



Grupo de investigación Ecología de Zonas Áridas

**CENTRO ANDALUZ PARA LA EVALUACIÓN Y  
SEGUIMIENTO DEL CAMBIO GLOBAL**



Modelling  
Workshops



# Funcionamiento, interfaz y formato de los datos en MaxEnt

**Elisa Liras**

Dpto. Biología Vegetal y Ecología  
Universidad de Almería  
eliras@ual.es

### ¿Qué es MaxEnt?

MaxEnt, un programa basado en una distribución de Máxima Entropía para la modelización de la distribución geográfica de las especies.

Ha sido escrito originalmente por Steven Phillips, Miro Dudik y Rob Schapire, con el respaldo de AT&T Labs-Research, Princeton University, y e Centro para la Biodiversidad y Conservación, Museo de Historia Natural de América.

### ¿Y qué hace?

MaxEnt empieza ajustando los datos a partir de una distribución uniforme, que va modificando hasta una distribución de Máxima Entropía. El modelo se basa en ajustar los parámetros de la distribución final.

### ¿Qué necesitamos?

- 1) **conjunto de localidades** (puntos) donde se sabe que la especie está presente
- 2) **coberturas geográficas** → parámetros ambientales que pueden, potencialmente, limitar la capacidad de supervivencia de la especie.

### ¿Qué significa Máxima Entropía?

Entropía = cantidad de información contenida por un sistema

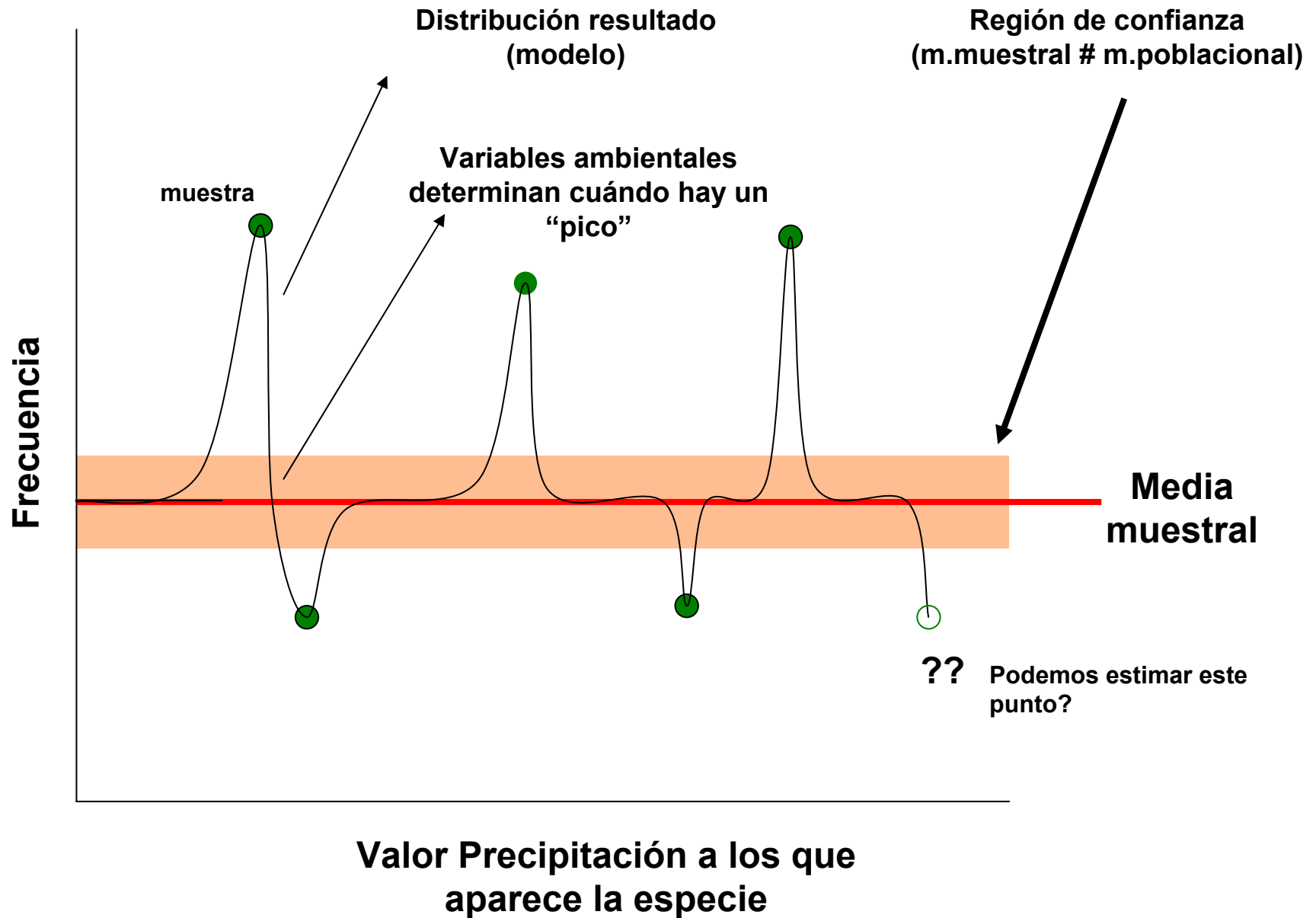
Máxima entropía = distribución más extendida, cercana a uniforme

### Entonces...

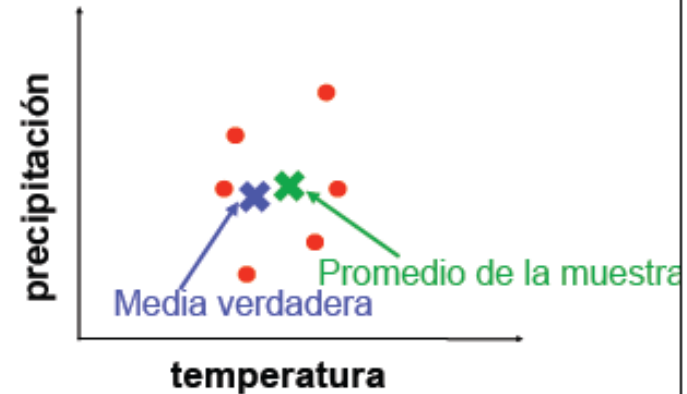
El método de Máxima Entropía: para estimar una distribución desconocida:

- Determinar lo que se conoce (restricciones)
- Entre las predicciones que se ajustan a las restricciones se favorece la que tiene máxima entropía
- Las restricciones están impuestas por los valores de las variables ambientales en las localidades conocidas de la especie.

# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

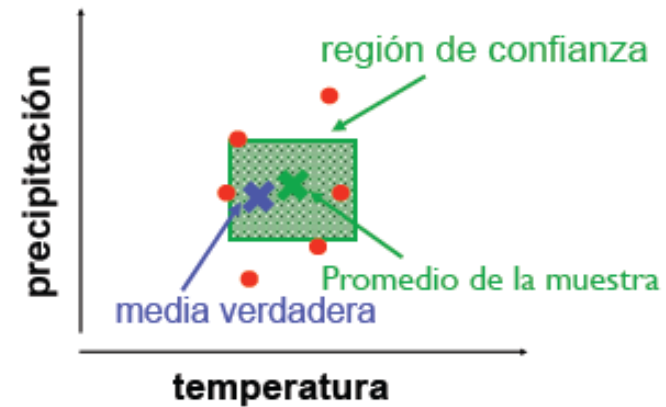


## Las variables ambientales imponen las restricciones



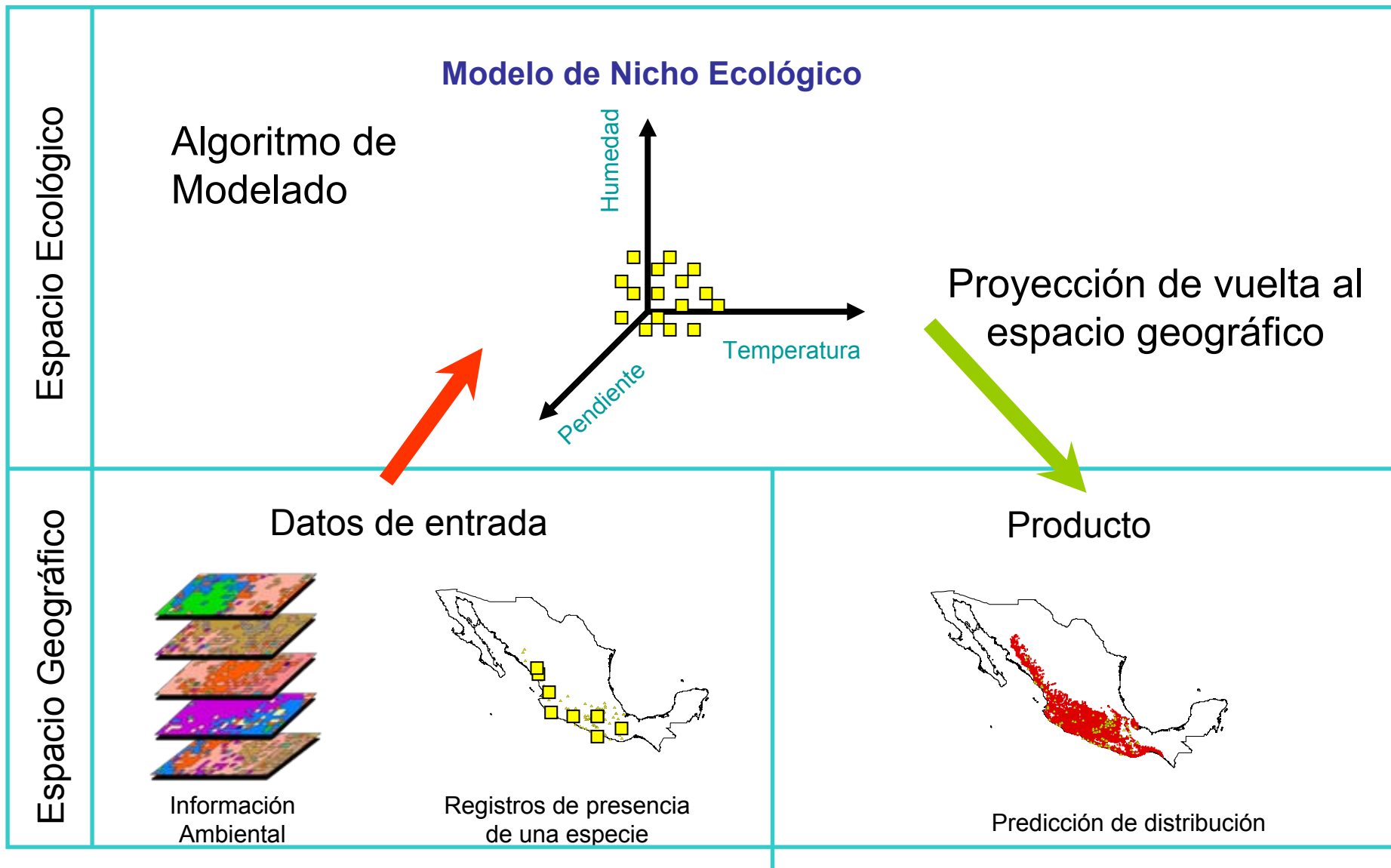
Encuentra la distribución  $\mathbf{p}$  de **máxima entropía** tal que para todos los atributos  $\mathbf{f}$ :  $E_{\mathbf{p}}[\mathbf{f}] =$  promedio de la muestra de  $\mathbf{f}$

## Regularización



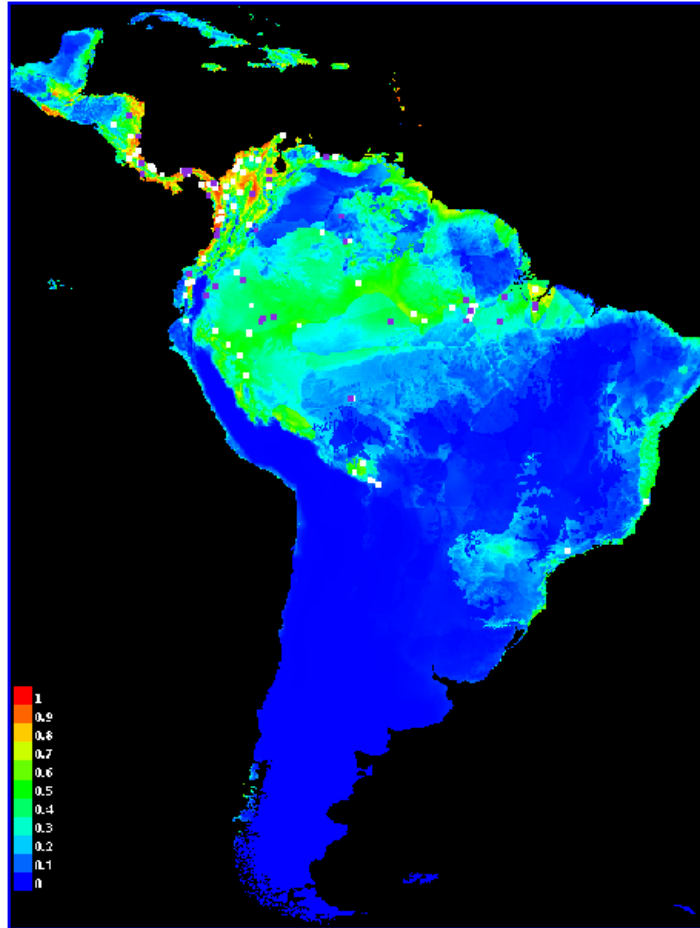
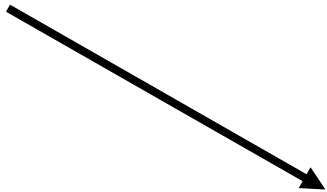
Encuentra la distribución  $\mathbf{p}$  de **máxima entropía** tal que  $E_{\mathbf{p}}[\mathbf{f}]$  en la **región de confianza** del promedio de la muestra de  $\mathbf{f}$

## Modelos de nicho:



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

Superficie de probabilidad



*Bradypus variegatus*

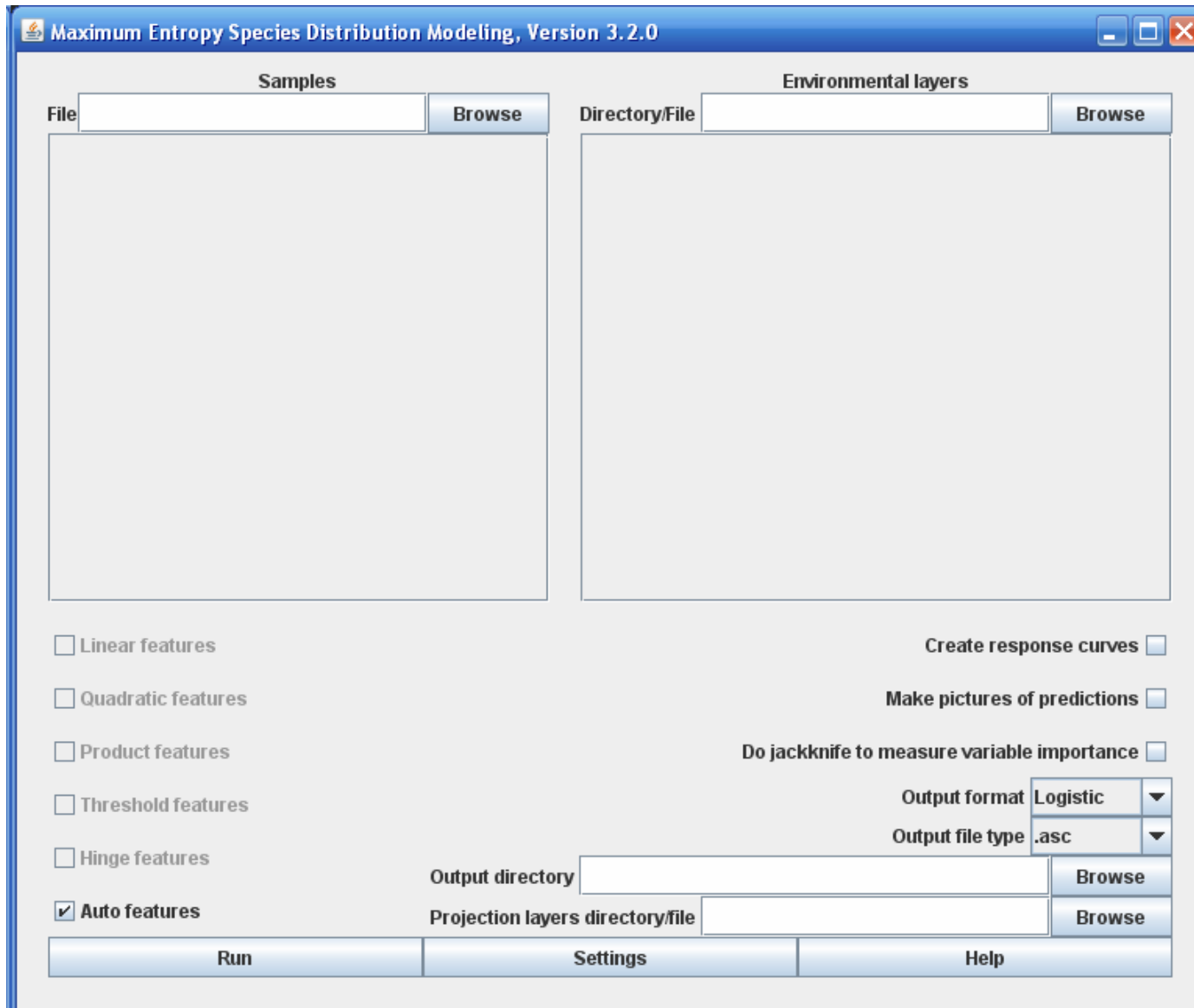


### Ventajas de MaxEnt:

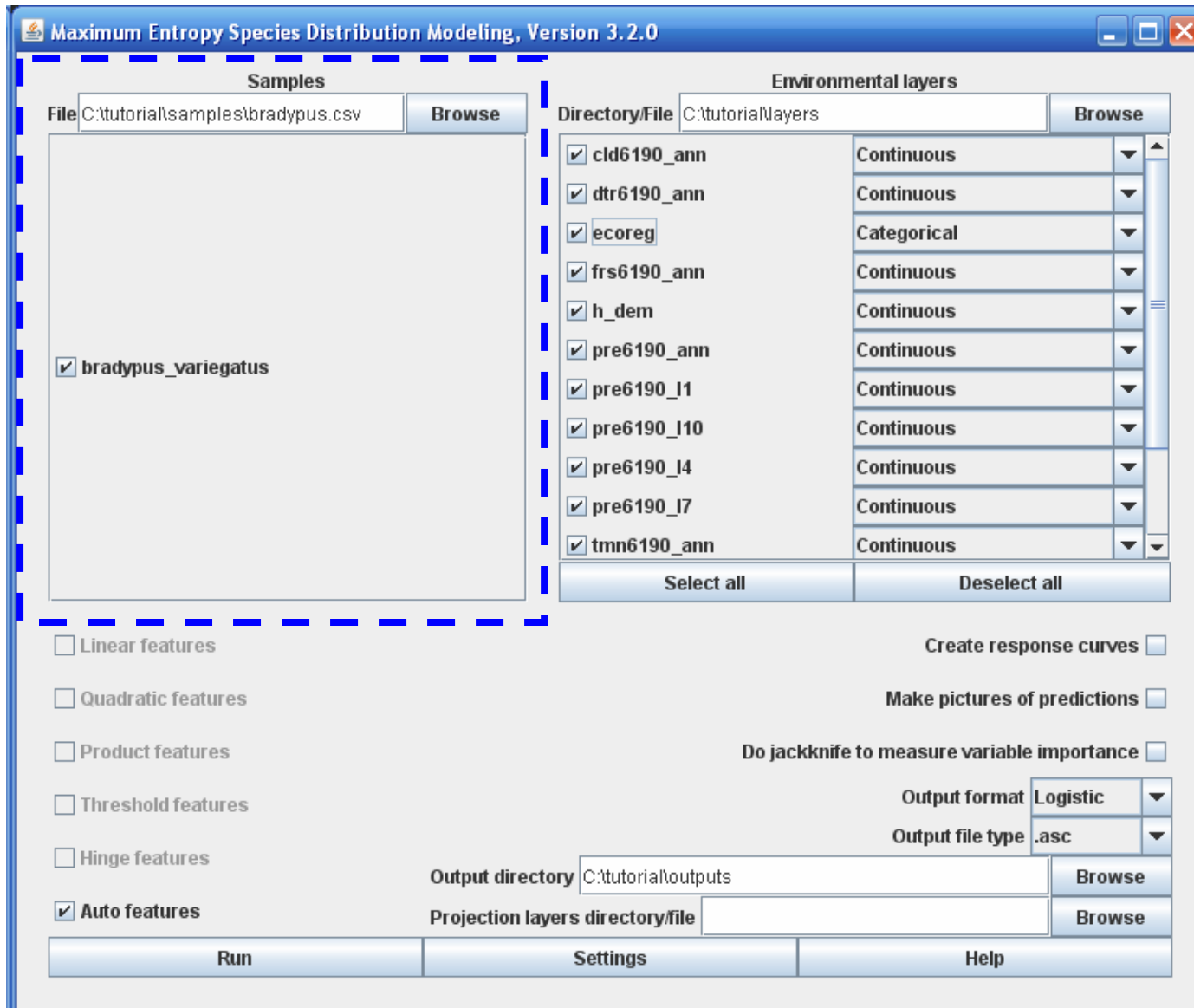
- Datos ambientales continuos y categóricos (trabaja con la frecuencia de aparición)
- Resultado continuo
- Comportamiento determinista (repetible)
- Capacidad interpretativa en dimensiones ecológicas (curvas de respuesta)
- Rápido
- Exacto (se ajusta a los datos)

*Steven Phillips, Miro Dudik & Rob Schapire*

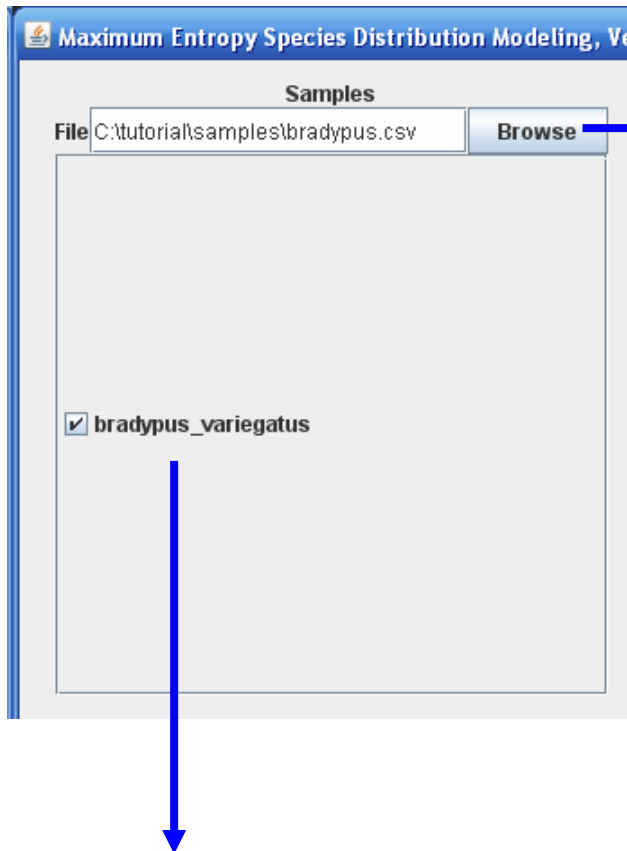
# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



Especificar archivo de datos

### Archivo .csv

species, longitude, latitude

bradypus\_variegatus, -65.4, -10.3833

bradypus\_variegatus, -65.3833, -10.3833

bradypus\_variegatus, -65.1333, -16.8

bradypus\_variegatus, -63.6667, -17.45

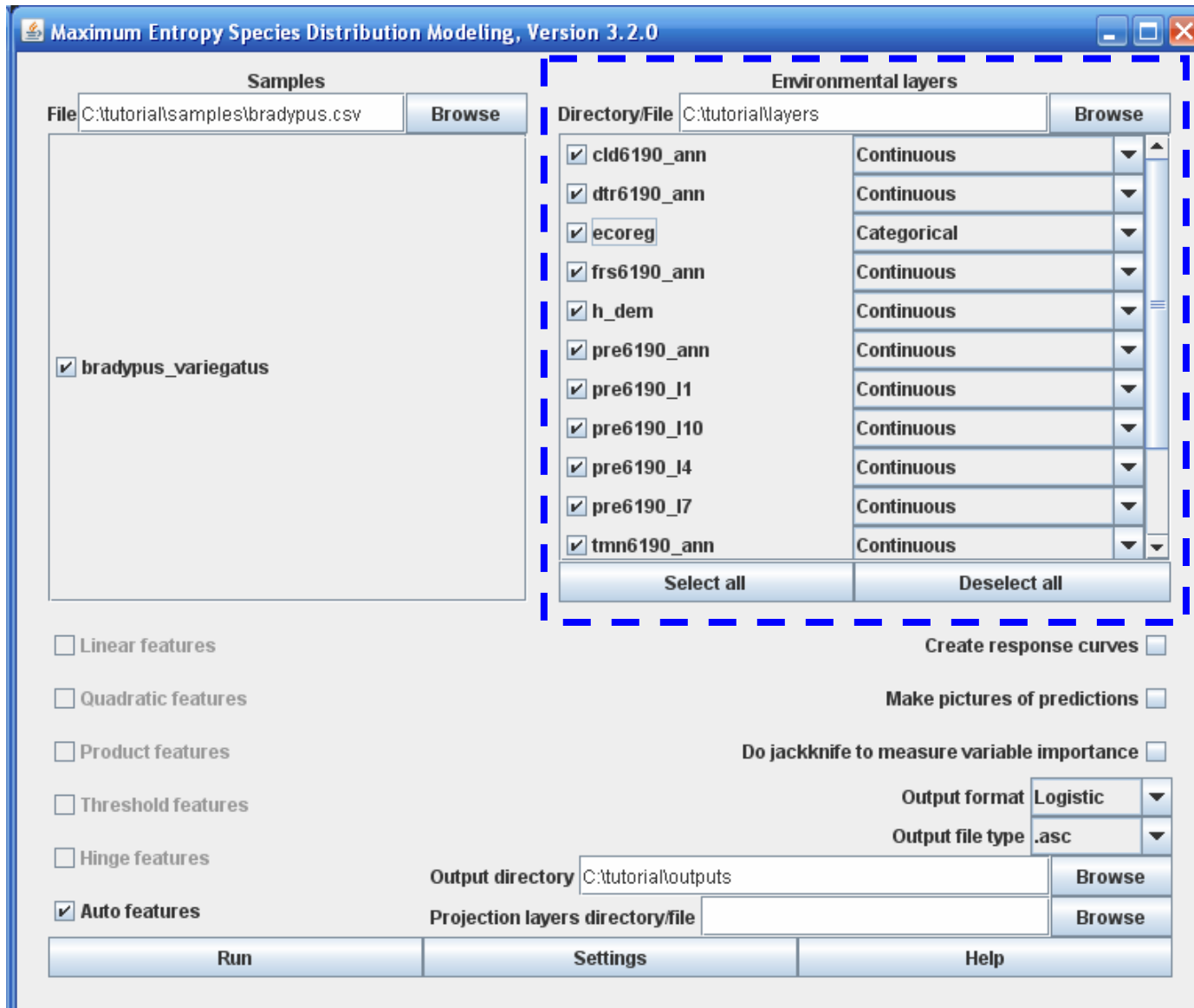
bradypus\_variegatus, -63.85, -17.4

Si hubiera más de una especie, lo veríamos aquí

(se pueden seleccionar o no, dependiendo de si queremos incluirlas en el modelo)

Opción: “Settings → Delete duplicates”.

# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



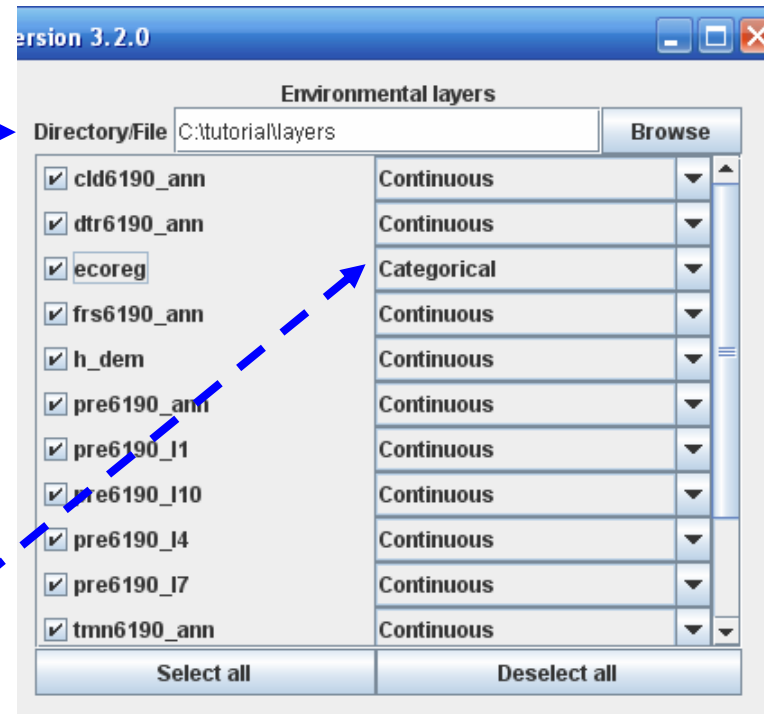
# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

Especificar archivo de datos

Formato de coberturas:  
ASCII grid de ESRI  
Hay que especificar si la  
variable es continua o  
categórica!!

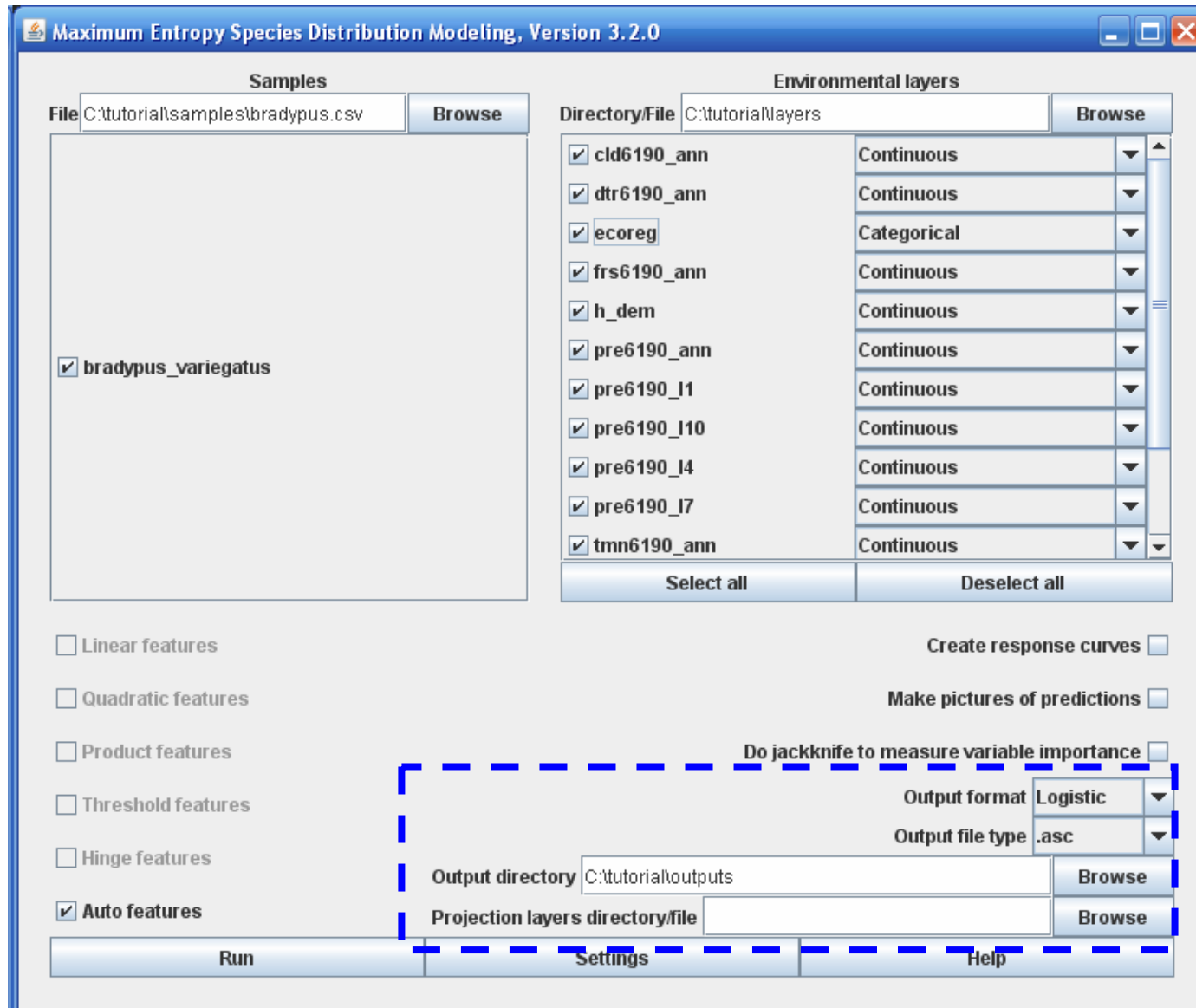
Variables categóricas:

indicadas preferiblemente por  
números (mejor que por letras o  
palabras).



(las coberturas se pueden  
seleccionar o no, dependiendo de si  
queremos incluirlas en el modelo)

# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

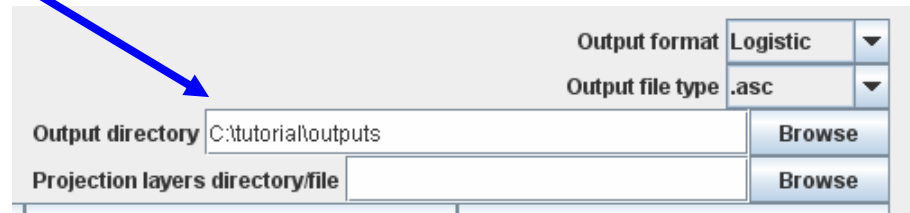


## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

Hay 3 tipos de formato de resultados  
(diferencias más adelante)

Especificar archivo de salida

Especificar archivo con las coberturas de proyección

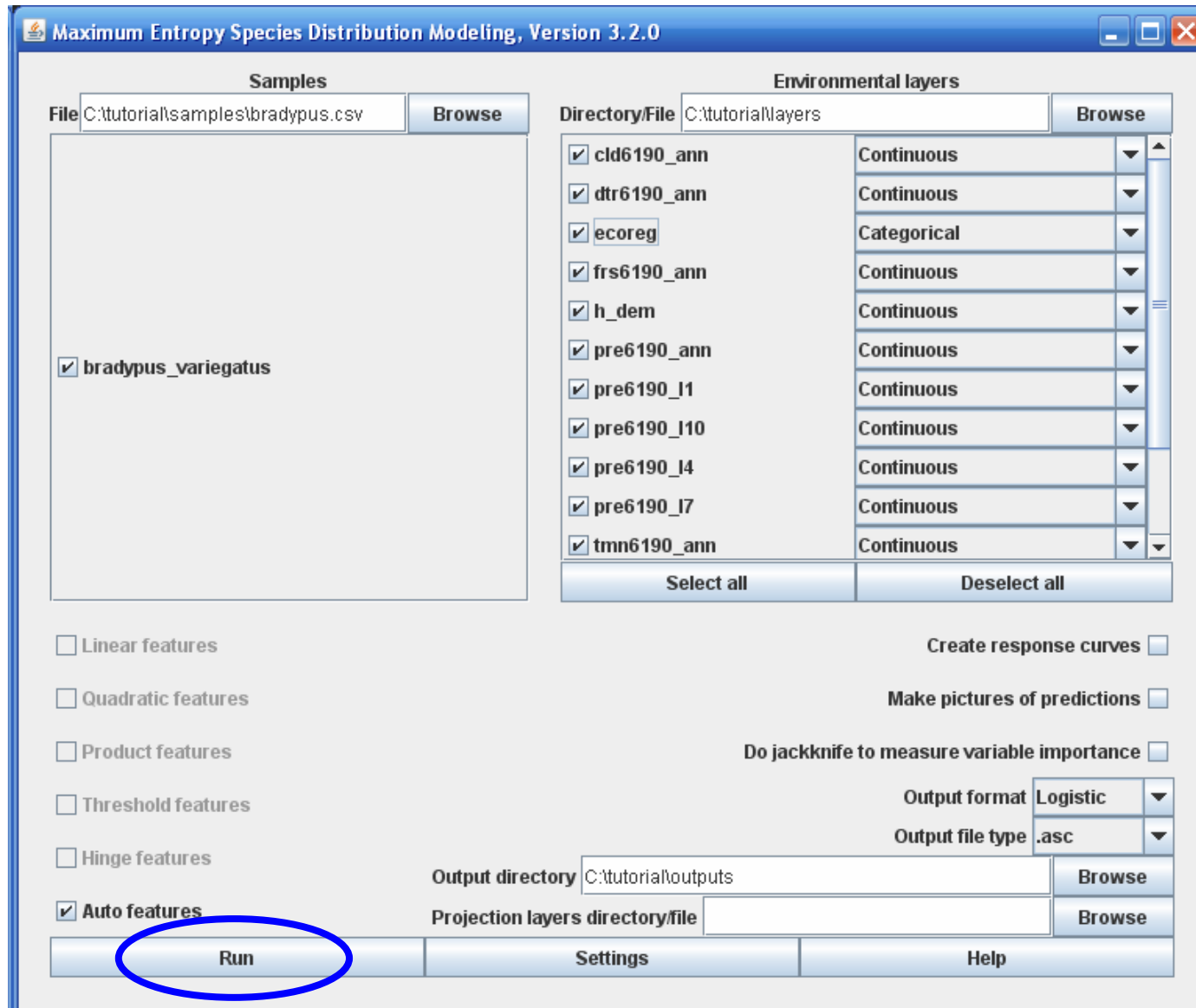


The screenshot shows the MaxEnt software interface with the following fields and controls:

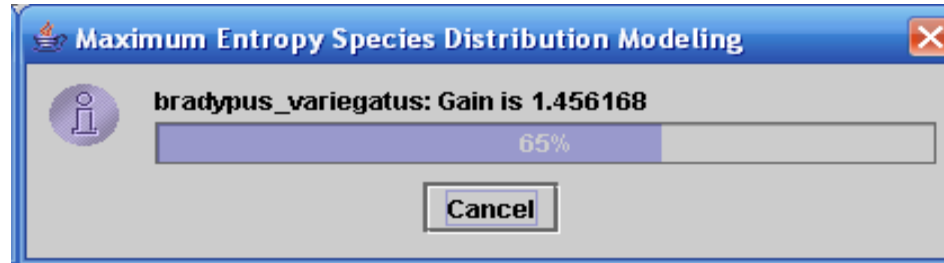
- Output format:** A dropdown menu currently set to "Logistic".
- Output file type:** A dropdown menu currently set to ".asc".
- Output directory:** A text input field containing "C:\tutorial\outputs" and a "Browse" button to the right.
- Projection layers directory/file:** A text input field and a "Browse" button to the right.



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

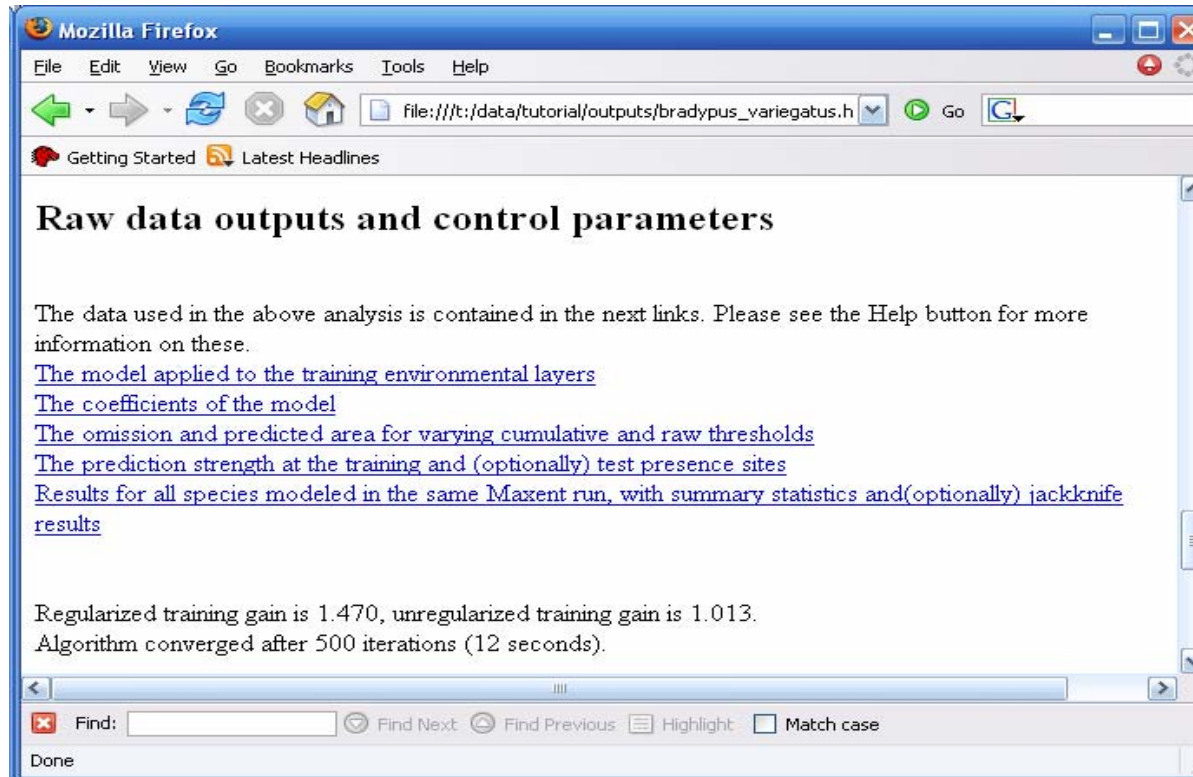


## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

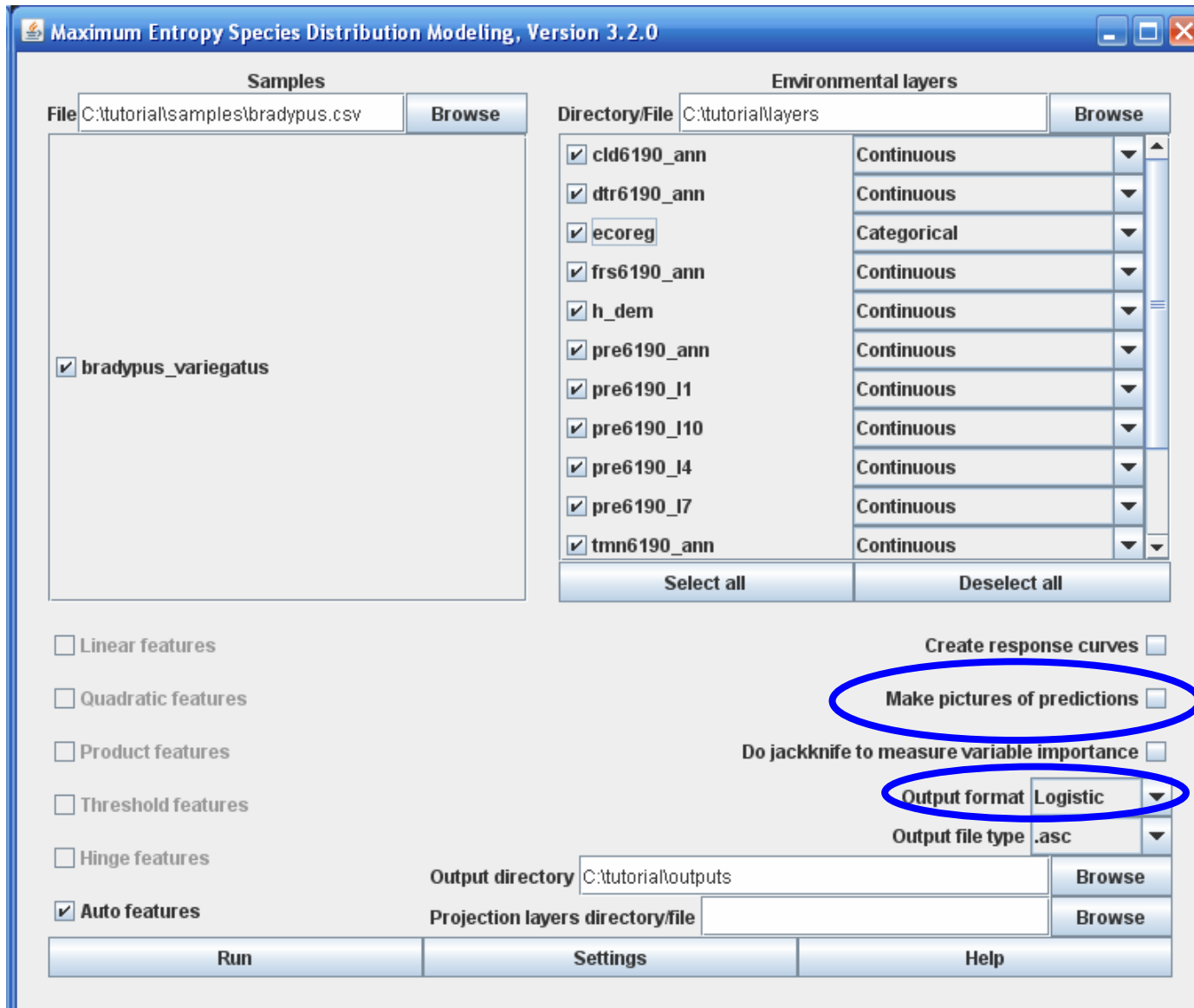


- El “gain” está relacionado con la desviación, una **medida de la bondad de ajuste** utilizada en GAM y GLM.
- Esta medida **empieza en el valor 0 y va aumentando** asintóticamente durante el proceso de modelización.
- Durante este proceso, MaxEnt está generando una distribución de probabilidad sobre los píxeles de la grid, empezando por una distribución uniforme y, de forma iterativa, mejorando el ajuste de la distribución a los datos de muestreo.
- Al final del proceso, **el valor del gain indica cómo de ajustado está el modelo a las muestras de presencia**; por ejemplo, si el gain toma valor de 2, ello significa que el valor medio de los píxeles que contienen un punto de presencia es  $\exp(2) \approx 7.4$  veces mayor que una distribución al azar sobre dicho píxel.
- Nótese que **MaxEnt no está calculando directamente la “probabilidad de ocurrencia”**. La probabilidad asignada por el modelo a cada píxel es, generalmente, muy pequeña, dado que la suma de los valores totales del grid debe de ser 1

### Archivos de salida (resultado de MaxEnt): html



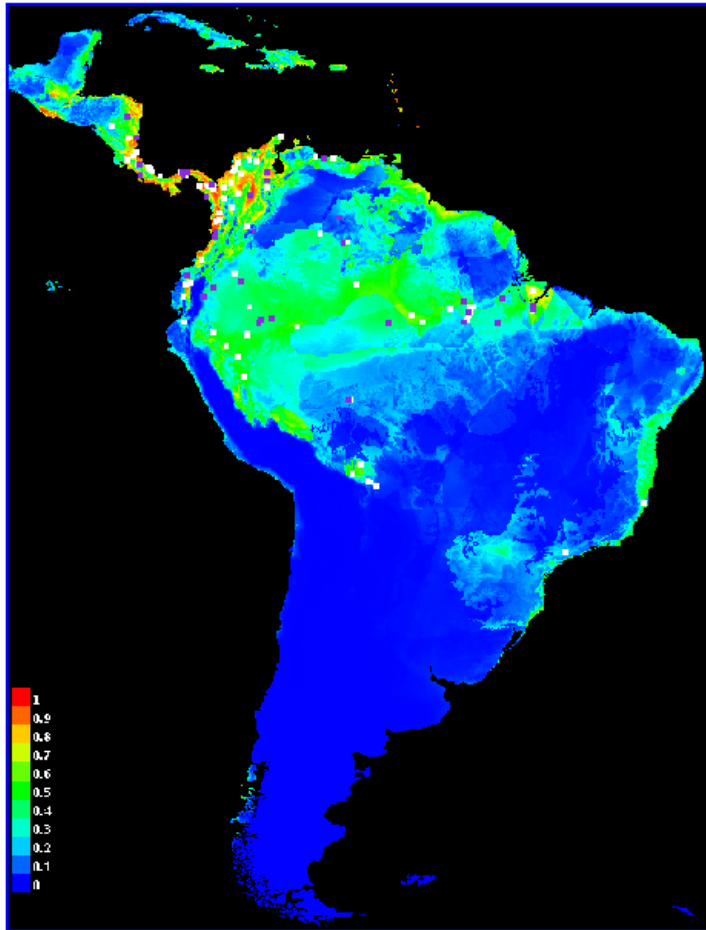
# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

## Pictures of the model

This is a representation of the Maxent model for *bradypus\_variegatus*. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.



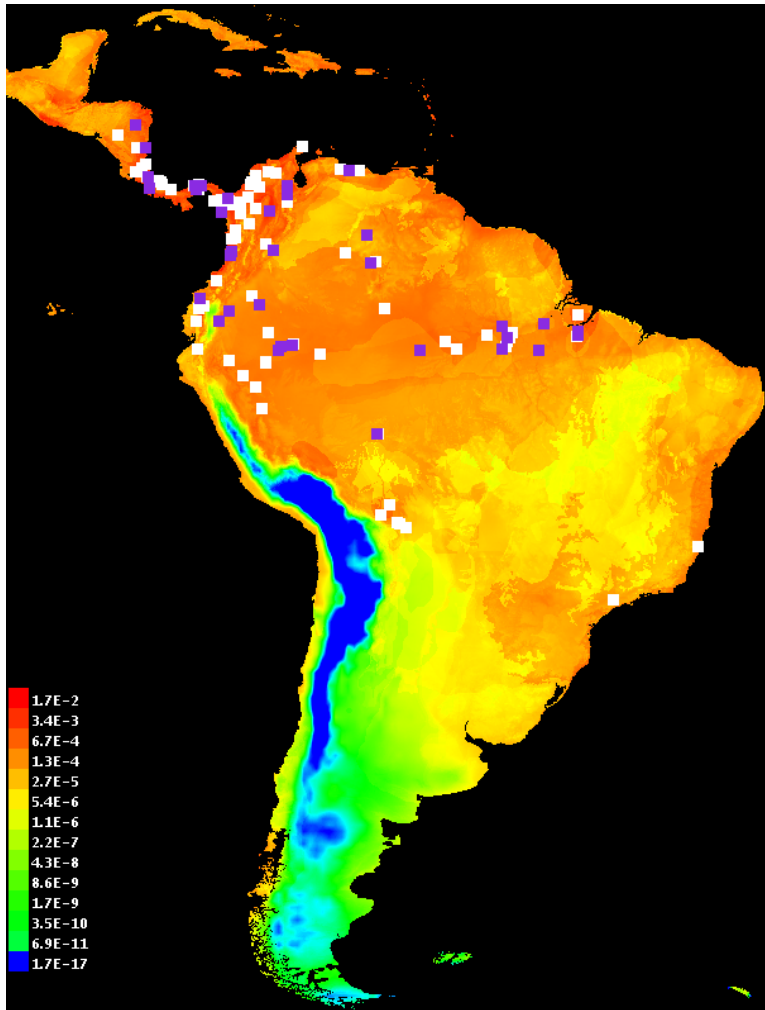
### Logístico:

**Valor por defecto**

**Da una estimación entre 0 y 1 de la probabilidad de presencia**

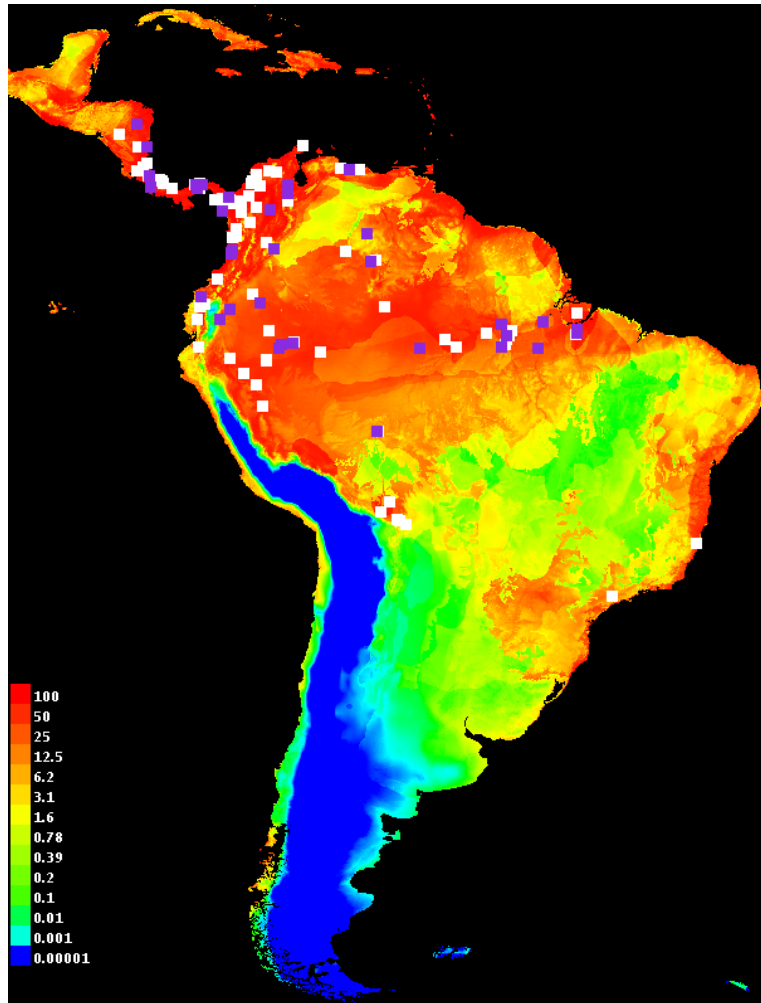
**Es el más fácil de interpretar**

## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



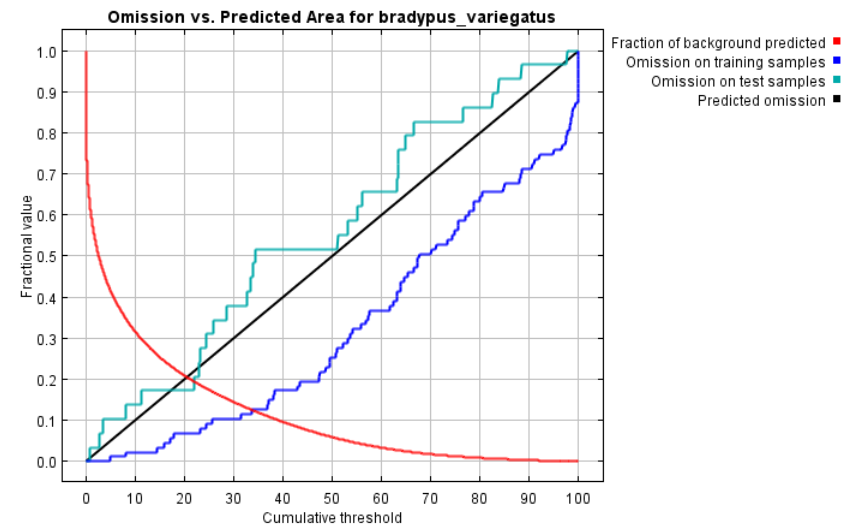
**Datos crudos (raw data):**

**Es simplemente el modelo  
exponencial de MaxEnt**

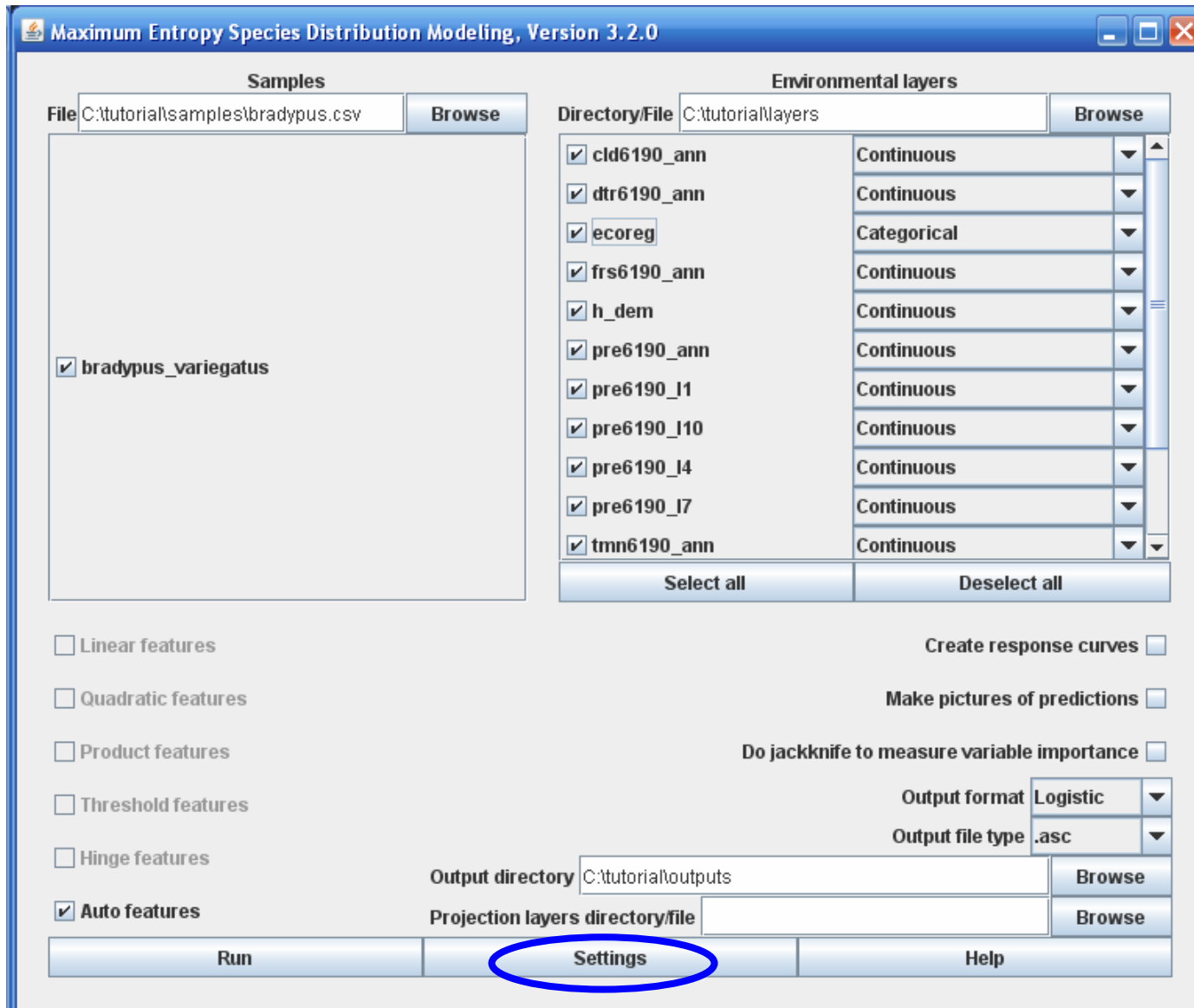


## Cumulative:

Es el resultado más fácil de interpretar en términos de tasa de predicción de omisión



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



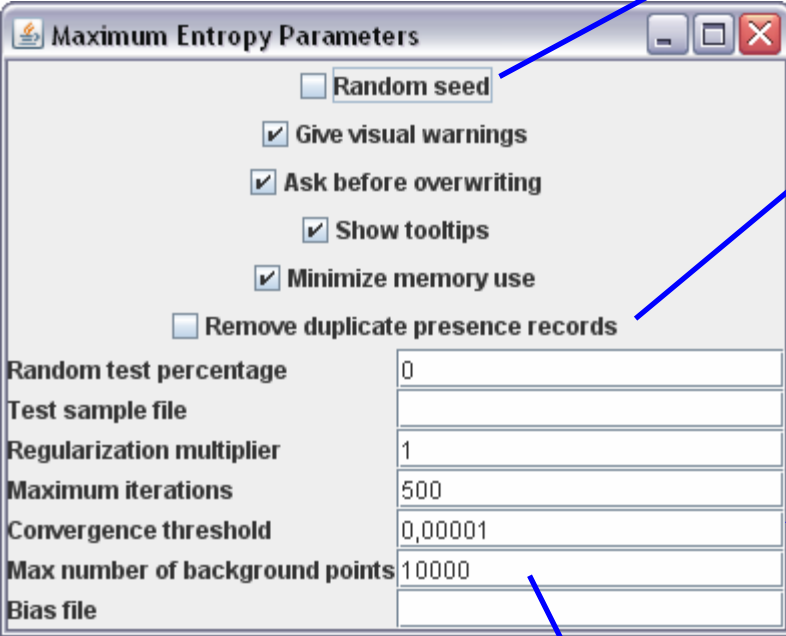


# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

% de datos que se "guardan" para realizar el test

Sin seleccionar: MaxEnt utiliza un conjunto de datos fijo

Archivo de datos independientes con los que realizar el test



Eliminar puntos duplicados

Ajuste del modelo (menor 1: más ajustado)

Límite de convergencia

Nº máximo de iteraciones

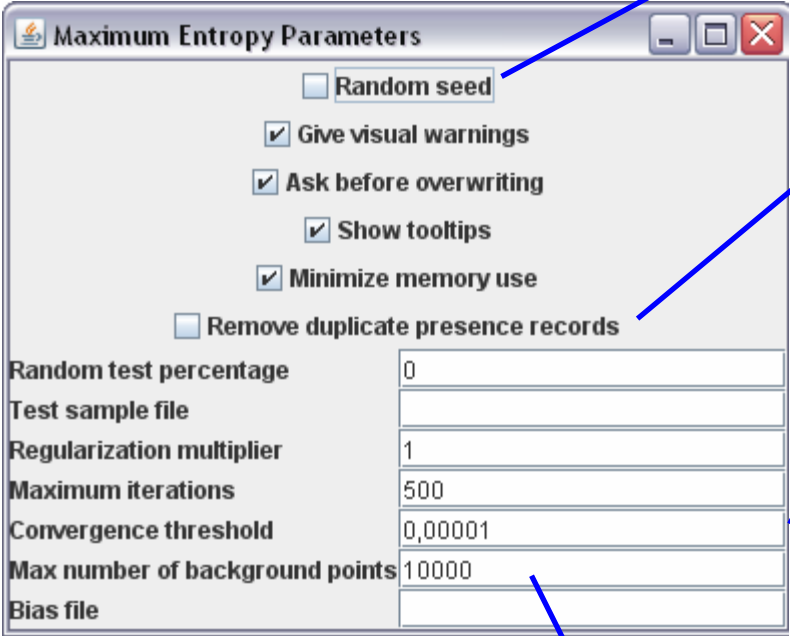
Nº máximo de "background" puntos (grid)

# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

% de datos que se "guardan" para realizar el test

Sin seleccionar: MaxEnt utiliza un conjunto de datos fijo

Archivo de datos independientes con los que realizar el test



Eliminar puntos duplicados

Ajuste del modelo (menor 1: más ajustado)

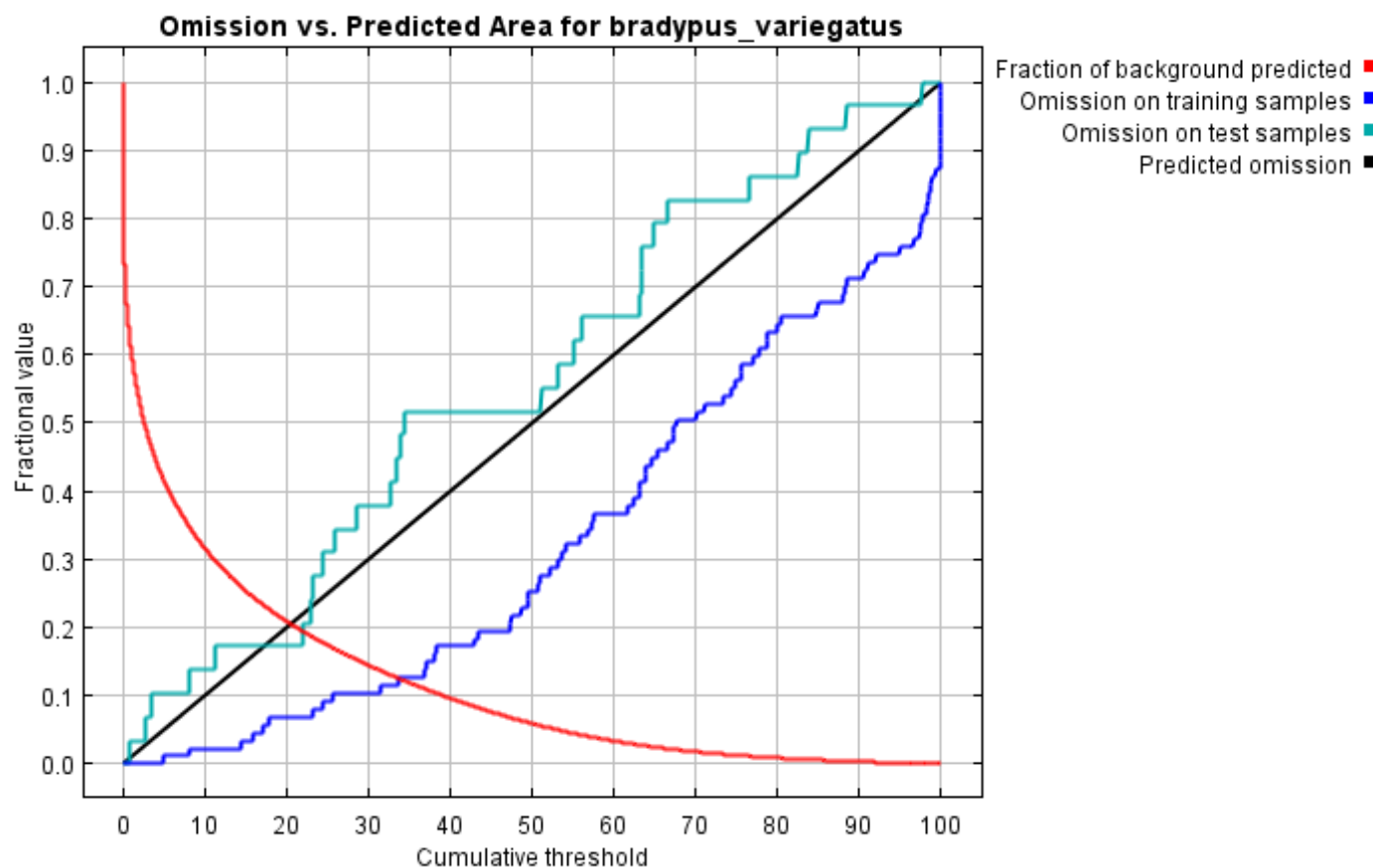
Límite de convergencia

Nº máximo de iteraciones

Nº máximo de "background" puntos (grid)

## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

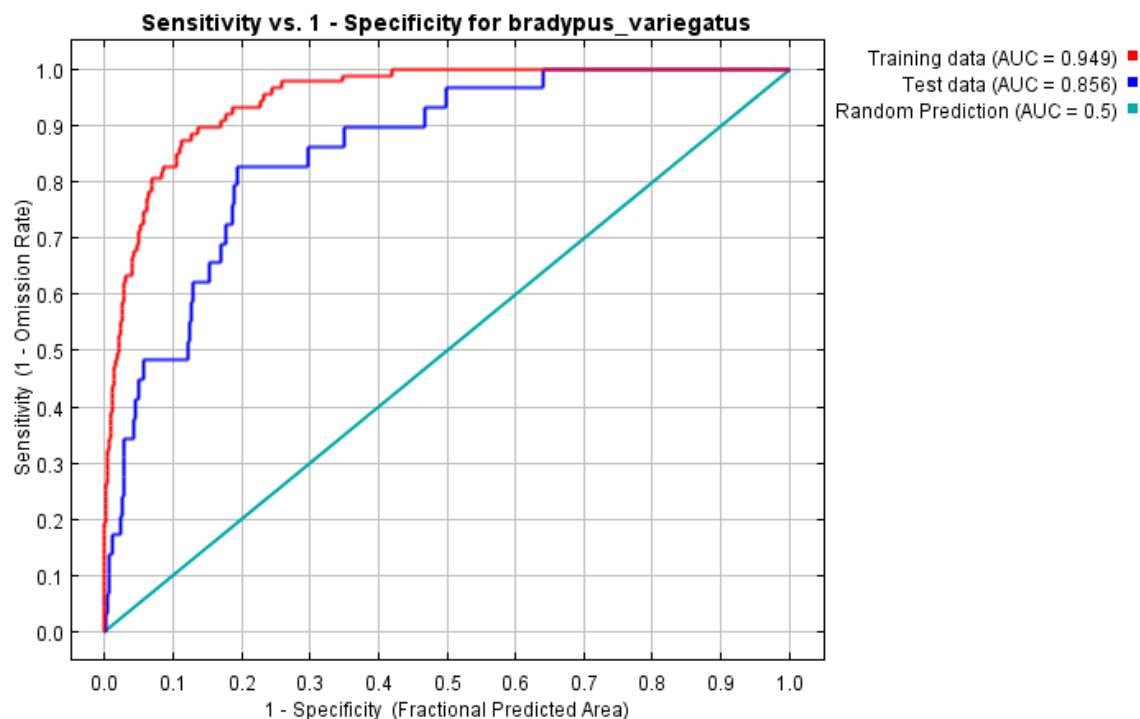
**Gráfico 1: Muestra cómo las omisiones calculadas a partir de los puntos de entrenamiento y los de test, y el área predicha como favorable varían según el valor límite cumulativo:**



## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

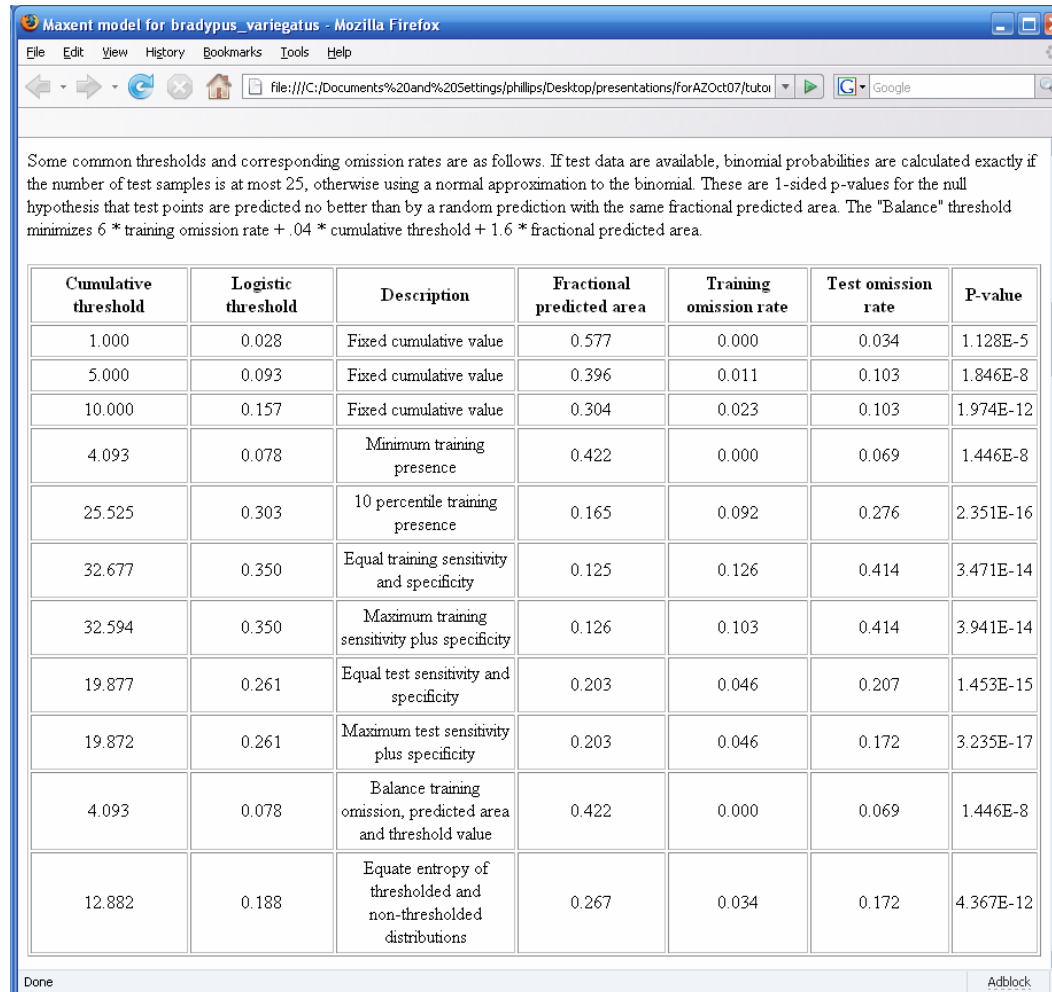
Gráfico 2: Curva operacional (curva ROC), para los 2 grupos de datos, el de test y el de entrenamiento, así como el área por debajo de la curva ROC (AUC).

La curva roja (entrenamiento) representa el ajuste del modelo a los datos de muestreo. La curva azul (test) indica el grado de ajuste del modelo a los datos de test, y supone el test real del poder predictivo del modelo. La línea turquesa representa la línea esperada si el modelo no fuese mejor que “por azar”. Si la curva azul (test) cae por debajo de la línea turquesa, indica que el modelo es peor que si se hubiese hecho al azar. Por el contrario, cuanto más se aproxime la curva azul a la esquina superior izquierda, mejor es el modelo para predecir las presencias de los datos de test.



## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

Si hay disponible un subconjunto de datos para el test, el programa calcula automáticamente la significancia estadística de la predicción utilizando un test binomial de omisión.

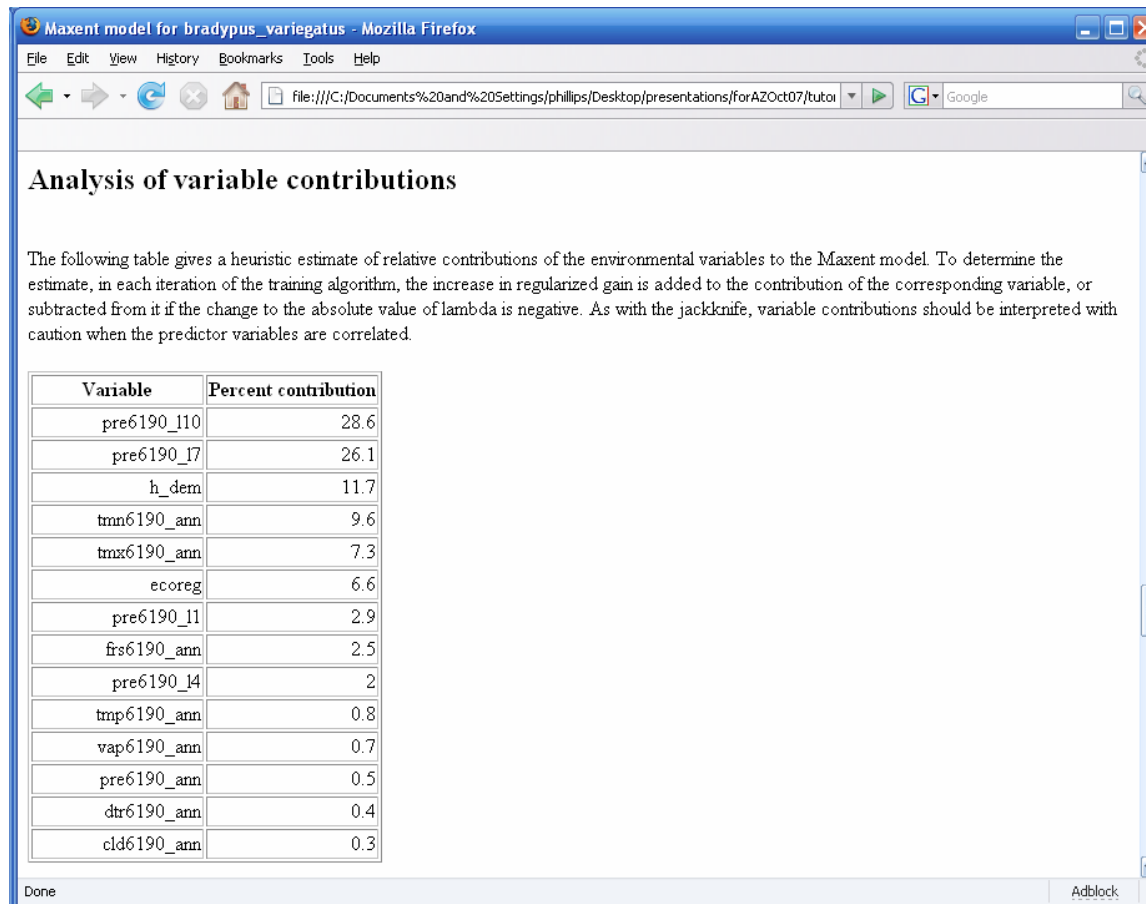


Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes  $6 * \text{training omission rate} + .04 * \text{cumulative threshold} + 1.6 * \text{fractional predicted area}$ .

| Cumulative threshold | Logistic threshold | Description   | Fractional predicted area | Training omission rate | Test omission rate | P-value   |
|----------------------|--------------------|---|---------------------------|------------------------|--------------------|-----------|
| 1.000                | 0.028              | Fixed cumulative value  | 0.577                     | 0.000                  | 0.034              | 1.128E-5  |
| 5.000                | 0.093              | Fixed cumulative value  | 0.396                     | 0.011                  | 0.103              | 1.846E-8  |
| 10.000               | 0.157              | Fixed cumulative value  | 0.304                     | 0.023                  | 0.103              | 1.974E-12 |
| 4.093                | 0.078              | Minimum training presence                                       | 0.422                     | 0.000                  | 0.069              | 1.446E-8  |
| 25.525               | 0.303              | 10 percentile training presence                                 | 0.165                     | 0.092                  | 0.276              | 2.351E-16 |
| 32.677               | 0.350              | Equal training sensitivity and specificity                      | 0.125                     | 0.126                  | 0.414              | 3.471E-14 |
| 32.594               | 0.350              | Maximum training sensitivity plus specificity                   | 0.126                     | 0.103                  | 0.414              | 3.941E-14 |
| 19.877               | 0.261              | Equal test sensitivity and specificity                          | 0.203                     | 0.046                  | 0.207              | 1.453E-15 |
| 19.872               | 0.261              | Maximum test sensitivity plus specificity                       | 0.203                     | 0.046                  | 0.172              | 3.235E-17 |
| 4.093                | 0.078              | Balance training omission, predicted area and threshold value   | 0.422                     | 0.000                  | 0.069              | 1.446E-8  |
| 12.882               | 0.188              | Equate entropy of thresholded and non-thresholded distributions | 0.267                     | 0.034                  | 0.172              | 4.367E-12 |

### ¿Qué variables importan más?

El programa asigna el incremento en el gain a las variables ambientales de las que depende la especie. Convirtiendo dichos valores a porcentajes, al final del proceso de modelado obtenemos la siguiente tabla:



Maxent model for bradypus\_variegatus - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///C:/Documents%20and%20Settings/phillips/Desktop/presentations/forAZOct07/tutoi Google

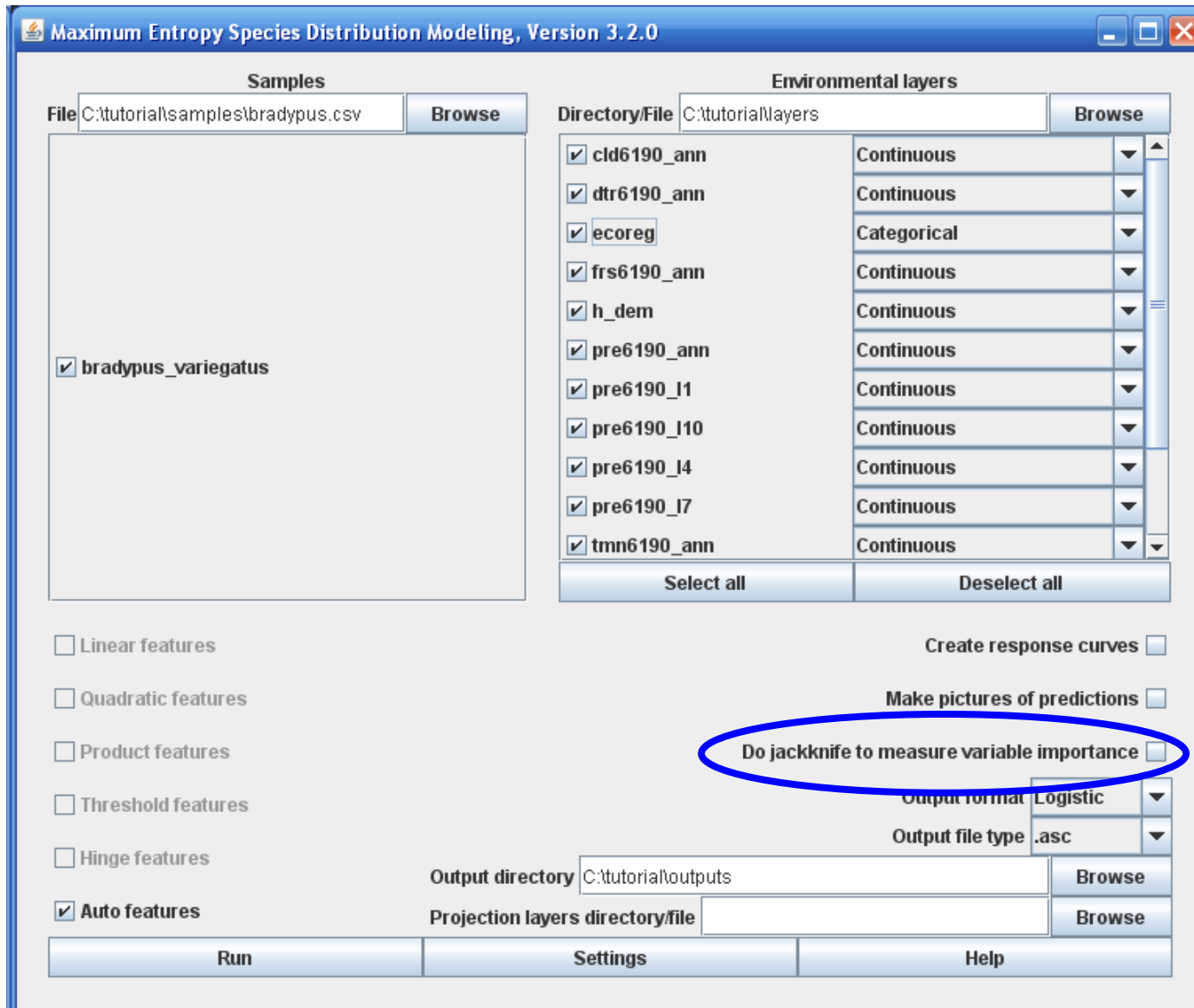
### Analysis of variable contributions

The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

| Variable     | Percent contribution |
|--------------|----------------------|
| pre6190_110  | 28.6                 |
| pre6190_17   | 26.1                 |
| h_dem        | 11.7                 |
| tmn6190_ann  | 9.6                  |
| tmx6190_ann  | 7.3                  |
| ecoreg       | 6.6                  |
| pre6190_11   | 2.9                  |
| firs6190_ann | 2.5                  |
| pre6190_14   | 2                    |
| tmp6190_ann  | 0.8                  |
| vap6190_ann  | 0.7                  |
| pre6190_ann  | 0.5                  |
| dtr6190_ann  | 0.4                  |
| cld6190_ann  | 0.3                  |

Done Adblock

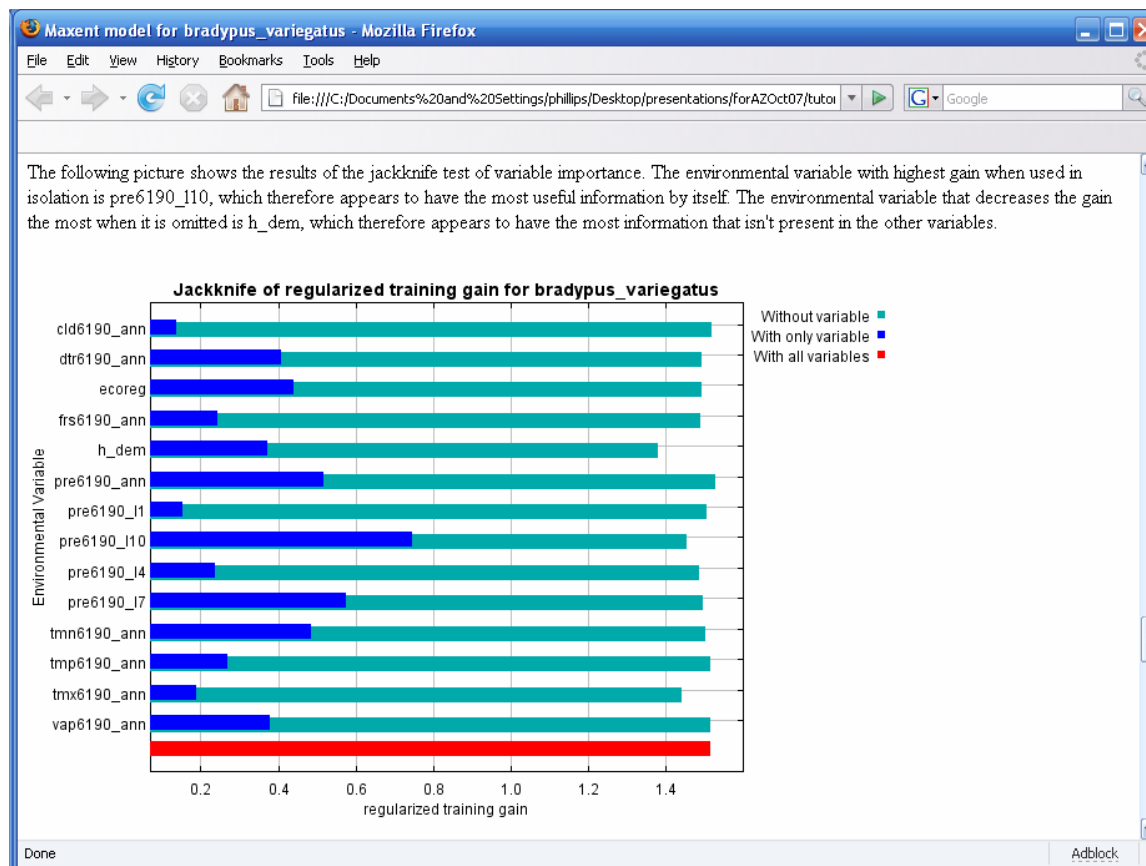
# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

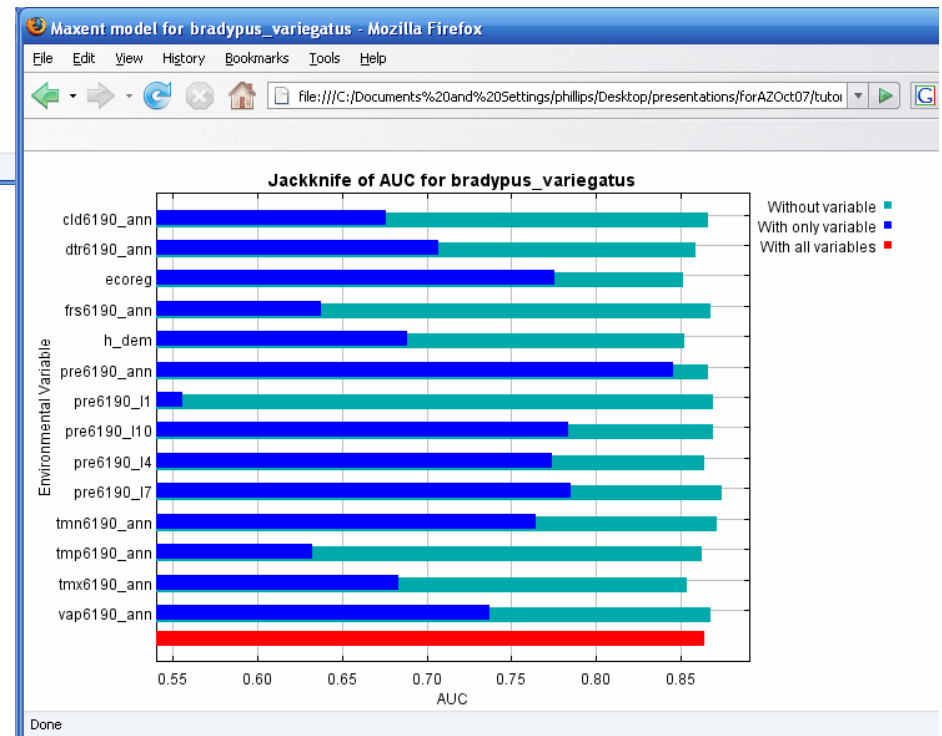
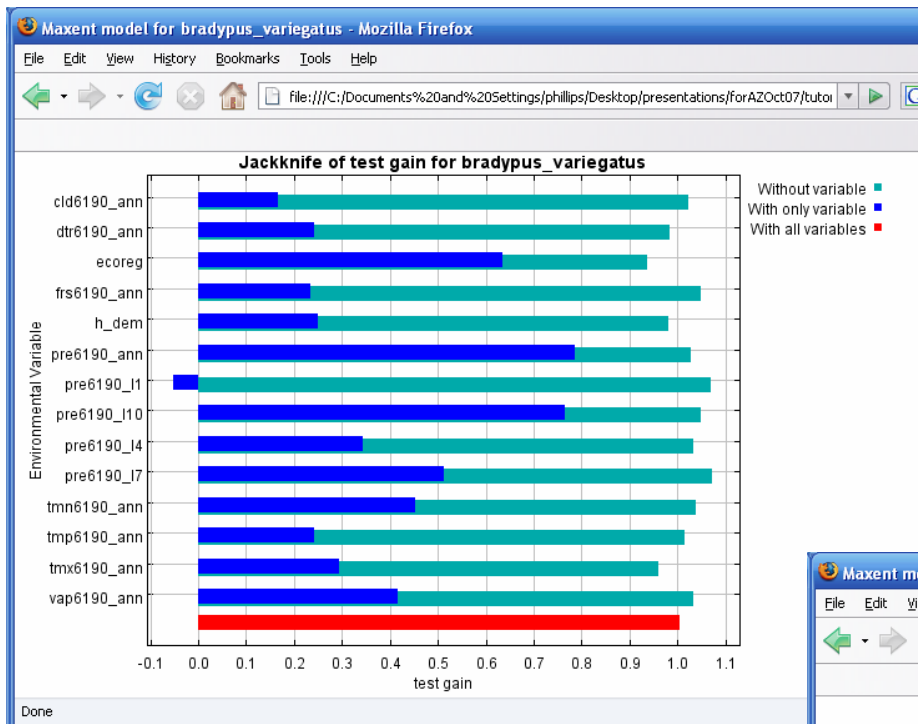
### ¿Qué variables importan más?

En cada vuelta del modelo se excluye una variable y se crea el modelo con las variables remanentes. Después, se crea un modelo con cada una de las variables por separado. De forma adicional, se crea un modelo utilizando todas las variables, como en el caso “normal” de ejecutar MaxEnt.

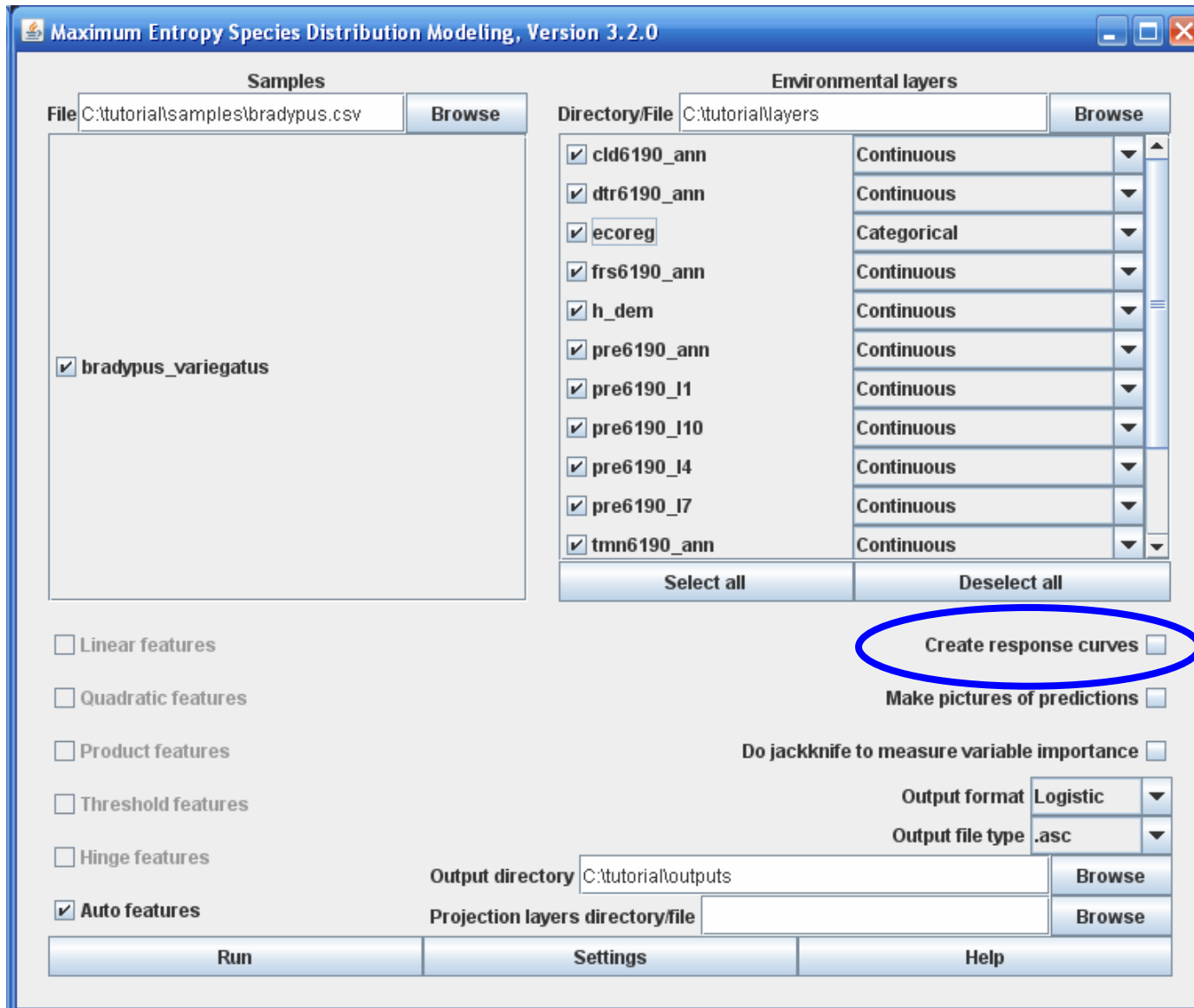




# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt

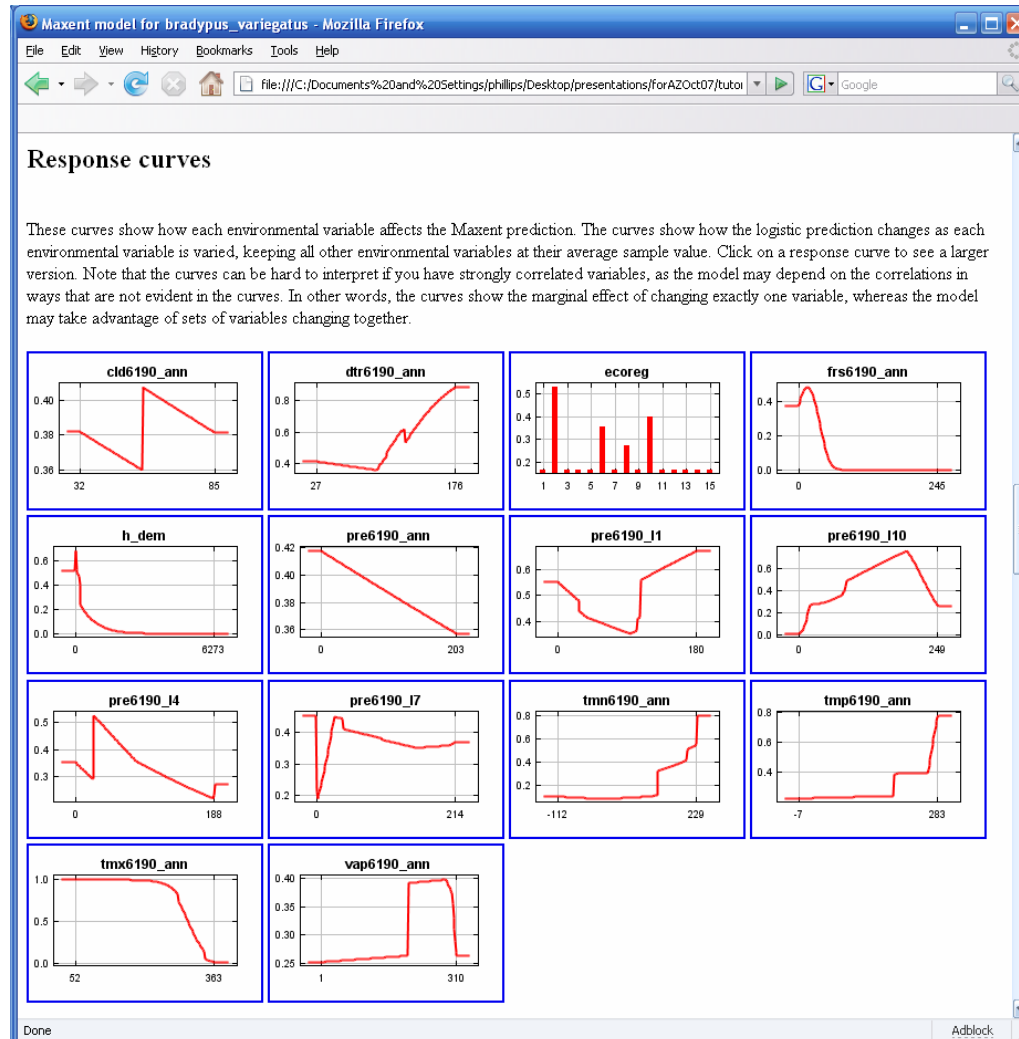


# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



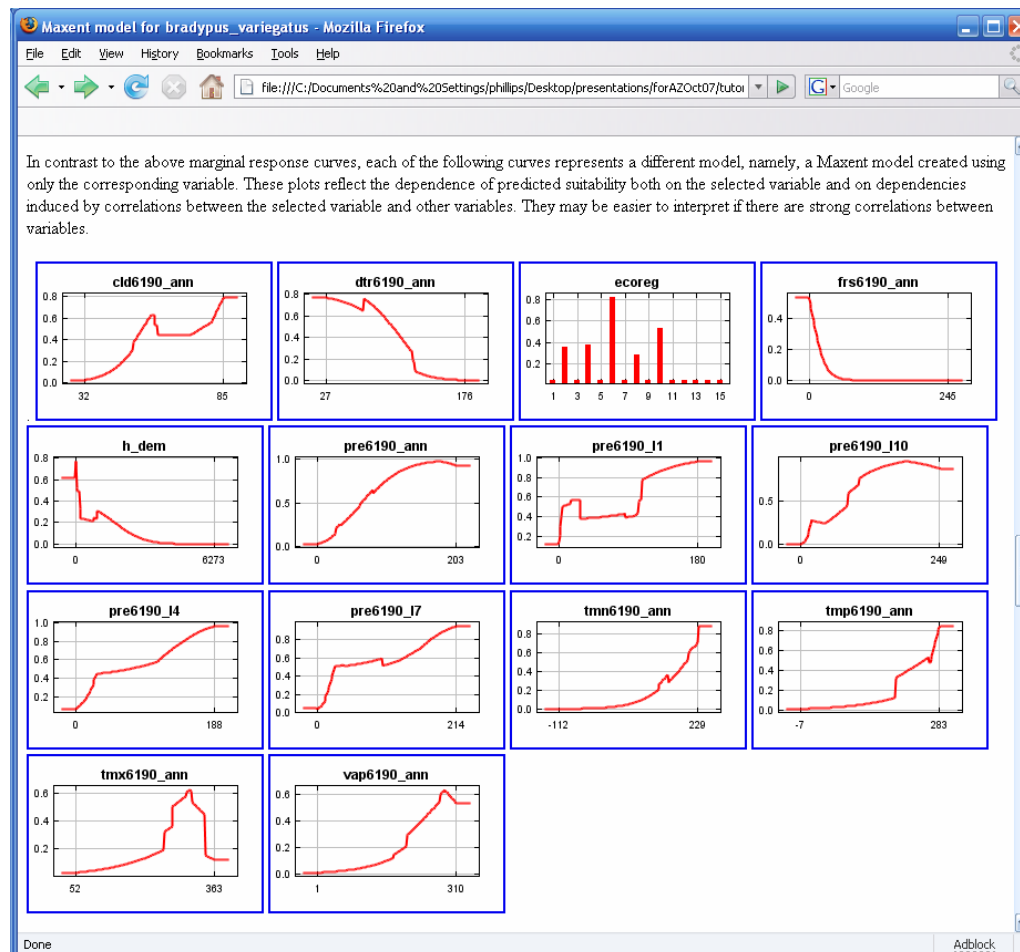
## ¿Cómo depende la predicción de las variables?

**Curvas de respuesta: se evalúa cada variables manteniendo el valor del resto fijo en su valor medio (cuidado con variables correlacionadas!!!)**

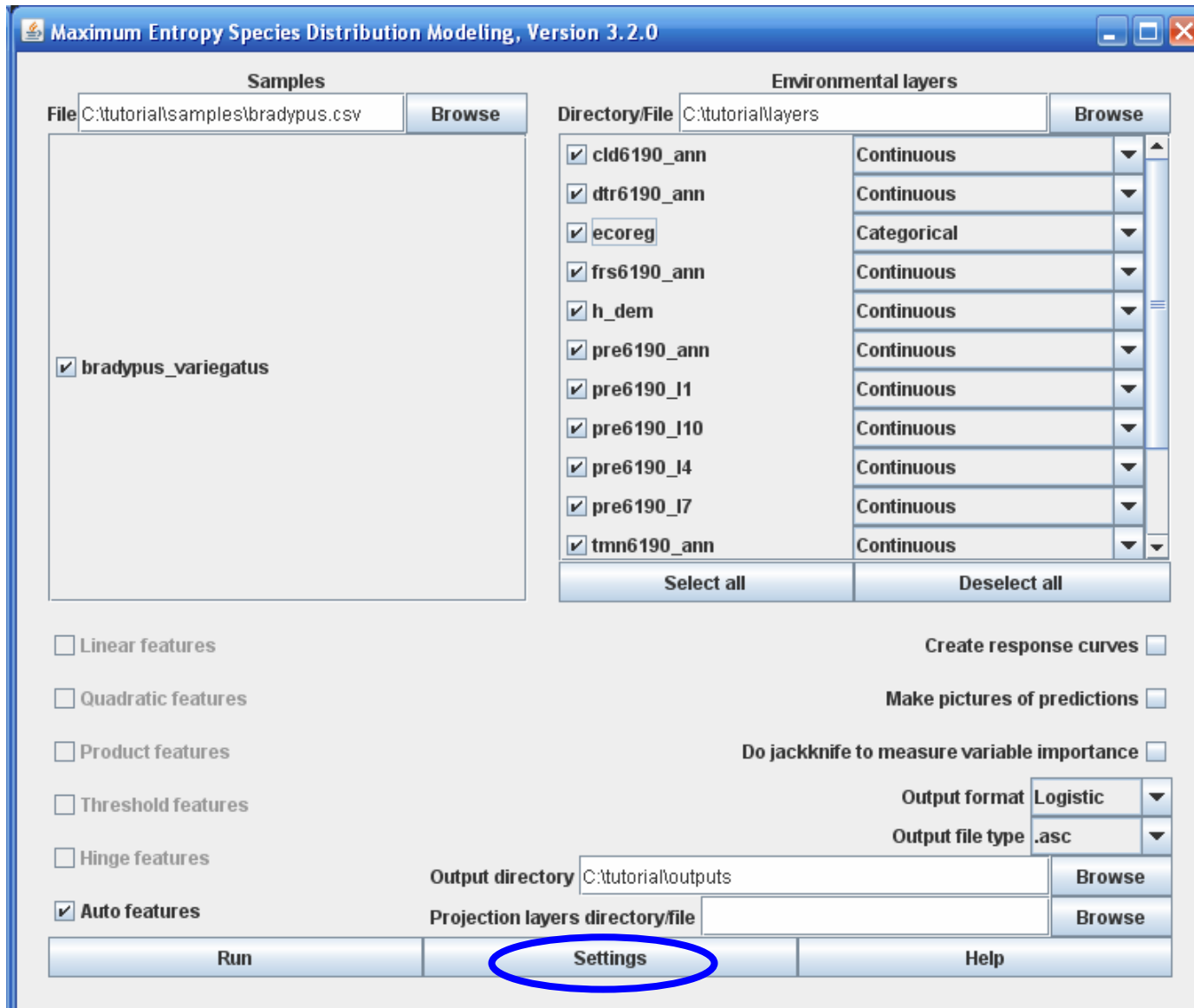


## ¿Cómo depende la predicción de las variables?

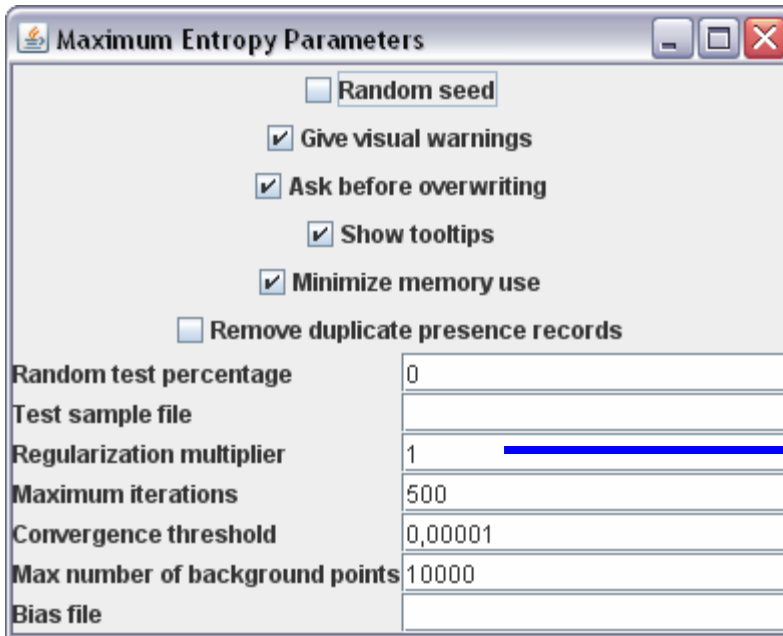
Contribución marginal de cada variable por sí sola al modelo (obviando el resto de variables)



# 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



## 13. Funcionamiento, interfaz y formato de los datos en MaxEnt



Maximum Entropy Parameters

- Random seed
- Give visual warnings
- Ask before overwriting
- Show tooltips
- Minimize memory use
- Remove duplicate presence records

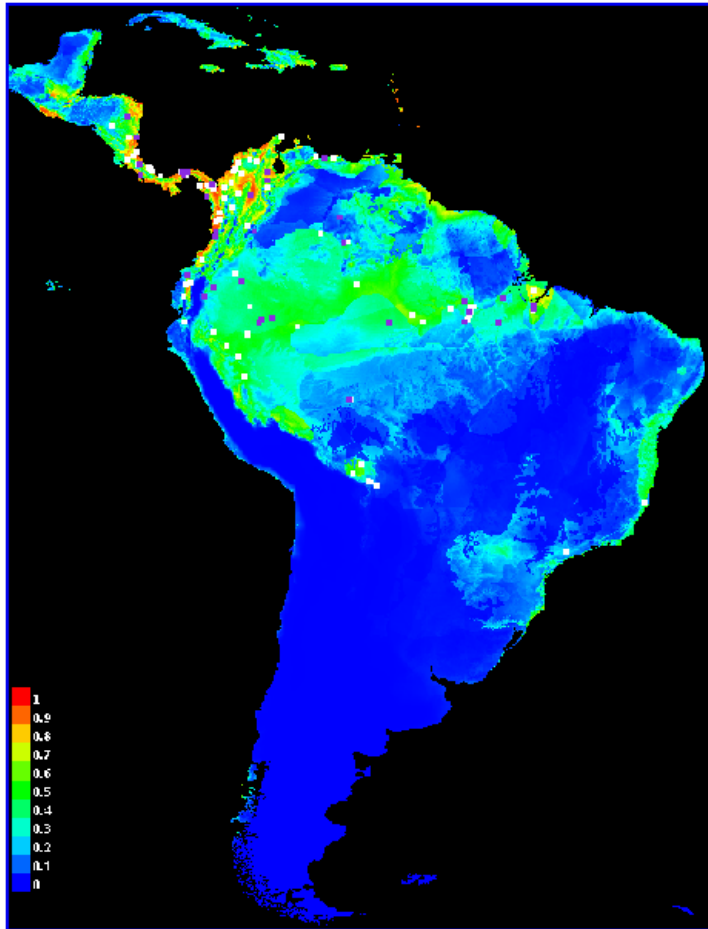
|                                 |         |
|---------------------------------|---------|
| Random test percentage          | 0       |
| Test sample file                |         |
| Regularization multiplier       | 1       |
| Maximum iterations              | 500     |
| Convergence threshold           | 0,00001 |
| Max number of background points | 10000   |
| Bias file                       |         |

**Si ponemos un valor más pequeño: modelo menos ajustado**

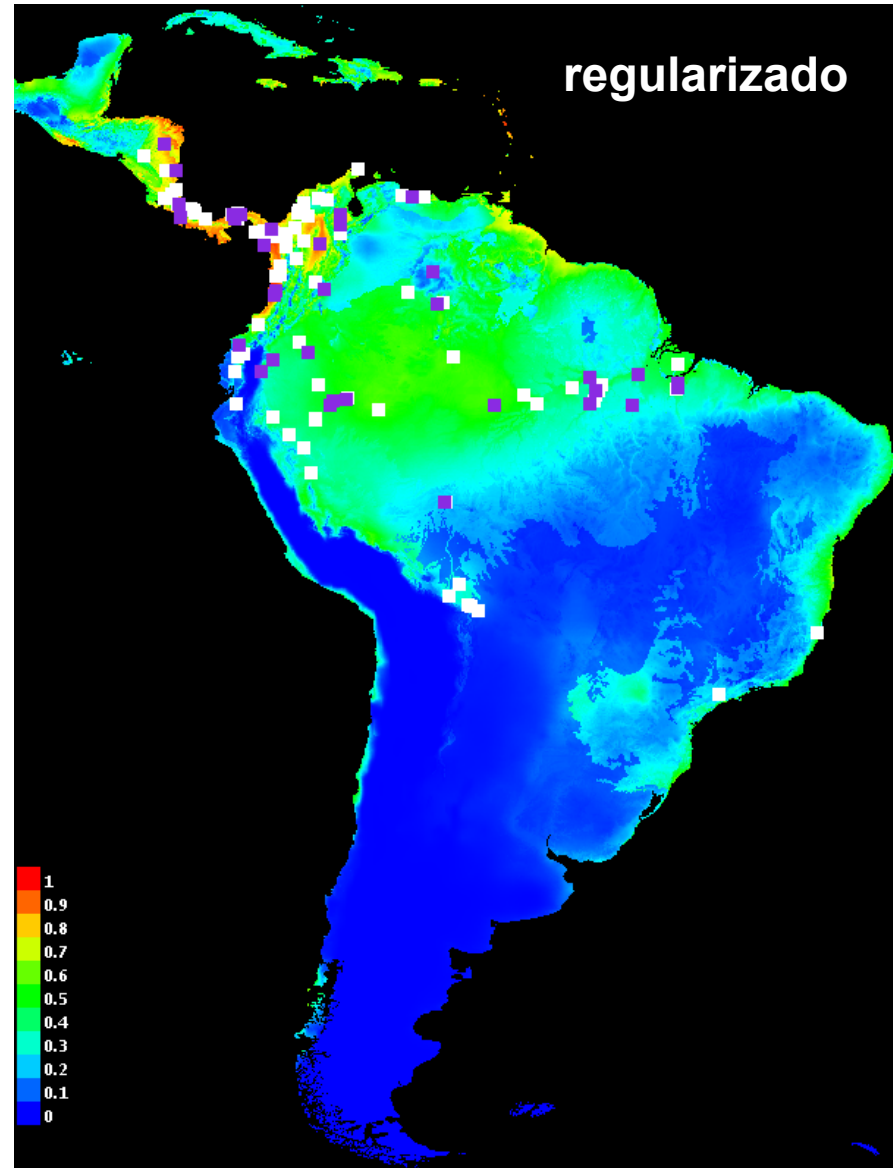
**Cuidado: distintos valores pueden dar lugar a modelos demasiado generales o modelos sobre-ajustados**

## Regularización: modelo más general (valor de regularización menor que 1)

original



regularizado



Para quien tenga un poco de experiencia en programación...

- 1) Cuando las grids ambientales son muy grandes se puede utilizar el formato de datos SWD (“*samples with data*”)
- 2) La modelización puede automatizarse con comandos de línea (*BATCH RUNNING*) cuando se necesitan generar muchos modelos cambiando los parámetros
- 3) Los resultado de MaxEnt se pueden importar en otros softwares estadísticos para realizar análisis adicionales (ejemplo: paquetes ROCR, vcd y boot de R)

→ **Todo esto está explicado en el manual**



**Muchas gracias**