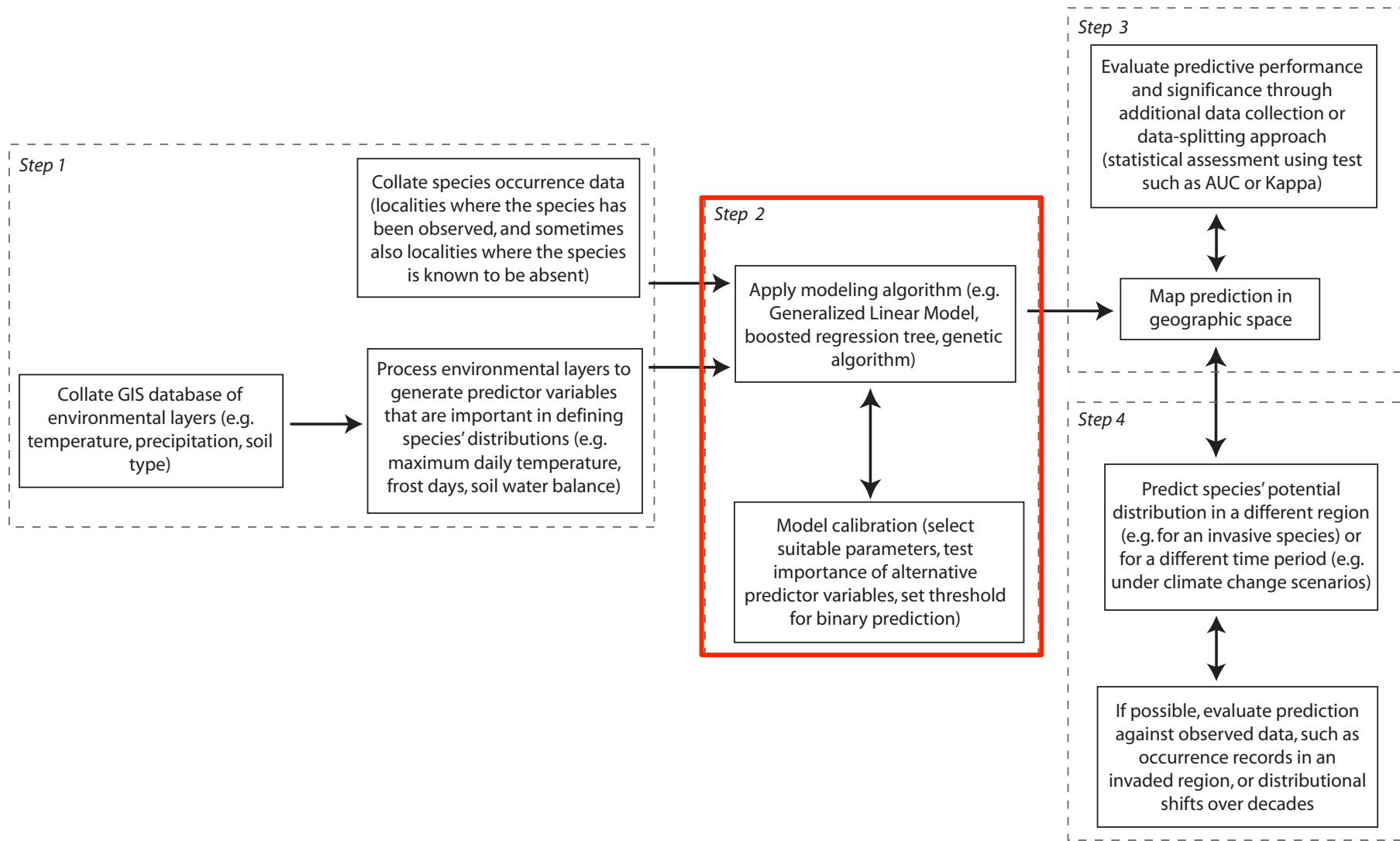




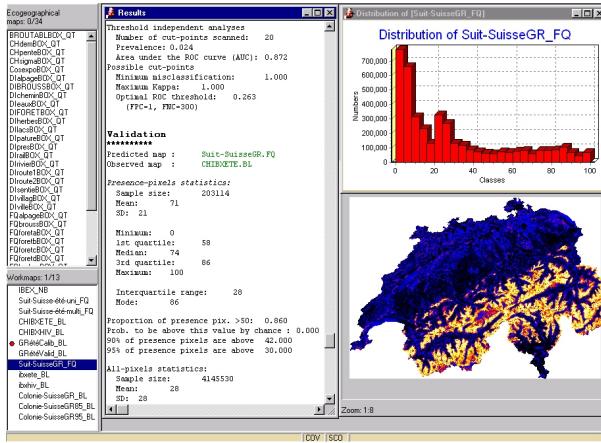
# Algoritmos de Modelación (o las máquinas de chorizos)

Enrique Martínez-Meyer  
[emm@ib.unam.mx](mailto:emm@ib.unam.mx)

# El proceso del modelado de nichos



Existe una gran variedad de algoritmos de modelación en el mercado (entre 20 y 30), así como una serie de plataformas en donde se han implementado



Global Change Biology (2003) 9, 1353–1362

**BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change**

WILFRIED THUILLER

## BIOdiversity MODelling



**BIOMOD – a platform for ensemble forecasting of species distributions**

Wilfried Thuiller, Bruno Lafourcade, Robin Engler and Miguel B. Araújo

Ecography 32: 369–373, 2009  
doi: 10.1111/j.1600-0524.2009.05748.x  
© 2009 The Authors. Journal compilation © 2009 Ecography  
Subject Editor: Carsten Rahbek. Accepted 30 December 2008

Ecography 32: 369–373, 2009

doi: 10.1111/j.1600-0524.2009.05748.x

© 2009 The Authors. Journal compilation © 2009 Ecography

Subject Editor: Carsten Rahbek. Accepted 30 December 2008

Software notes

## Plataformas con múltiples algoritmos

**Biomod:** <http://cran.r-project.org/web/packages/biomod2/index.html>

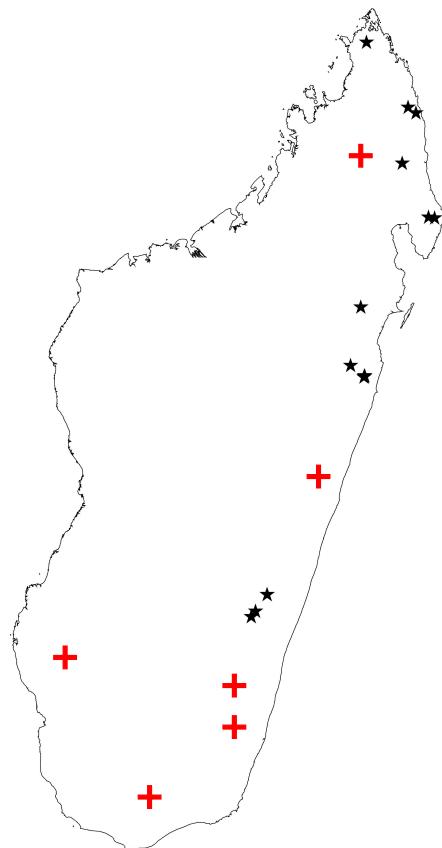
**Dismo:** <http://cran.r-project.org/web/packages/dismo/index.html>

**OpenModeller:** <http://openmodeller.sourceforge.net/>

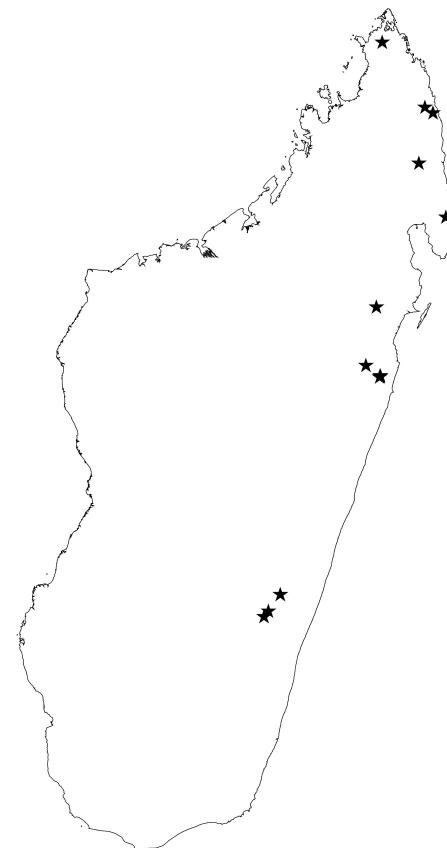
**ModEco:** <http://gis.ucmerced.edu/ModEco/>

**Div-GIS:** <http://www.diva-gis.org>

Para tomar una decisión en cuál algoritmo usar, primero hay que saber el tipo de datos que tenemos



Presencia/ausencia  
(e.g. ANN)



Sólo presencia  
(e.g. BIOCLIM)



Presencia/pseudo-  
ausencia  
(e.g. GARP)

Después hay que saber como funcionan.....

# Clasificación de métodos para modelar

Algoritmo	Ontogenia	Tipo de datos de entrada
Bioclim, Aquamaps, Envelope Score, SVM, Surface range envelopes	Envolturas ambientales	Sólo Presencia
Euclidian, Mahalanobis, Manhattan, Chebishev	Distancias ambientales	Sólo Presencia
Regresiones Logit, GLM, GAM, MARS, CART	Regresiones y clasificaciones	Presencia/ausencia
Redes Neuronales (ANN), BRT, Algoritmos genéticos, SVM	Inteligencia Artificial	Presencia/ausencia
GARP	Algoritmos Genéticos	Presencia/pseudo-ausencia
ENFA, Climate Space Model (PCA), Multiple Discriminant Analysis	Estadística multivariada: Estructura de varianzas	Presencia/entorno
MaxEnt	Máxima Entropía	Presencia/entorno

**Algoritmos que usan datos de  
solo presencia**

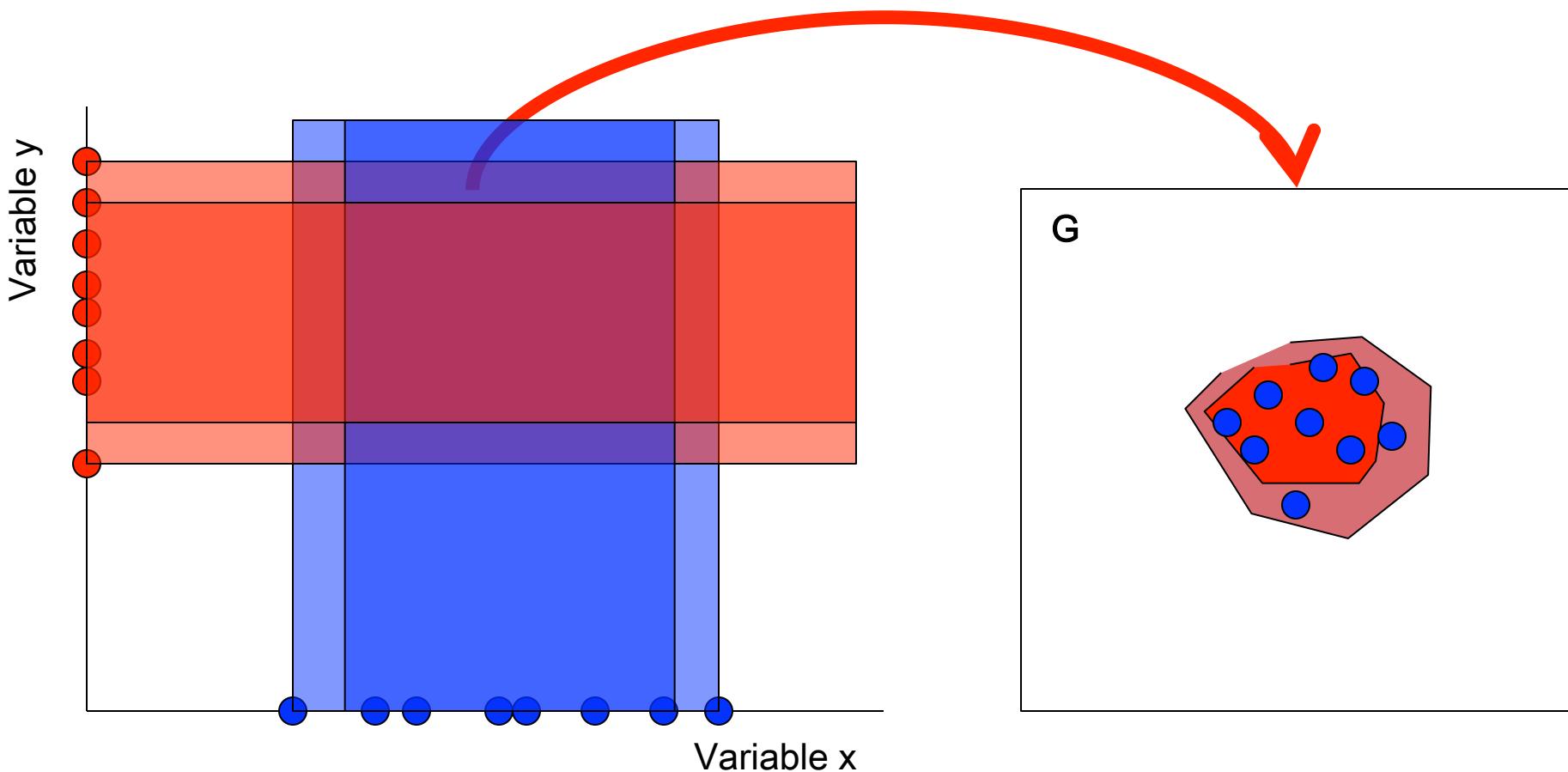
## BIOCLIM

Fue el primer algoritmo de modelado y fue desarrollado en Australia a finales de los 70's.

Es bastante simple pues construye una caja alrededor de los puntos de presencia en cada variable, resultando en una envoltura ambiental multidimensional rectilínea

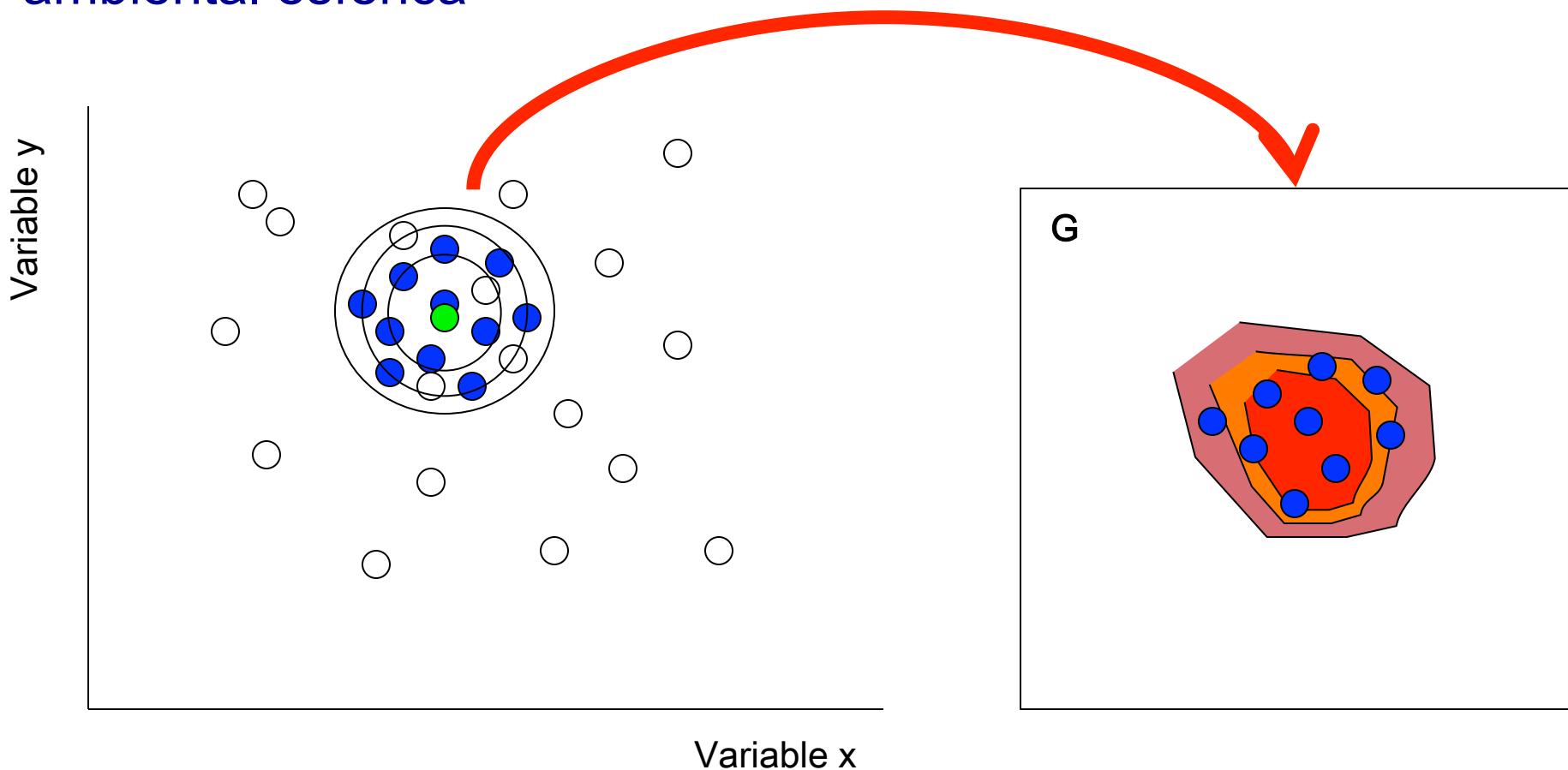
# BIOCLIM

Para evitar el efecto negativo de sobrepredicción provocado por los datos extremos, la envoltura ambiental se puede reducir a diferentes desviaciones estándar



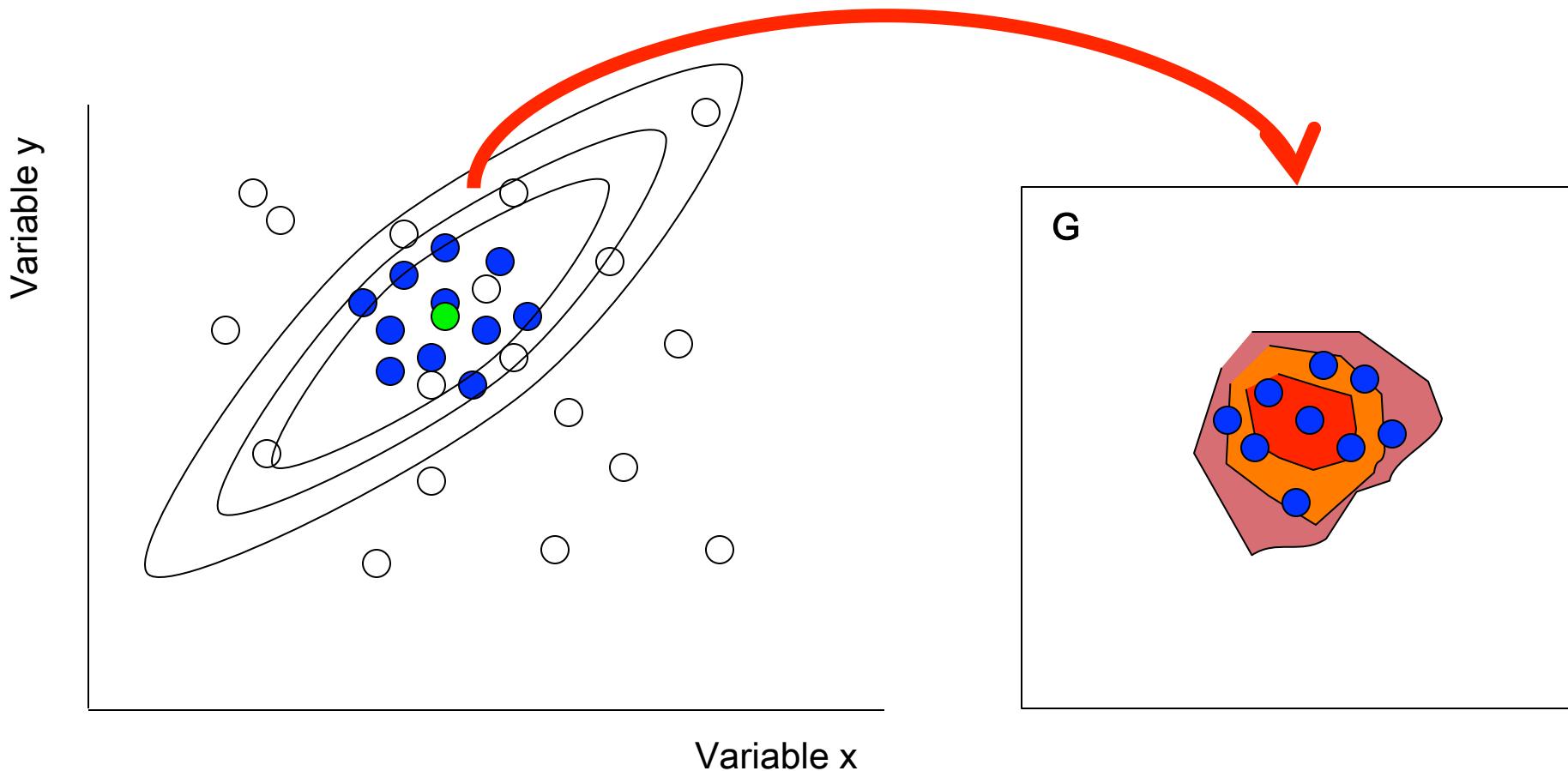
# Distancias Ambientales: Euclídea

Calcula la distancia euclídea multidimensional para identificar los pixeles que pertenecen a una envoltura ambiental esférica



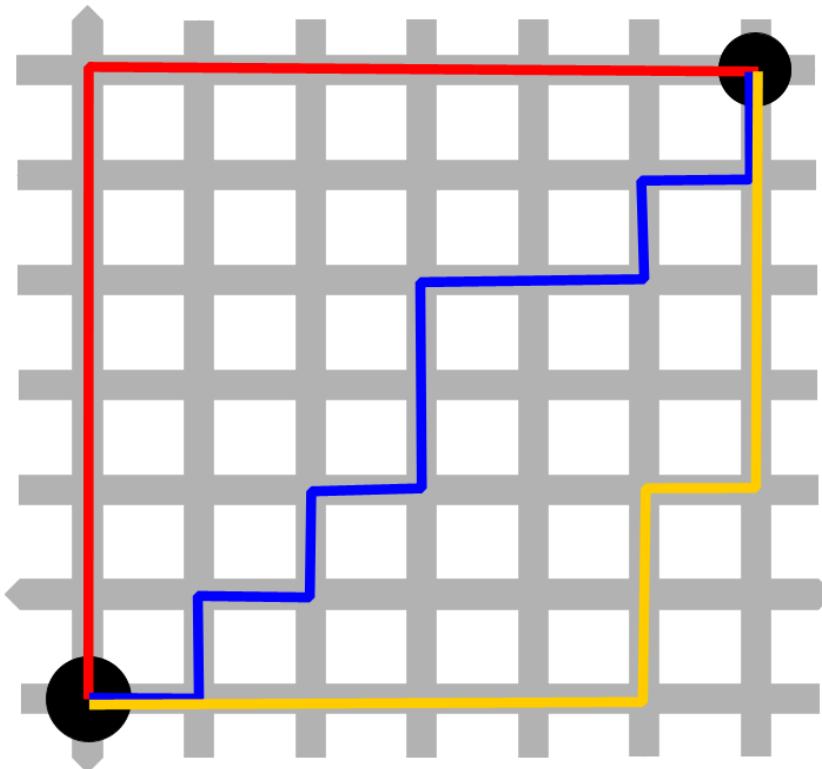
# Distancias Ambientales: Mahalanobis

Calcula la distancia de Mahalanobis multidimensional para identificar los pixeles que pertenecen a una envoltura ambiental elíptica, dada la correlación entre las variables

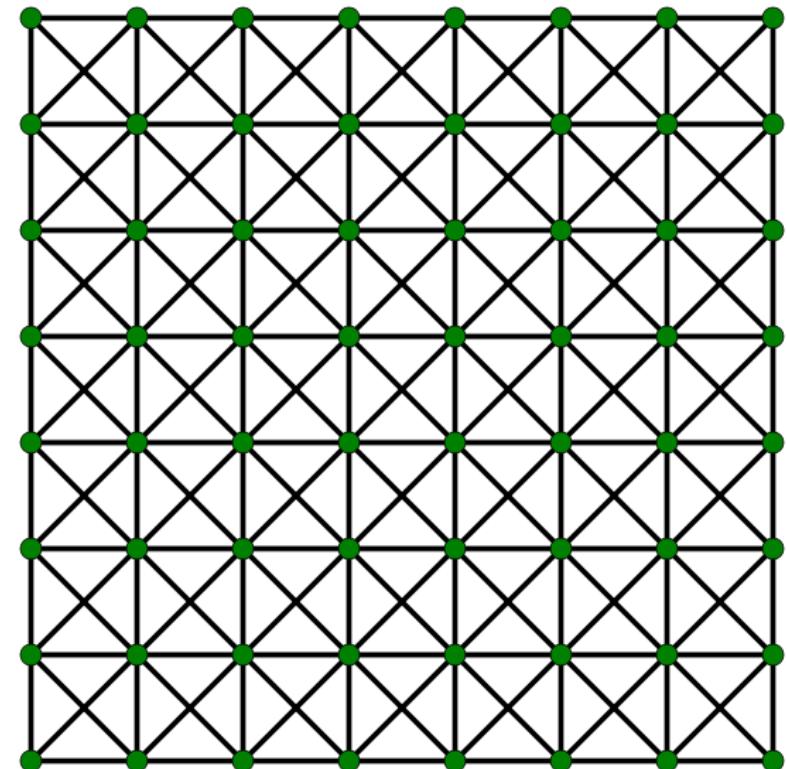


# Otras Distancias Ambientales

Gower / Manhattan  
(Domain)



Chebyshev

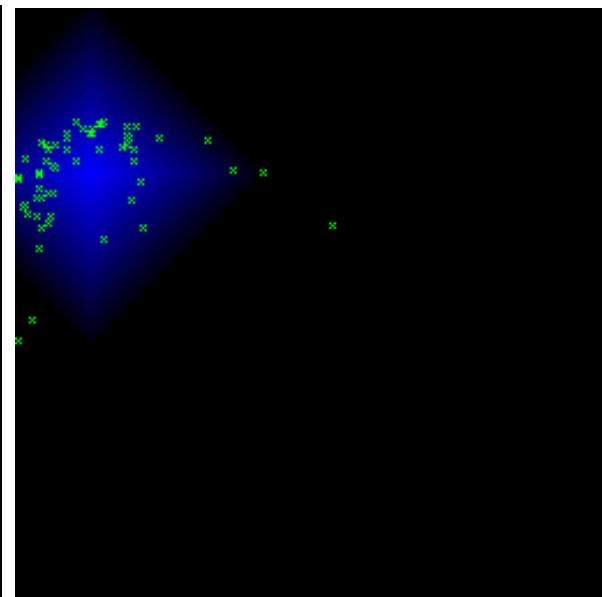
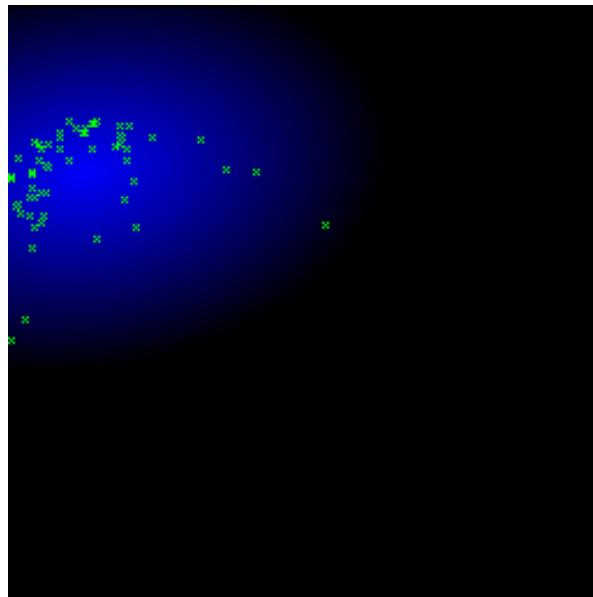
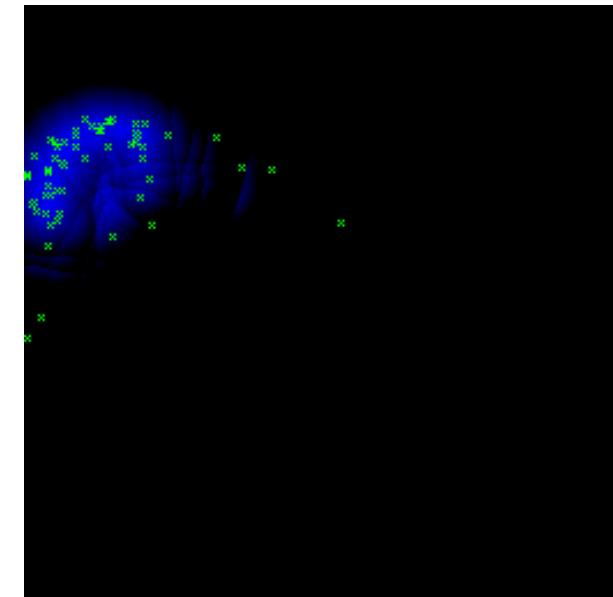


## Diferencia entre las Distancias Ambientales

Euclidiana

Mahalanobis

Gower



Precipitación

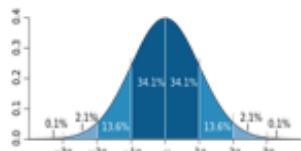
Temperatura

# Algoritmos que usan datos de presencia-ausencia

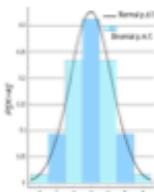
## Modelos Lineales Generalizados (GLM)

Es una forma de regresión. Busca una relación estadística entre una variable de respuesta y variables predictoras por medio de una función de enlace.

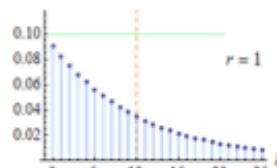
Permite trabajar con distribuciones de distinta naturaleza, así como con datos no lineales y no constantes



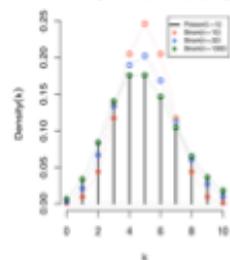
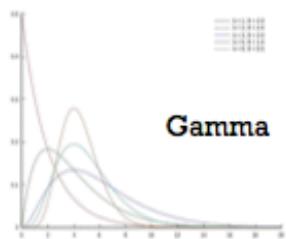
Normal: datos continuos



Binomial: datos discretos (0-1)



Binomial negativa: datos discretos



Poisson: datos discretos en intervalos de tiempo/espacio fijos

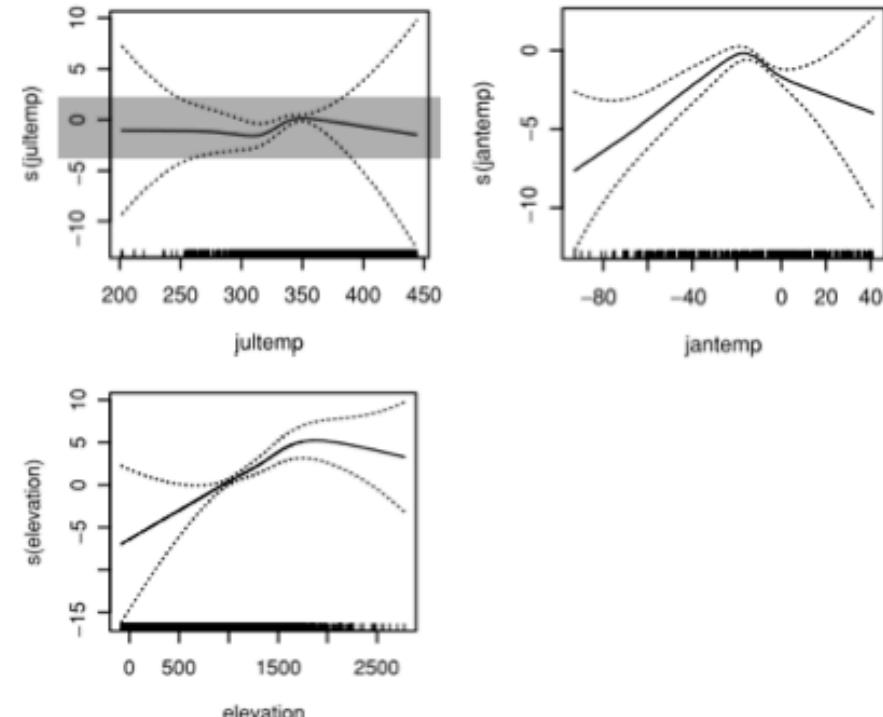
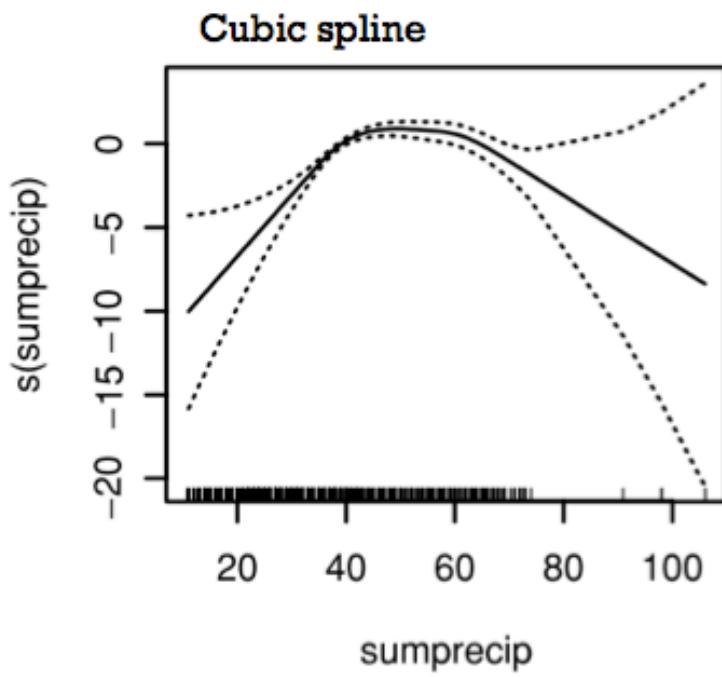
La variable de respuesta es de tipo binomial (0-1)



## Modelos Aditivos Generalizados (GAM)

Es una extensión no paramétrica del GLM que es más flexible para identificar y describir relaciones no lineales entre los predictores y la variable de respuesta. Usa una función de suavizado para modelar mejor el comportamiento de los datos.

GAM (logit link, datos binarios –presencia-ausencia)

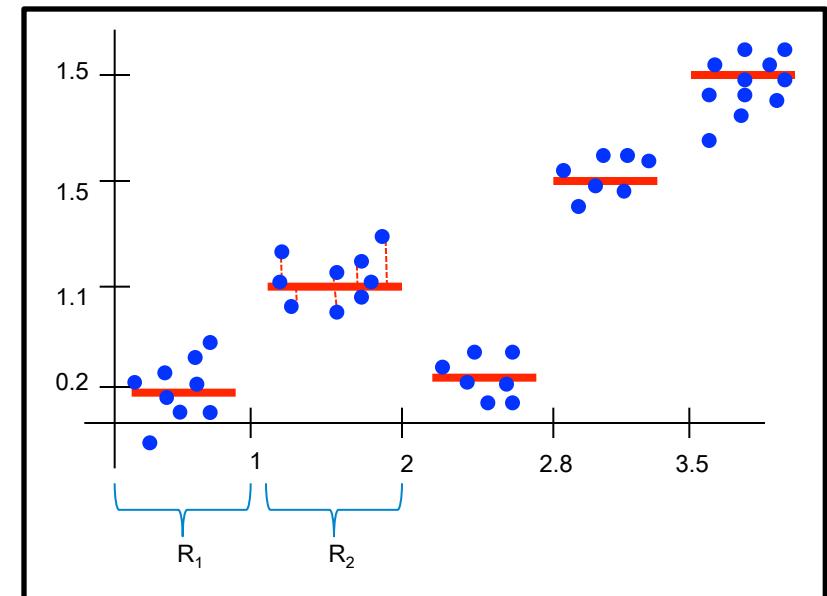
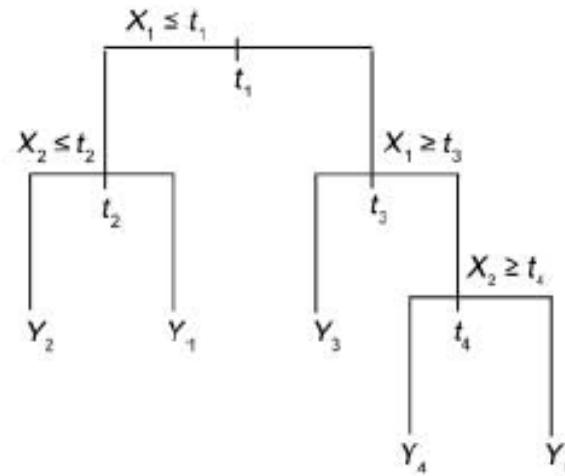


# Boost Regression Trees (BRT)

BRT es una técnica que busca mejorar el rendimiento de un modelo único, mediante el ajuste de muchos modelos, los cuales combina para la predicción. Utiliza de forma iterativa árboles de regresión para construir ensambles de árboles

Utiliza dos algoritmos:

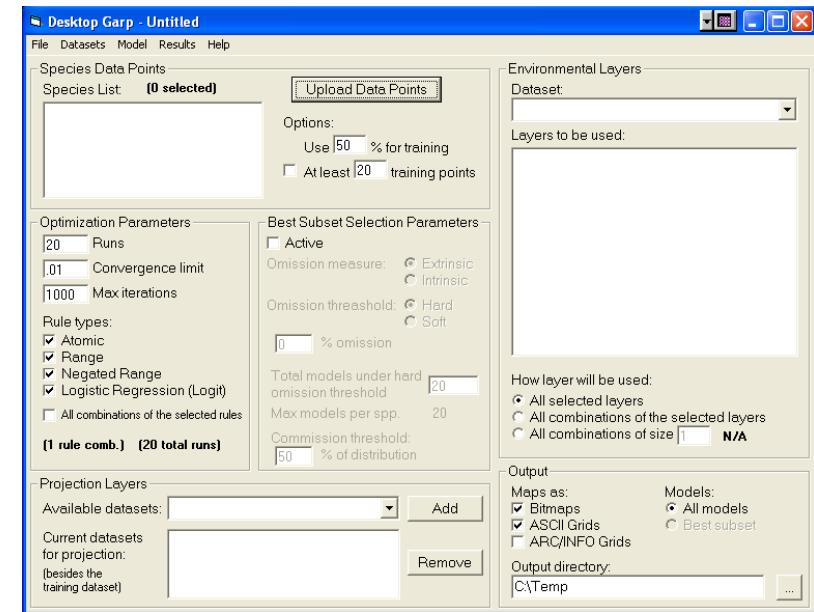
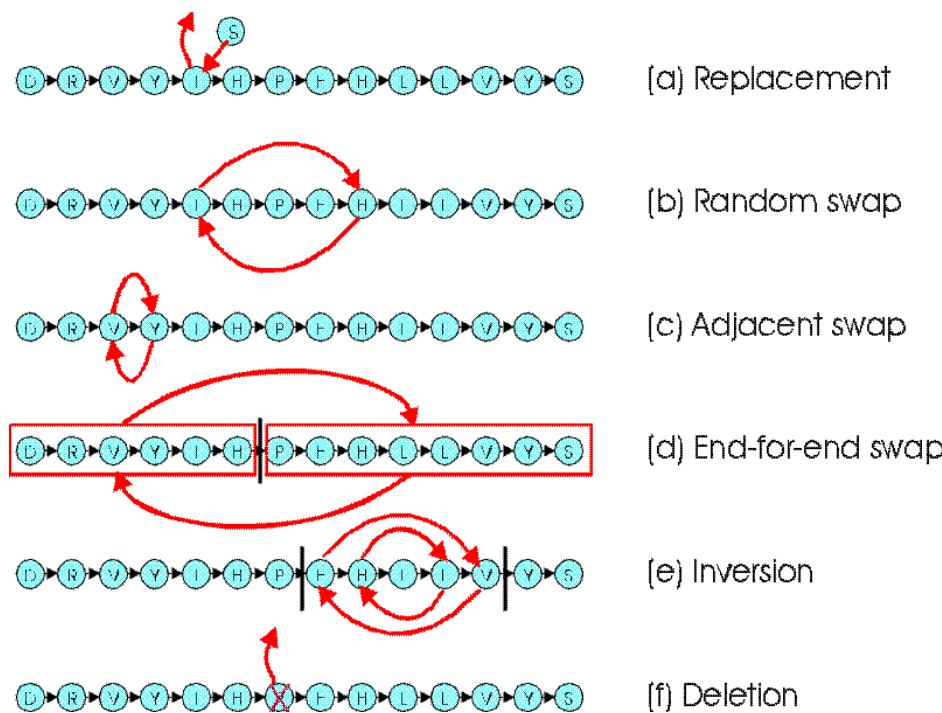
1. Arboles de regresión (árboles de decisión).
2. Boosting (combina los modelos).



**Algoritmos que usan datos de  
presencia-pseudoausencia**

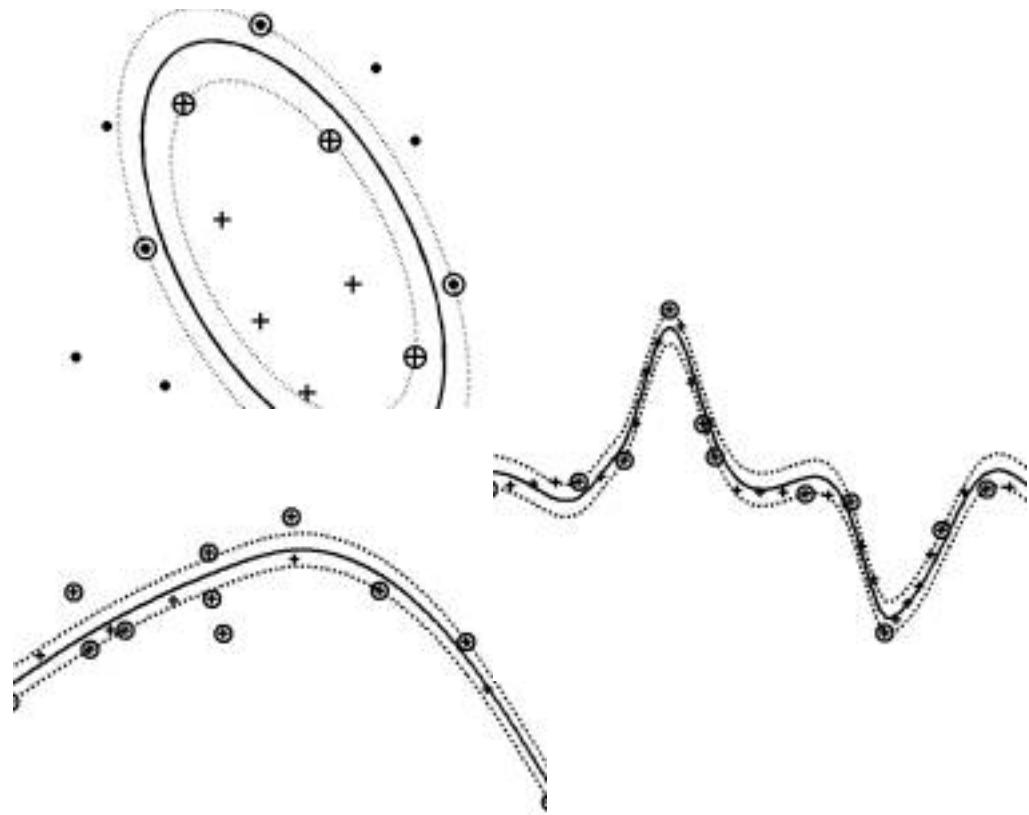
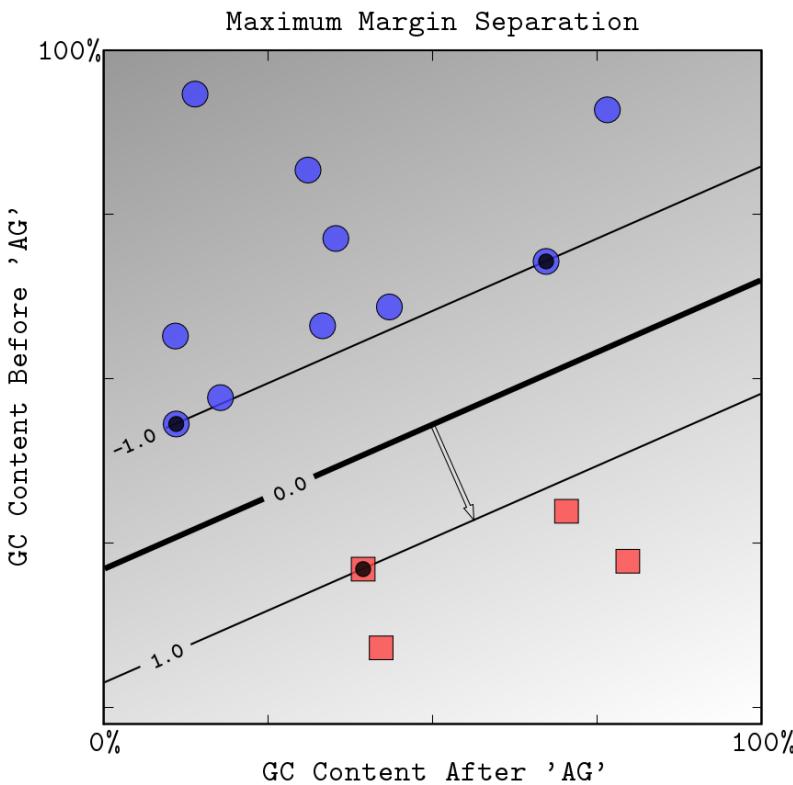
## Genetic Algorithm for Rule-set Prediction (GARP)

Es un método heurístico de aprendizaje de máquina que usa un algoritmo genético para construir “reglas” que describen el nicho de la especie con base en el principio de selección natural. Es un meta-algoritmo porque usa métodos de bioclim y regresión logística para construir las reglas iniciales



# Support Vector Machine

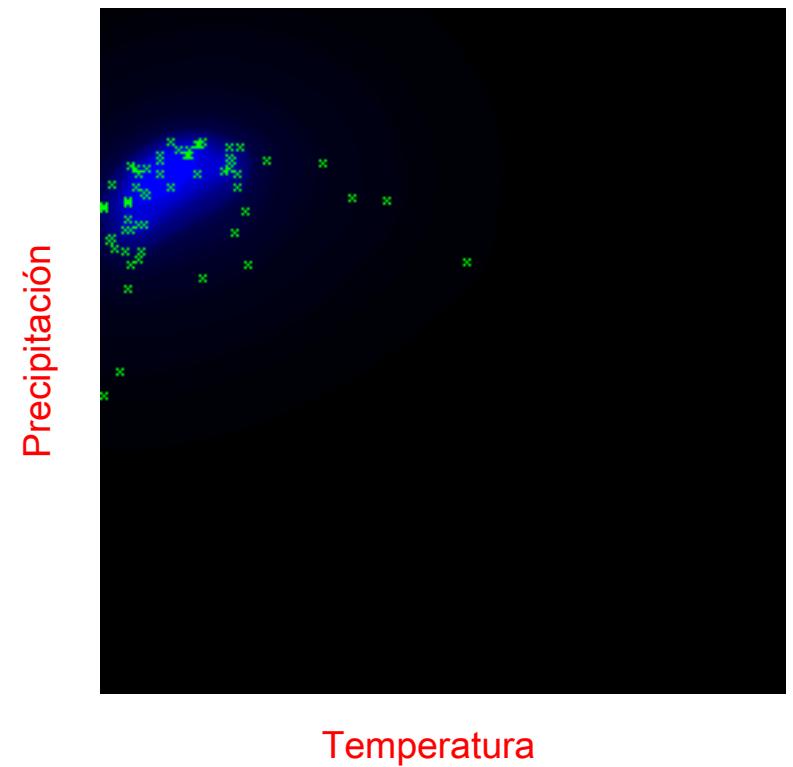
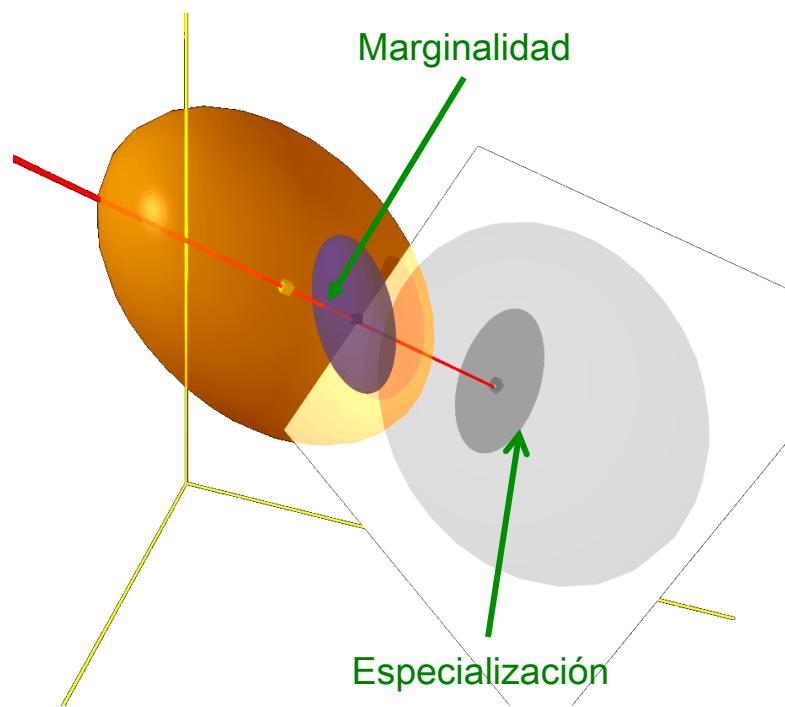
Es un método heurístico de clasificación no lineal que se basa en estadística y teoría de la optimización. Es un método de aprendizaje de máquina que es útil para cuando se tienen sólo datos de presencia o de presencia y ausencia.



# Algoritmos que usan datos de presencia-entorno

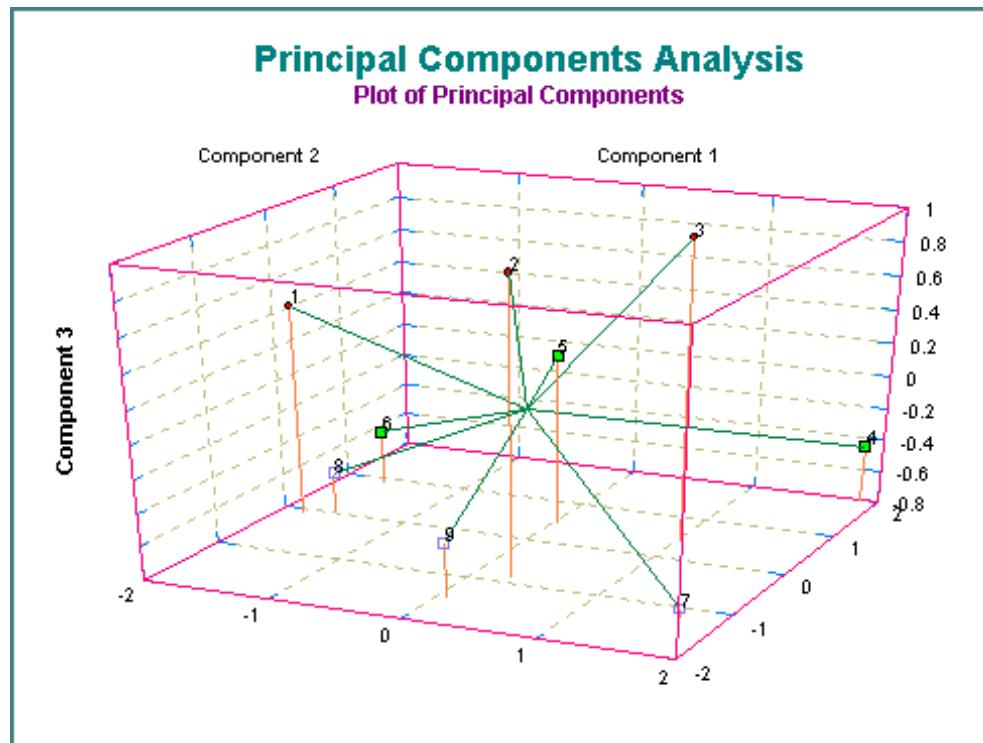
## Ecological Niche Factor Analysis (ENFA)

Implementa un análisis de factores para identificar la “marginalidad” y la “especialización” de los datos con respecto al área de análisis (M)



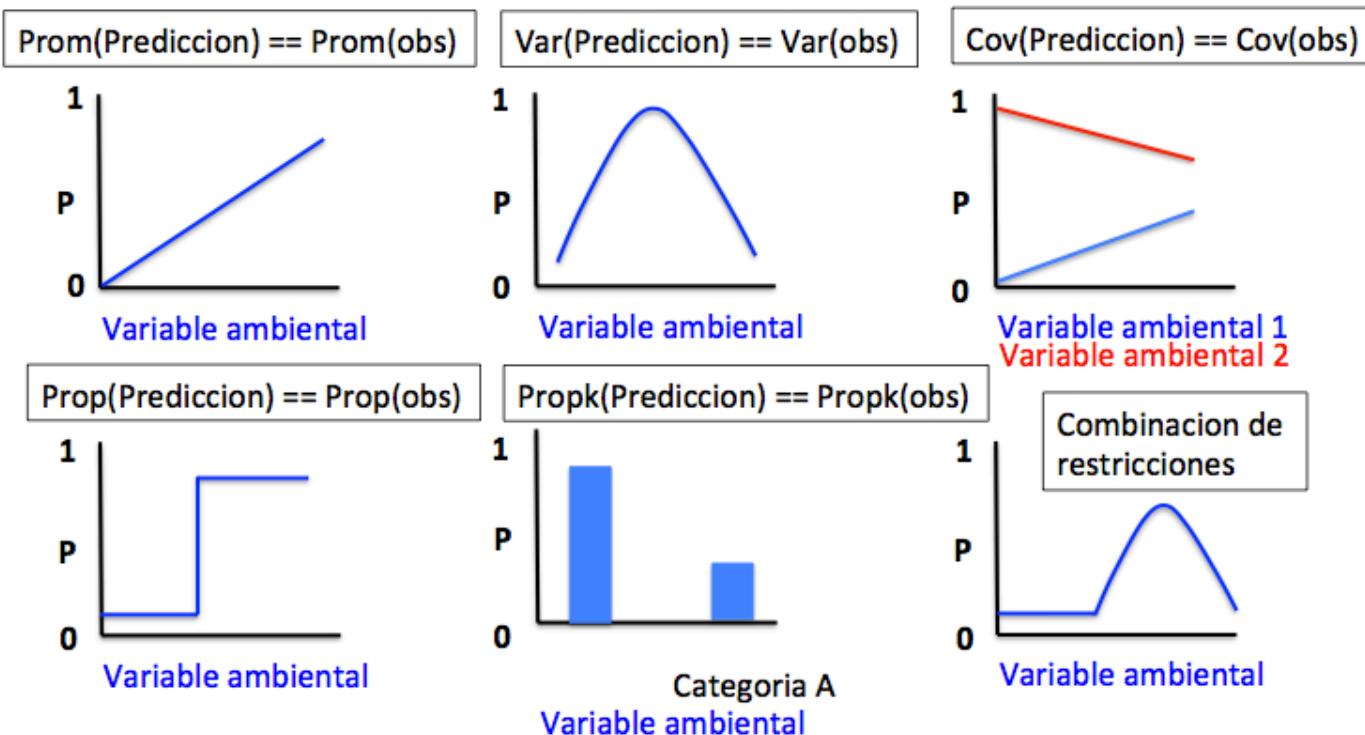
## Climate Space Model

Implementa un análisis de componentes principales (PCA)

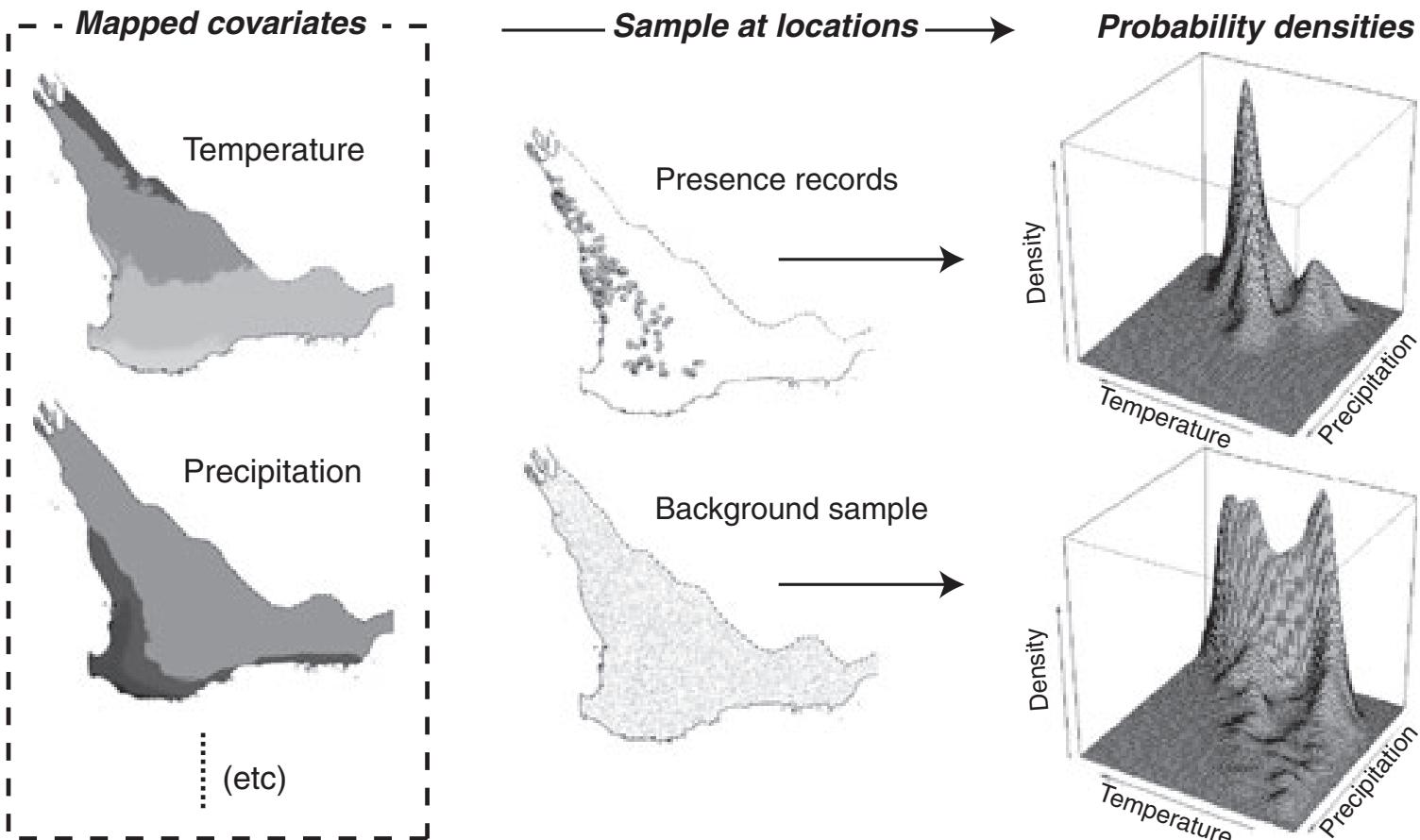


# Máxima Entropía (MaxEnt)

Es una técnica de aprendizaje de máquina que combina estadística, máxima entropía y métodos bayesianos, cuyo propósito es estimar distribuciones de probabilidad sujeto a restricciones dadas por la información ambiental.

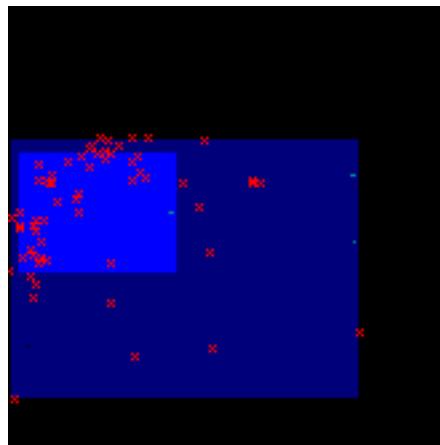


## Máxima Entropía (MaxEnt)

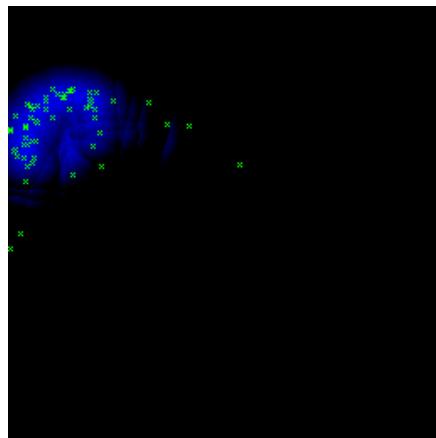


# Algoritmos

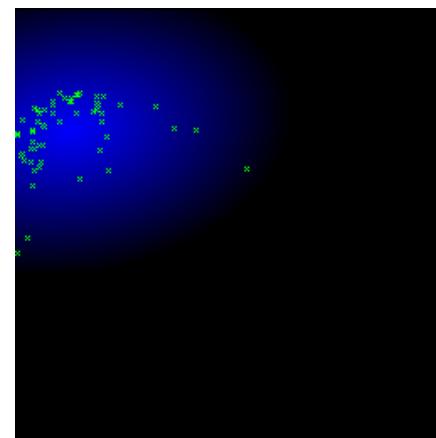
Bioclim



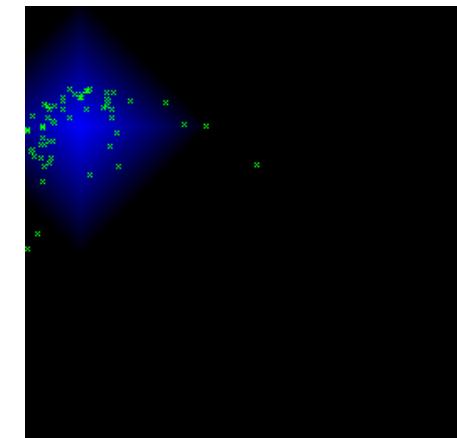
D. Euclidiana



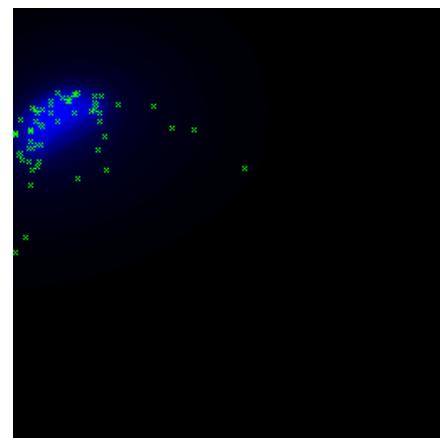
D. Mahalanobis



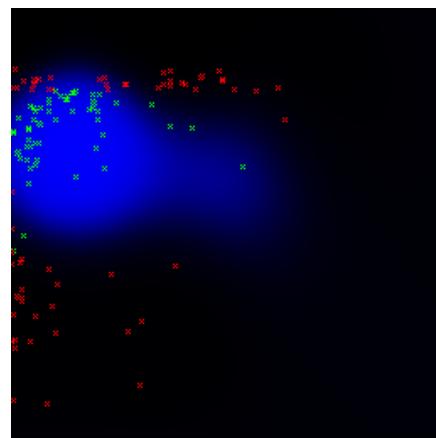
D. Gower



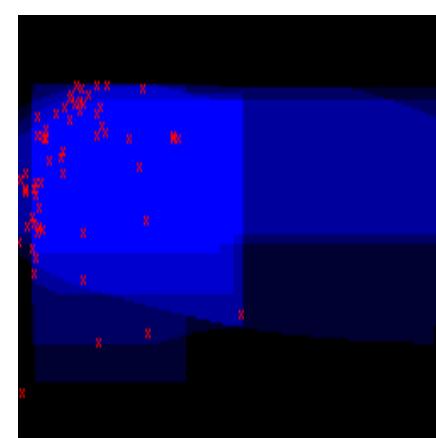
ENFA



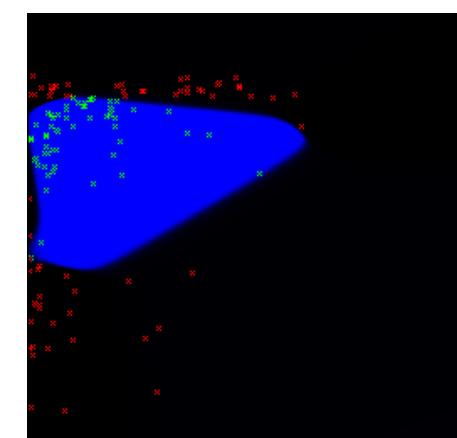
Support Vector Machine



GARP-BS



Red Neuronal (ANN)



**¿Y cuál uso?**

**Hay varios estudios  
en donde comparan  
el desempeño de  
los algoritmos**

ECOSPHERE

MaxEnt versus MaxLike:  
empirical comparisons with ant species distributions

MATTHEW C. FITZPATRICK,<sup>1,†</sup> NICHOLAS J. GOTELLI,<sup>2</sup> AND AARON M. ELLISON<sup>3</sup>

PLoS One. 2013; 8(7): e68823.

PMCID: PMC3706317

Published online Jul 9, 2013. doi: [10.1371/journal.pone.0068823](https://doi.org/10.1371/journal.pone.0068823)

**How Can Model Comparison Help Improving Species Distribution Models?**

[Emmanuel Stephan Gritti](#),<sup>1,2,\*</sup> [Cédric Gaucherel](#),<sup>3,4</sup> [Maria-Veronica Crespo-Perez](#),<sup>1</sup> and [Isabelle Chuine](#)<sup>1</sup>

*Journal of Biogeography (J. Biogeogr.)* (2007) **34**, 1455–1469

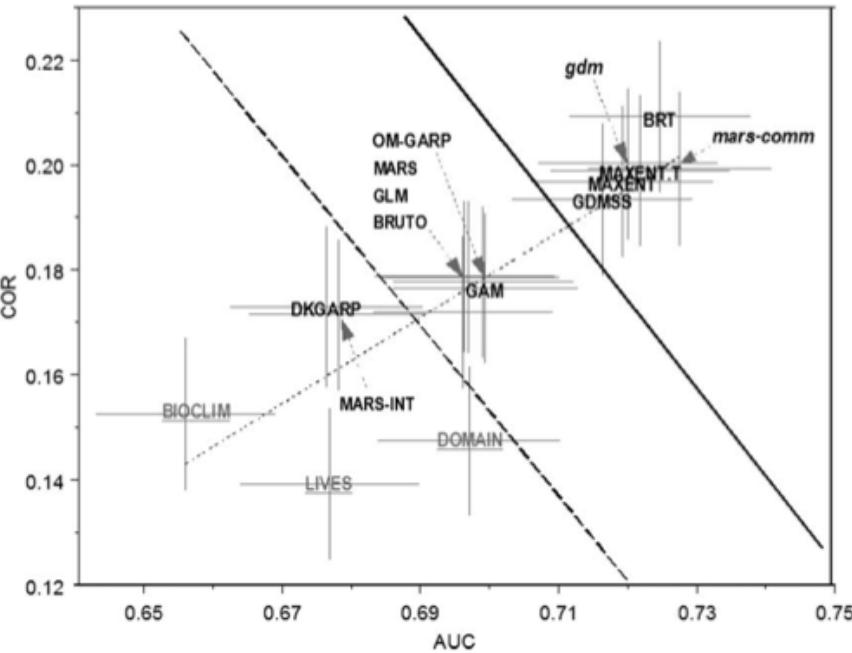


**Predicting species distributions: a critical comparison of the most common statistical models using artificial species**

Christine N. Meynard<sup>1,2,\*</sup> and James F. Quinn<sup>1</sup>

# ¿Qué método es mejor?

Novel methods improve prediction of species' distributions from occurrence data



A la fecha no existe un algoritmo que sea mejor que los demás para todos los tipos de datos. Algunos funcionan mejor cuando el número y la calidad de registros es baja y otros son más robustos cuando la calidad de los datos mejora.

Jane Elith\*, Catherine H. Graham\*, Robert P. Anderson, Miroslav Dudik, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Schachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz and Niklaus E. Zimmermann

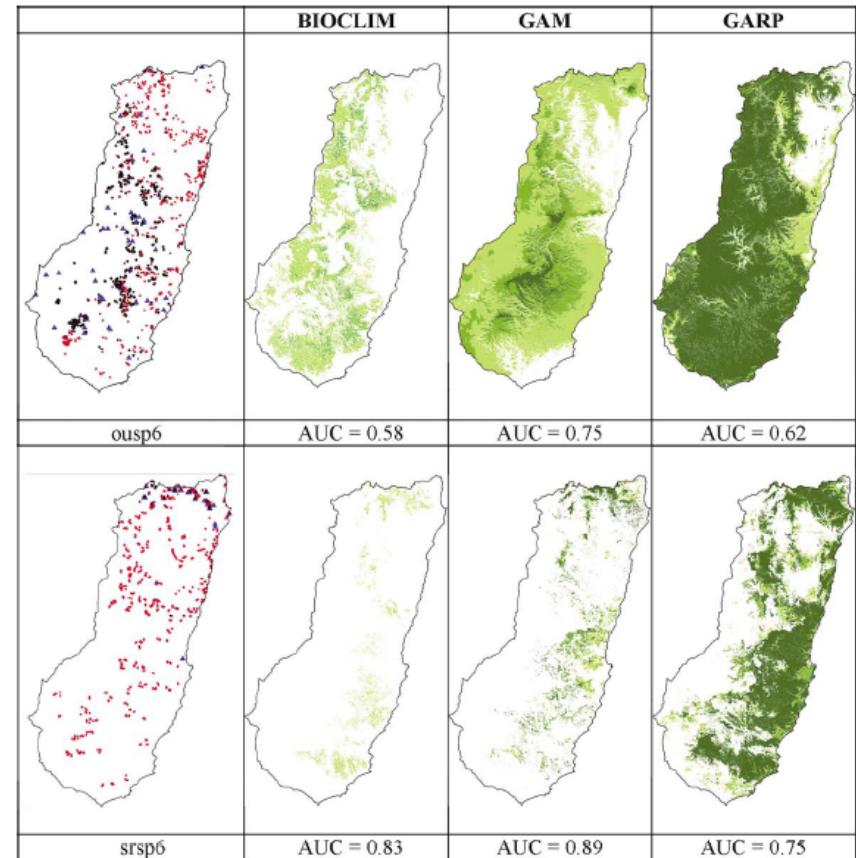


Fig. 1. Maps for two species from NSW for each of three selected techniques. Details: ousp6, *Poa sibiriana* (53 records for modelling and 512 presence/797 absence for evaluation); srsp6 *Ophiocincus truncatus* (79 model, 74/932 eval). The first column shows modelling sites (grey triangles) and evaluation sites: presence = black circle, absence = black cross

# Además, su comportamiento es distinto cuando los modelos se transfieren en espacio o tiempo

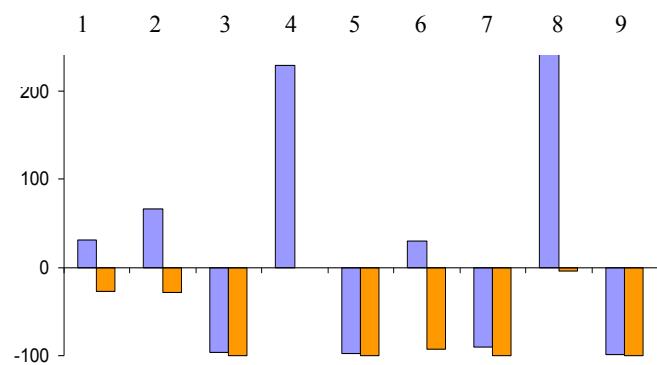
Journal of Biogeography (J. Biogeogr.) (2006) 33, 1704–1711



## Model-based uncertainty in species range prediction

Richard G. Pearson<sup>1\*</sup>, Wilfried Thuiller<sup>2</sup>, Miguel B. Araújo<sup>3,4†</sup>, Enrique Martinez-Meyer<sup>5</sup>, Lluís Brotons<sup>6</sup>, Colin McClean<sup>7</sup>, Lera Miles<sup>8</sup>, Pedro Segurado<sup>9</sup>, Terence P. Dawson<sup>10</sup> and David C. Lees<sup>11</sup>

*C. Leucosperma tomentosum*



1. ANN1 (SPECIES)
2. ANN2 (PLUS)
3. BIOCLIM
4. CTA
5. GA
6. GAM
7. GARP
8. GLM
9. DOMAIN

# Evaluación de la Incertidumbre

Se ha propuesto al ensamble de modelos para reducir la incertidumbre



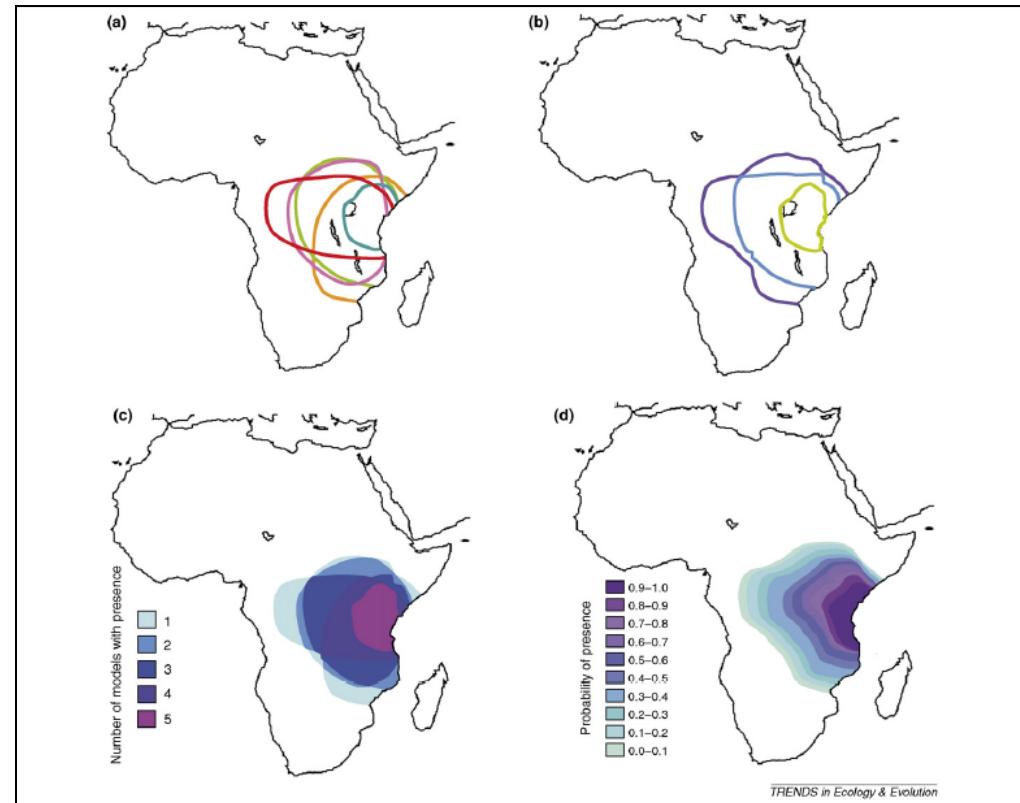
Review

TRENDS in Ecology and Evolution Vol.22 No.1

Full text provided by www.sciencedirect.com  
ScienceDirect

## Ensemble forecasting of species distributions

Miguel B. Araújo<sup>1</sup> and Mark New<sup>2</sup>



Es práctica común que la decisión del método de modelación se base en criterios ajenos a los datos; por ejemplo, por moda, por escuela, por dogma de fe, por las (in)habilidades personales.

La modelación de nichos y distribuciones no es como una técnica de laboratorio en donde se tienen protocolos, dosis y recetas ya estandarizados, y seguramente no los habrá.

# En conclusión

- La selección del método de modelación es un paso crítico en el proceso
- Para tomar una decisión es necesario conocer el tipo de datos con los que uno cuenta y el funcionamiento de los algoritmos de modelado
- Probar el desempeño de más de un método es una buena práctica, pero hay que ser cuidadosos en el proceso de evaluación