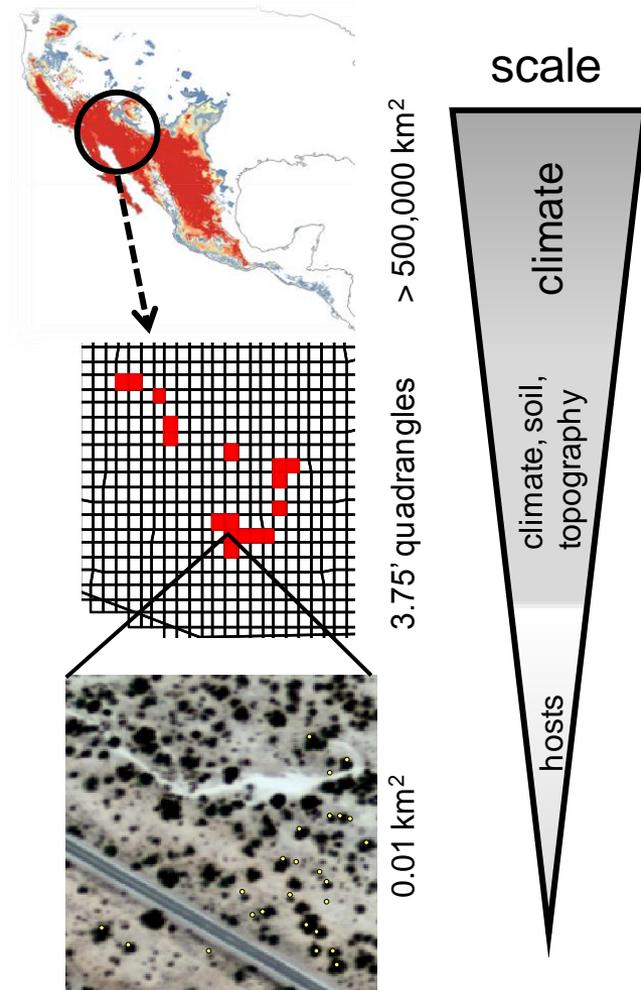
**Abstract.** Because factors affecting distributional areas of species change as scale (extent and grain) changes, different environmental and biological factors must be integrated across geographic ranges at different resolutions, to understand fully the patterns and processes underlying species' ranges. We expected climate factors to be more important at coarse resolutions and biotic factors at finer resolutions. We used data on occurrence of a parasitic plant (*Phoradendron californicum*), restricted to parts of the Sonoran Desert. We developed a process-based model of dispersal and colonization with a multiscale framework. We used a more layered approach to predictor variables, where more layers were important at finer resolutions. We found that models based on climate only did not change when based on climate and biotic factors. A clearer and more comprehensive mechanistic understanding was derived from the process-based algorithm.



**Key words:** colonization; dispersal; extinction; geographic range; metapopulation; *Phoradendron californicum*.



# Framework



- 1.- How does scaling affect the factors that determine the area of distribution of a species?
- 2.- Moving from a correlative (ENM) approach to a process-based approach.
- 3.- Climate is more important at coarse resolutions and biotic factors at finer resolutions.

# Fieldwork

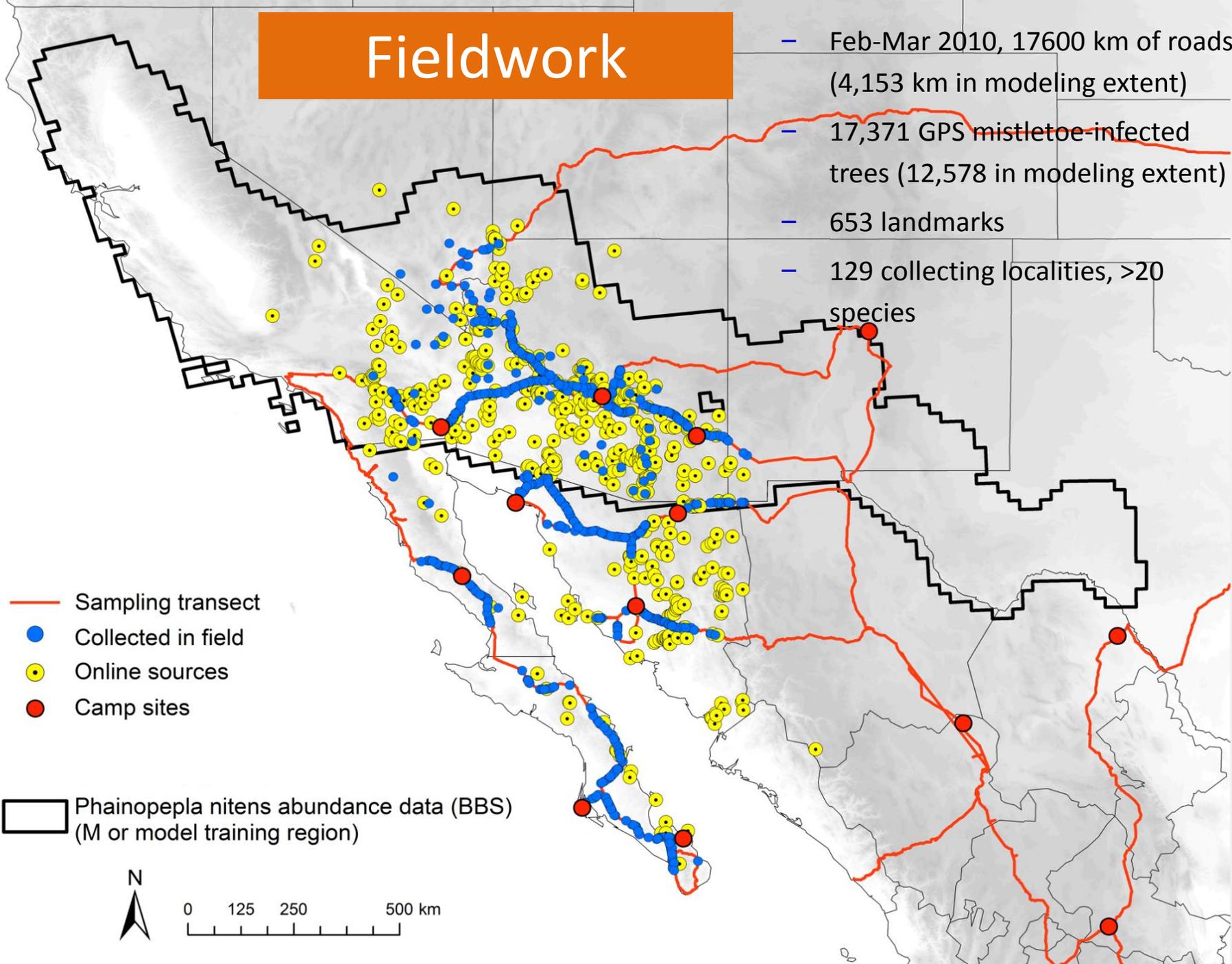
- Feb-Mar 2010, 17600 km of roads (4,153 km in modeling extent)
- 17,371 GPS mistletoe-infected trees (12,578 in modeling extent)
- 653 landmarks
- 129 collecting localities, >20 species

- Sampling transect
- Collected in field
- Online sources
- Camp sites

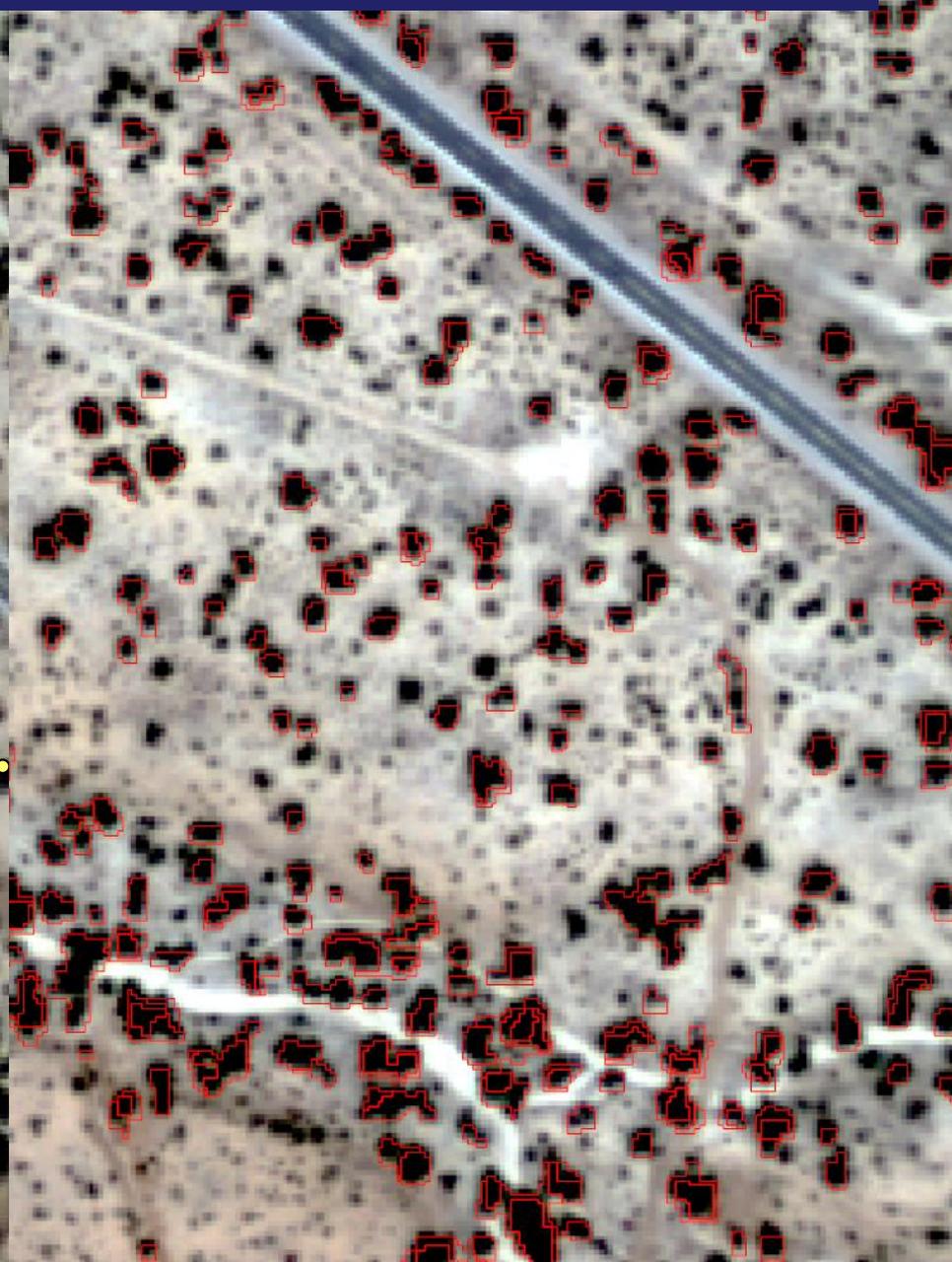
□ Phainopepla nitens abundance data (BBS)  
(M or model training region)



0 125 250 500 km



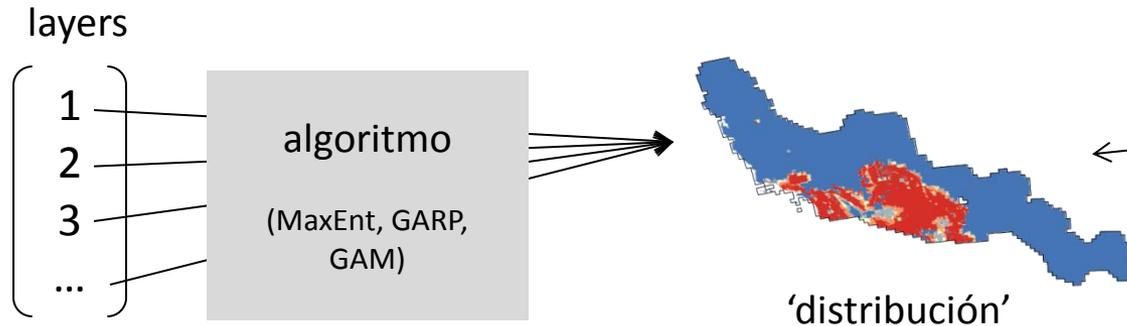
# Remote sensing



- Object-oriented classification (eCognition 3): number of host trees within a 300 m buffer along highways
- >60 1 m resolution NAIP images (81 km<sup>2</sup>) from 2007-2009
- Random forest: number of trees

# Modelos correlativos y mecanísticos

## Correlativos



## Mecanísticos y orientados a procesos

$$p_j = \frac{c_j}{c_j + e_j}$$

'distribución'  
(índice de favorabilidad-suitability)  
(probabilidad)

e.g., metapoblaciones  
Hanski (1999)

5 resoluciones espaciales (1, 5, 10, 20, 50 km)

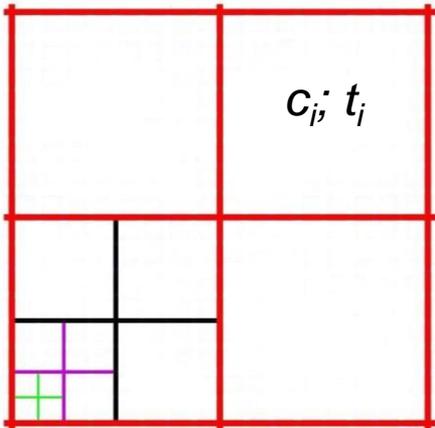
- Asumen la influencia de factores o interacciones entre especies mediadas por correlaciones ambientales **PERO** no establecen explícitamente una hipótesis del proceso o mecanismo.
- Permiten hacer hipótesis explícitas sobre la interacción de la especie con factores ambientales u otras especies.

# Process-based model

## ● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp \left( -d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}} \right)$$

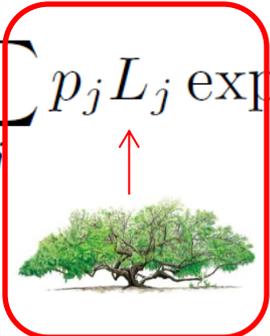
## ● extinction



$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$

# Process-based model

## ● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp \left( -d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}} \right)$$


## ● extinction

$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$

# Process-based model

## ● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp \left( -d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}} \right)$$

move cost  
↓  
 $\beta_0$

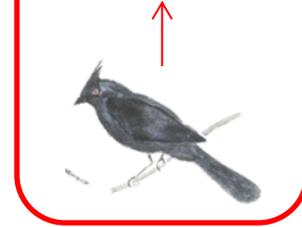
## ● extinction

$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$

# Process-based model

## ● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp \left( -d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}} \right)$$



## ● extinction

$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$

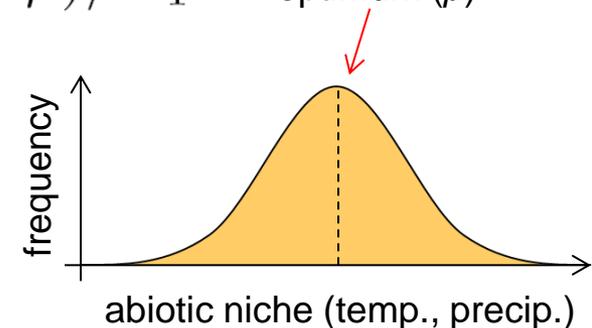
# Process-based model

## ● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp \left( -d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}} \right)$$

## ● extinction

$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$



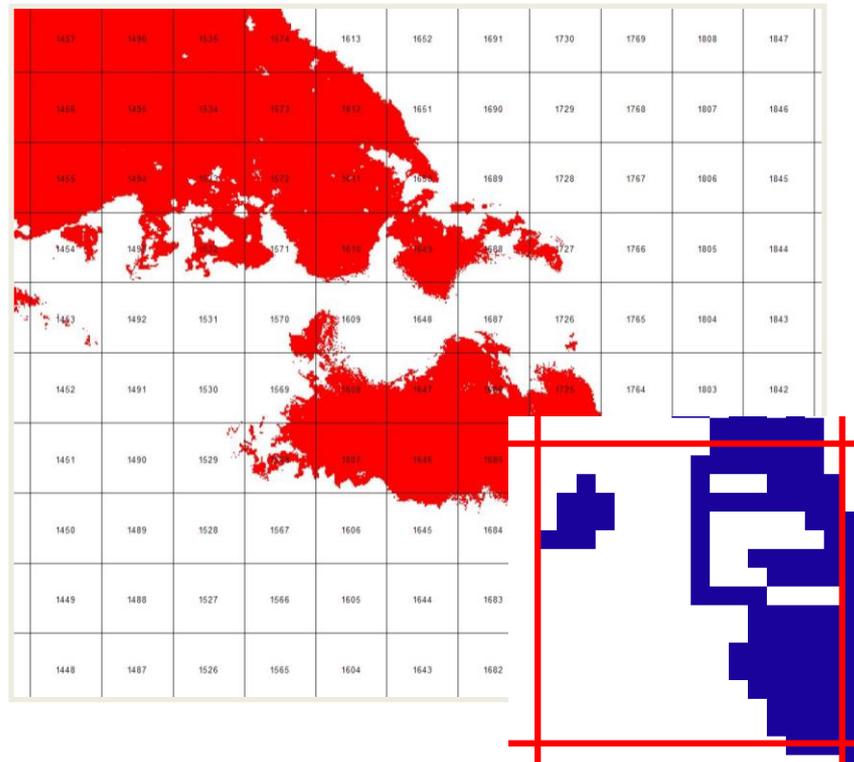
# Process-based model

● colonization

$$c_i = \beta_{d+1} \sum_j p_j L_j \exp\left(-d_{ij} \frac{\beta_0}{1 + \sum_{k=1}^s \beta_k b_{jk}}\right)$$

● extinction

$$t_i = (\mathbf{x}_i - \mu)' \Sigma^{-1} (\mathbf{x}_i - \mu) / K_1$$



# Potential distributions

1 km

50 km

Process-based



temp.  
precip.



MaxEnt

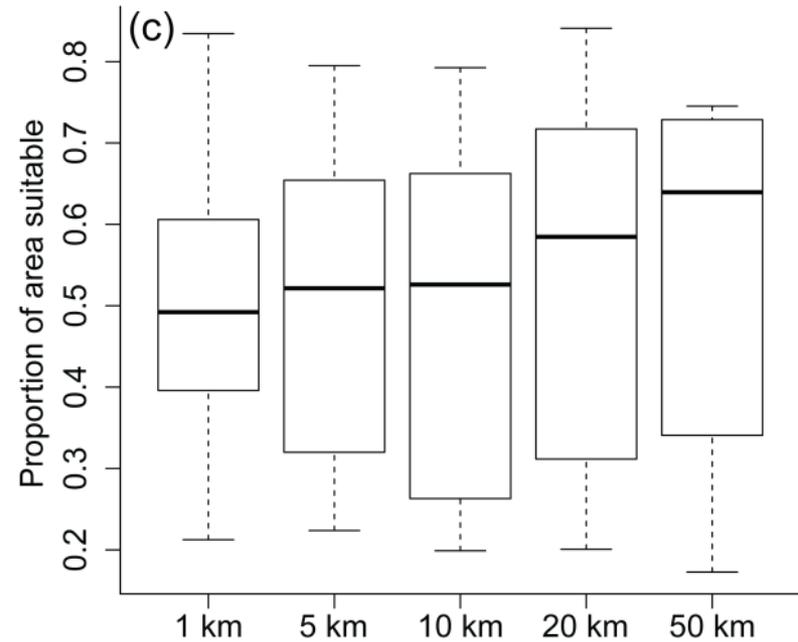
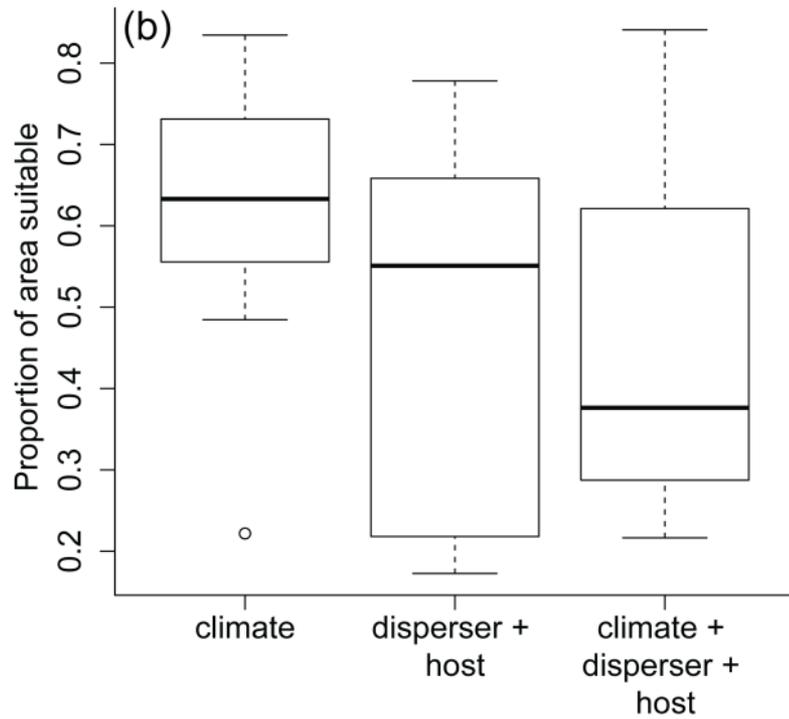
disperser  
trees



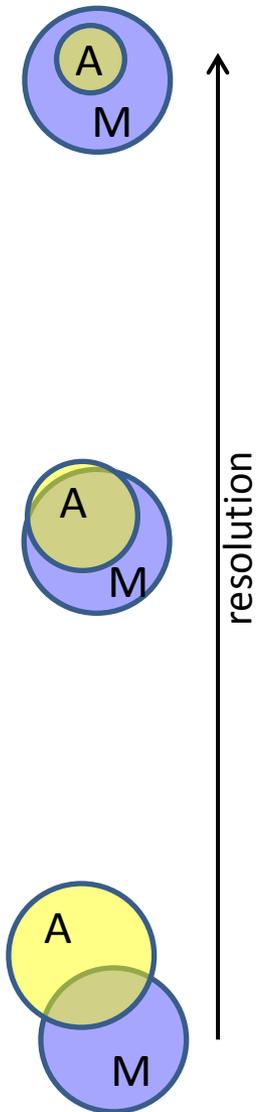
temp.  
precip.  
disperser  
trees



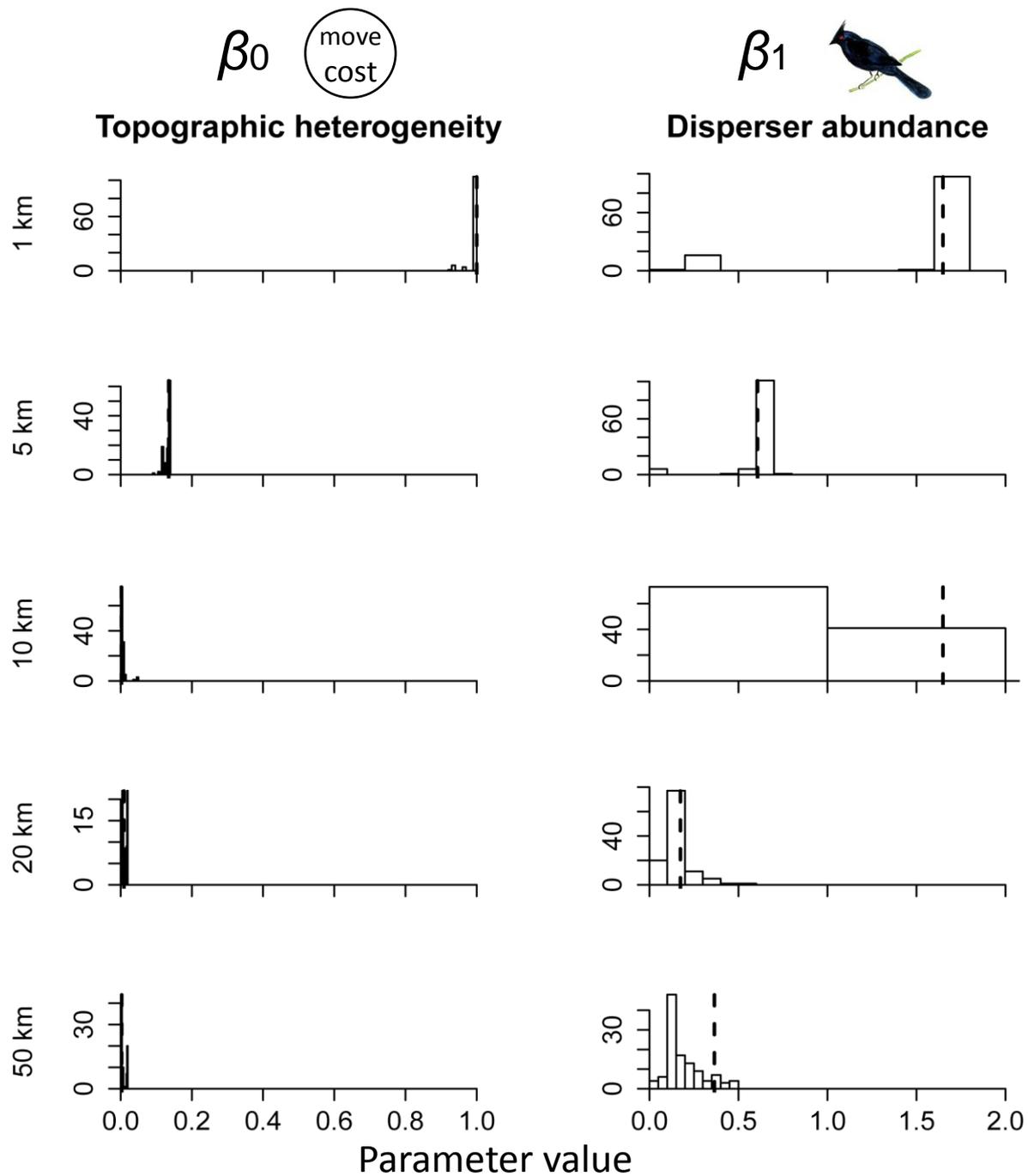
# Model performance



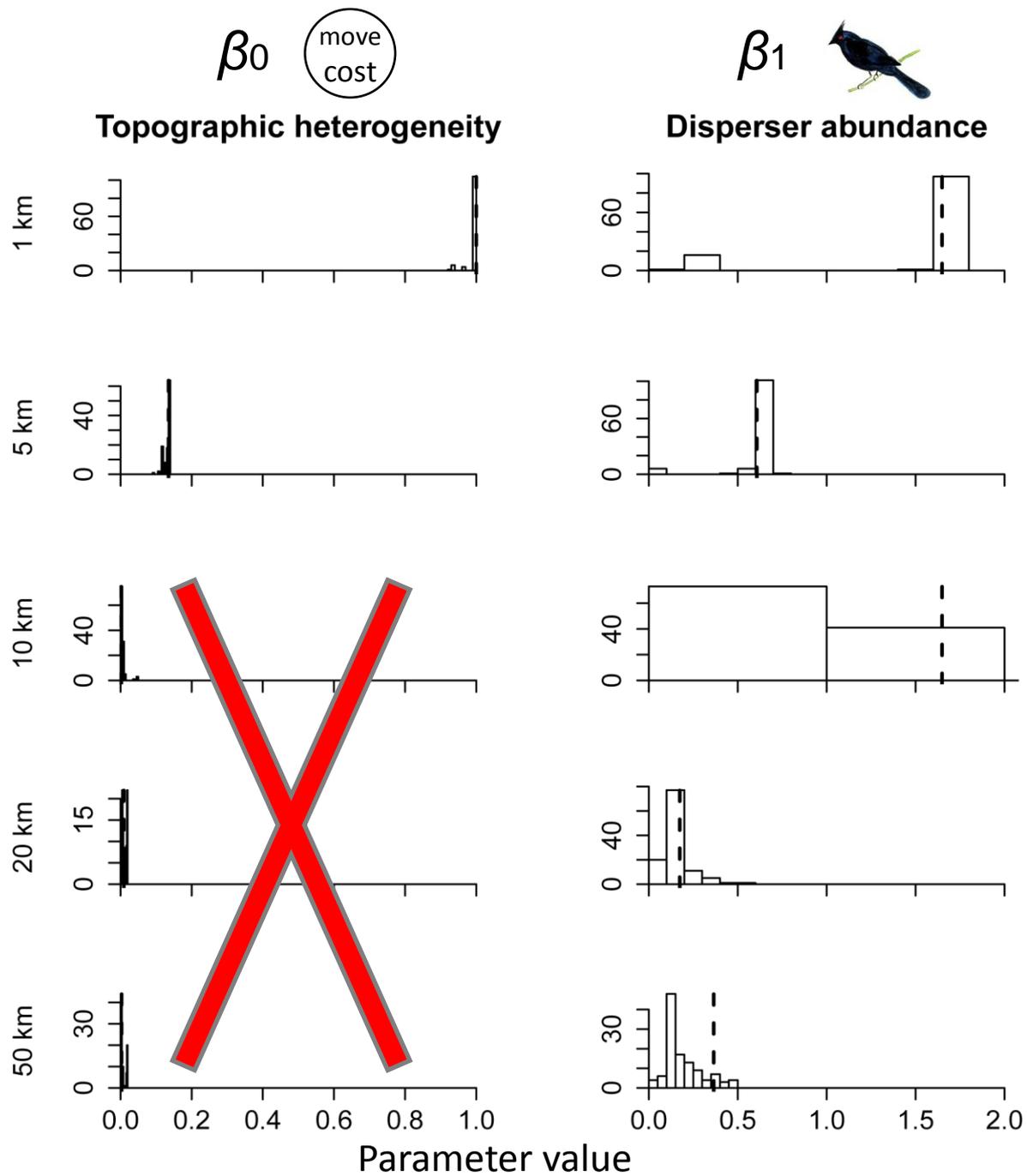
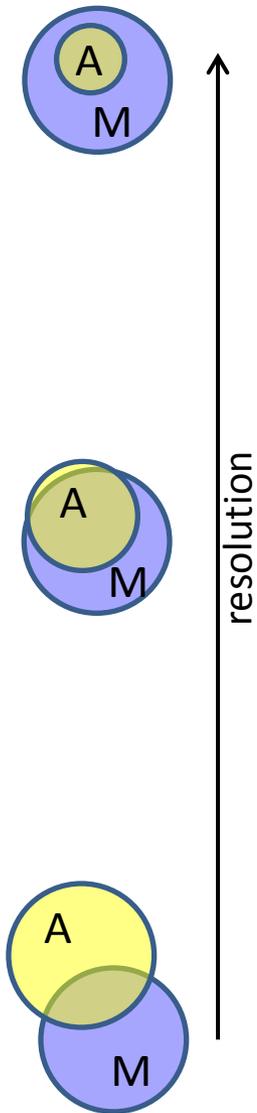
# Colonization parameters



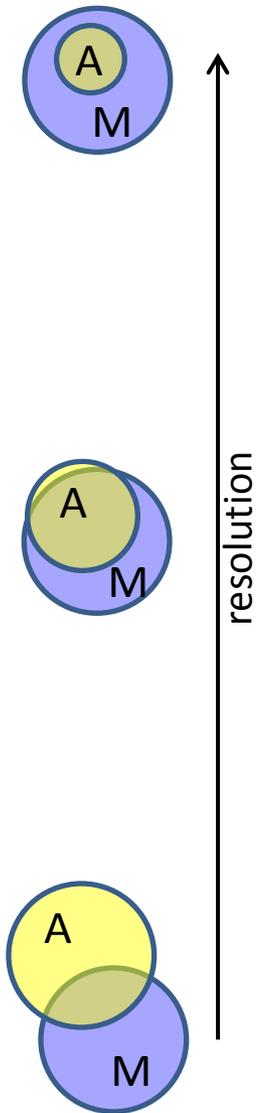
resolution



# Colonization parameters

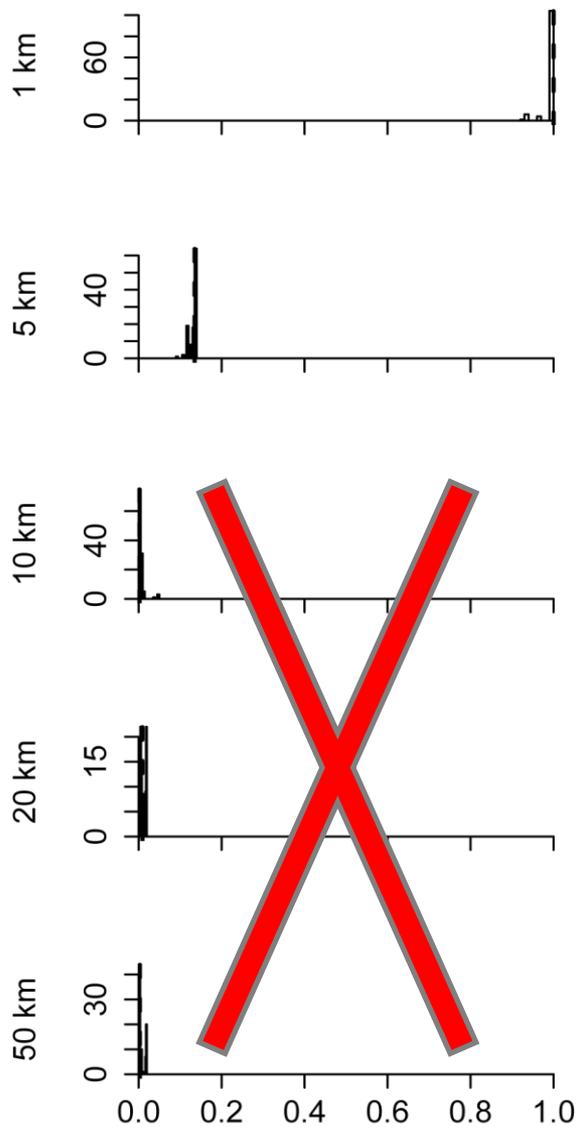


# Colonization parameters



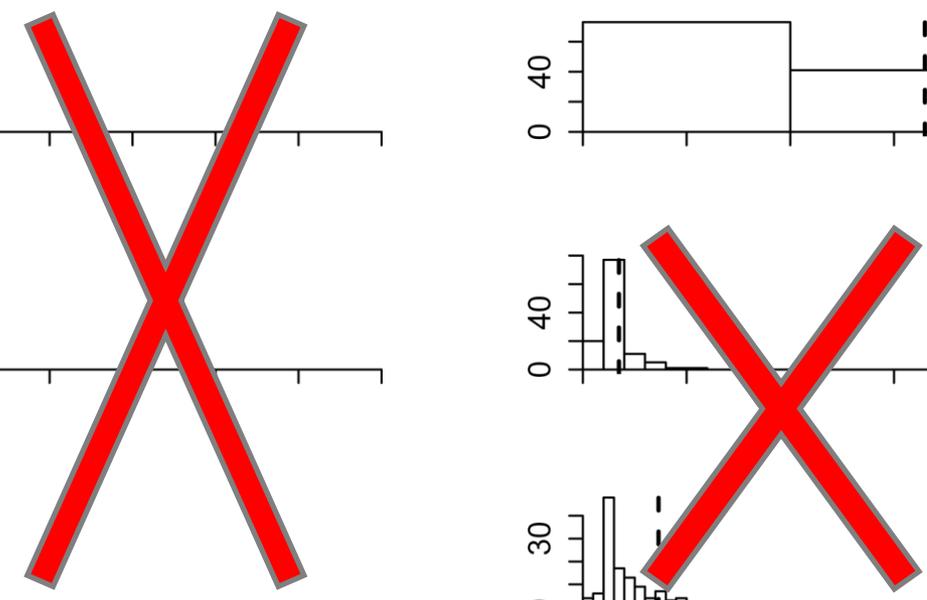
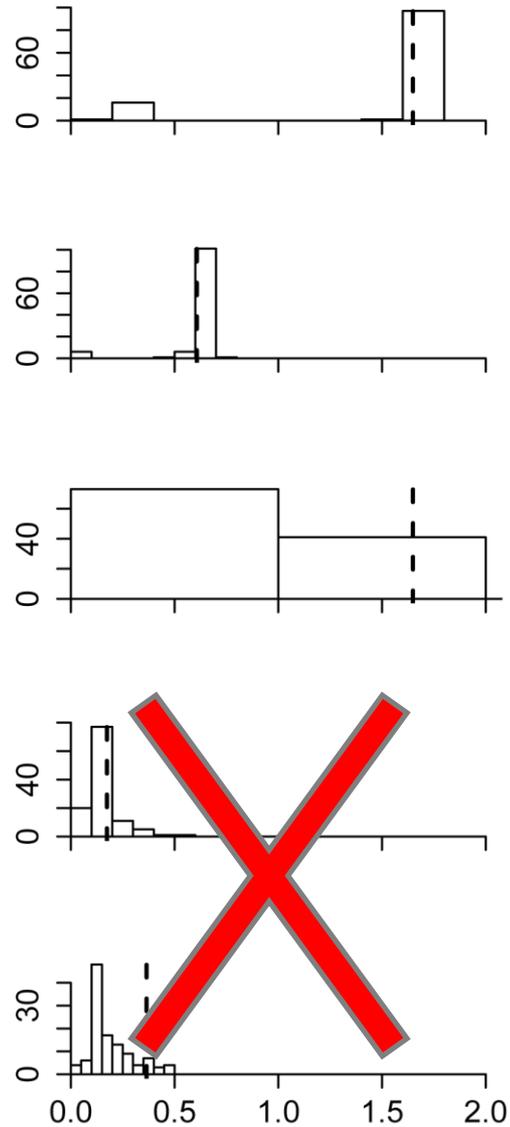
$\beta_0$  (move cost)

Topographic heterogeneity

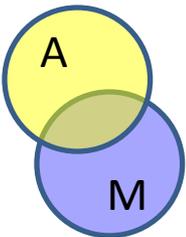
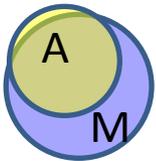
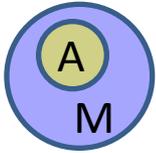


$\beta_1$  (bird icon)

Disperser abundance



# Extinction parameters



resolution ↑

	1 km	5 km	10 km	20 km
cent. T 1 km				
cent. T 5 km	0.70798			
cent. T 10 km	0.51133	0.76370		
cent. T 20 km	0.48072	0.42342	0.32195	
cent. T 50 km	0.34926	0.31525	0.24539	0.72072

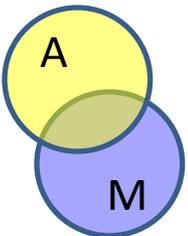
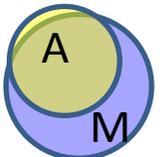
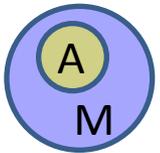
	1 km	5 km	10 km	20 km
cent. PP 1 km				
cent. PP 5 km	0.34944			
cent. PP 10 km	0.62765	0.34944		
cent. PP 20 km	0.71373	0.62765	0.49596	
cent. PP 50 km	0.47808	0.71373	0.32780	0.75283

	1 km	5 km	10 km	20 km
var T 1 km				
var T 5 km	0.99958			
var T 10 km	0.86275	0.99877		
var T 20 km	0.55034	0.99684	0.75631	
var T 50 km	0.99971	0.99967	0.99980	1.00000

	1 km	5 km	10 km	20 km
var PP 1 km				
var PP 5 km	0.99971			
var PP 10 km	1.00000	1.00000		
var PP 20 km	0.98449	0.98448	1.00000	
var PP 50 km	0.99990	0.99990	1.00000	0.99829

	1 km	5 km	10 km	20 km
cov T,PP 1 km				
cov T,PP 5 km	0.94617			
cov T,PP 10 km	0.98294	0.98270		
cov T,PP 20 km	0.99825	0.99384	0.98306	
cov T,PP 50 km	0.99862	0.99967	0.98307	0.99826

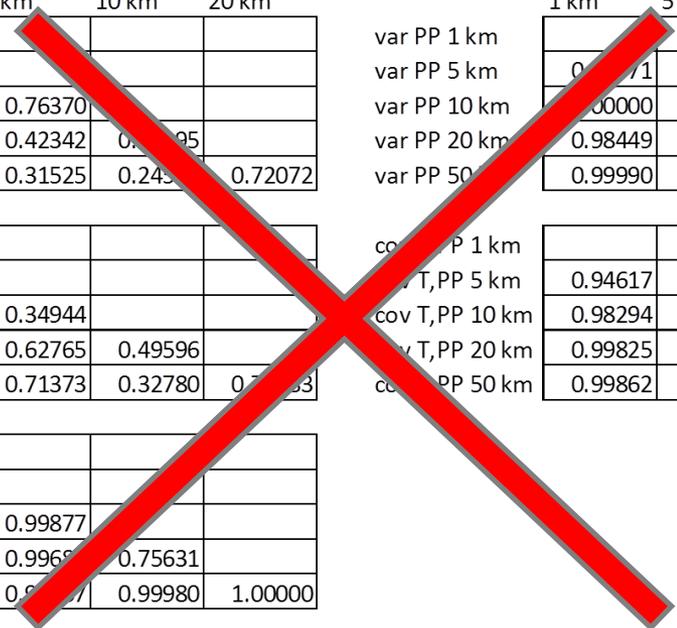
# Extinction parameters



resolution ↑

	1 km	5 km	10 km	20 km
cent. T 1 km				
cent. T 5 km	0.70798			
cent. T 10 km	0.51133	0.76370		
cent. T 20 km	0.48072	0.42342	0.7095	
cent. T 50 km	0.34926	0.31525	0.243	0.72072
cent. PP 1 km				
cent. PP 5 km	0.34944			
cent. PP 10 km	0.62765	0.34944		
cent. PP 20 km	0.71373	0.62765	0.49596	
cent. PP 50 km	0.47808	0.71373	0.32780	0.71373
var T 1 km				
var T 5 km	0.99958			
var T 10 km	0.86275	0.99877		
var T 20 km	0.55034	0.9968	0.75631	
var T 50 km	0.99971	0.99971	0.99980	1.00000

	1 km	5 km	10 km	20 km
var PP 1 km				
var PP 5 km	0.99971			
var PP 10 km	1.00000	1.00000		
var PP 20 km	0.98449	0.98448	1.00000	
var PP 50 km	0.99990	0.99990	1.00000	0.99829
cov T, PP 1 km				
cov T, PP 5 km	0.94617			
cov T, PP 10 km	0.98294	0.98270		
cov T, PP 20 km	0.99825	0.99384	0.98306	
cov PP 50 km	0.99862	0.99967	0.98307	0.99826



# Conclusions Process-based model

- Dispersal related parameters are more relevant at finer scales (1-5 km) and less relevant at coarse scales (10-50 km).
- Abiotic factors associated with extinction do not change with scale.
- Correlative models overfit to occurrence data when more variables are used.
- **Original hypothesis partially supported: dispersal related variables dominate at fine resolutions; effect of climate doesn't change with resolution.**

## Ventajas y desventajas de los modelos mecánicos, orientados a procesos y correlativos

	<b>Ventajas</b>	<b>Desventajas</b>
<b>Correlativos</b>	<p>Económicos en disponibilidad de algoritmo, tiempo de ejecución y tipo de datos.</p> <p>Permite incorporar datos provenientes de modelos mecánicos.</p>	<p>Asumen la influencia de variables ambientales en procesos fisiológicos y distribuciones de especies.</p>
<b>Mecánicos</b>	<p>Permite estimar el nicho fundamental (<math>N_F</math>) cuando usa datos fisiológicos y de adecuación.</p> <p>Informan de manera explícita sobre tolerancias fisiológicas.</p>	<p>Requieren datos más “costosos” (observaciones controladas en laboratorio o campo).</p> <p>Requieren de diseño y programación de acuerdo al problema.</p>
<b>Orientados a procesos</b>	<p>Permiten estimar importancia de procesos bióticos.</p>	