

Réplica del III taller sobre modelización de nichos ecológicos de GBIF

Madrid, España. 14-16 marzo 2007



Francisco Javier Bonet García, 2007

■ Generalidades sobre modelización: Culturas de modelización

■ Cultura algorítmica



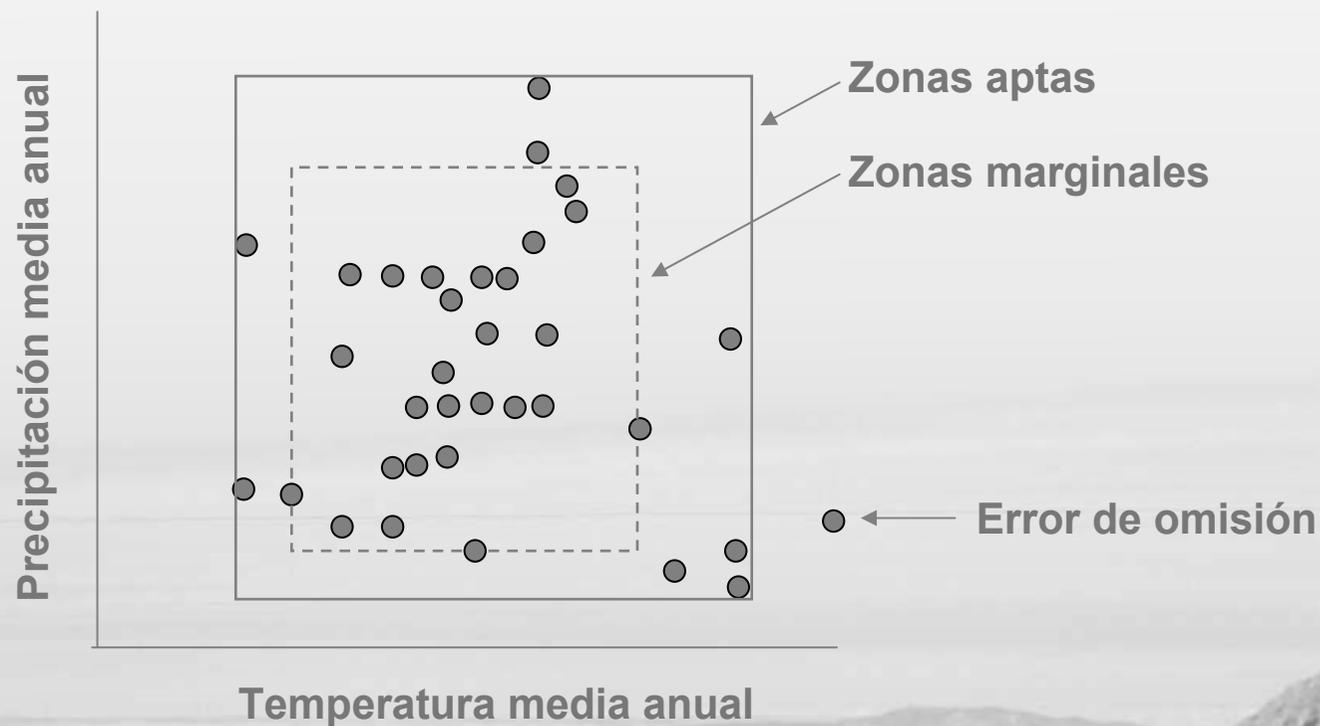
- El contenido de la caja es complejo y desconocido. Es difícil interpretar el funcionamiento del modelo.
- Dentro de la caja hay “comportamientos muy humanos”
- Los resultados se validan probando el ajuste predictivo del modelo.

BIOCLIM

- Se basa en el algoritmo de “envolventes bioclimáticas”.
- La idea básica de BIOCLIM es encontrar una regla sencilla que identifique todas las áreas con un clima similar en las zonas de ocurrencia de la especie.
- En principio se usan 35 variables climáticas para determinar la distribución potencial de una especie. Pero hay aplicaciones (openmodeller) donde no se condiciona el número ni tipo de variables utilizadas. Se pueden usar variables no climáticas.
- Para cada variable ambiental, el algoritmo encuentra la media y la desviación típica (asumiendo una distribución normal) asociadas con los puntos de presencia de la especie.
- La distribución potencial se identifica interpolando las variables climáticas para cada celda de un MDE y comparándolas con el perfil climático de la especie.
- BIOCLIM se puede utilizar para tres tareas fundamentales:
 - Describir las características ambientales bajo las cuales las especies pueden vivir
 - Identificar otras localidades donde las especies podrían vivir
 - Identificar los lugares donde la especie podría vivir bajo escenarios climáticos diferentes

BIOCLIM

- Ejemplo de una envolvente bidimensional. Los puntos representan la temperatura media anual y la precipitación media anual para una serie de lugares de presencia de una especie dada. BIOCLIM clasificará todas las ubicaciones con valores incluidos en los extremos de la envolvente (línea continua) como adecuados para la especie. El recuadro discontinuo excluye las áreas que están fuera del percentil 95 de la especie.

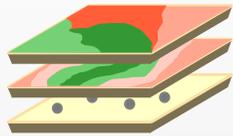


BIOCLIM

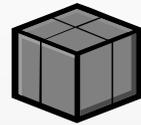
■ Desventajas

- No tiene en cuenta las posibles relaciones entre las variables utilizadas
- Todos los predictores tienen el mismo “peso”
- Todas las condiciones ambientales incluidas en la envolvente, son igualmente adecuadas para la especie
- No permite el uso de información categórica

BIOCLIM



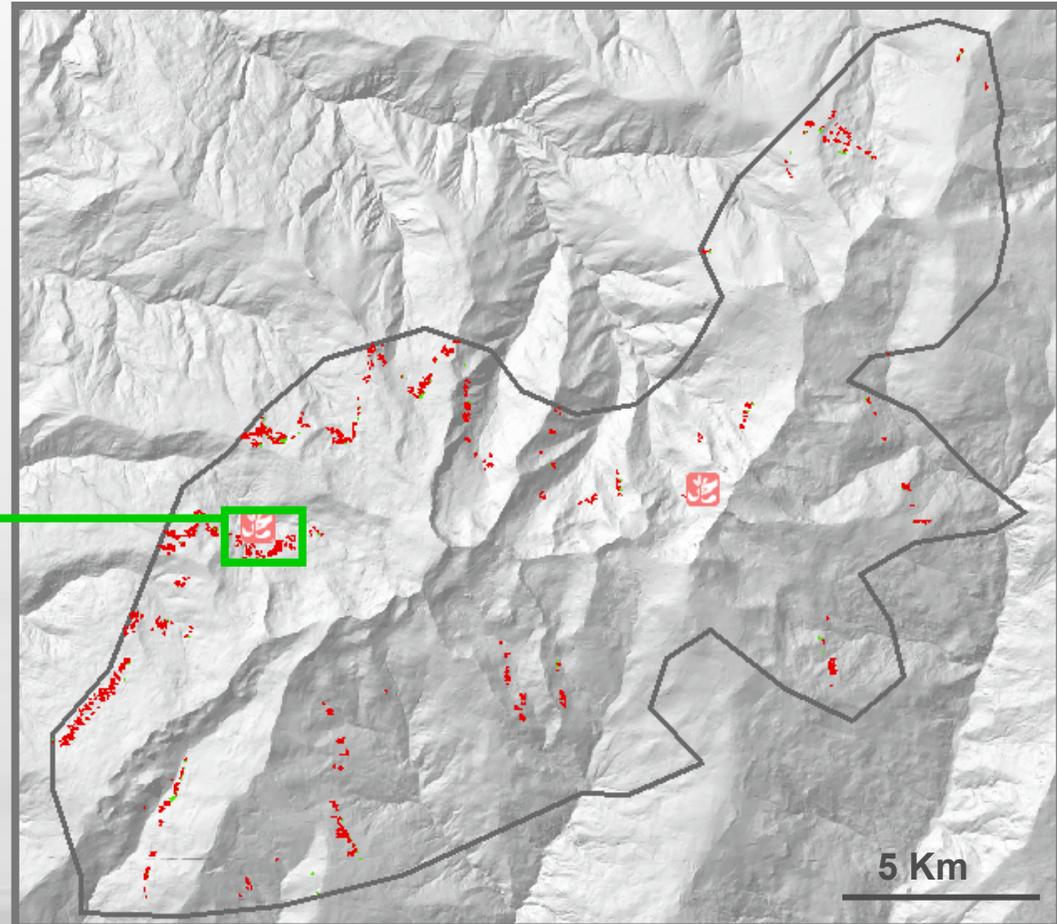
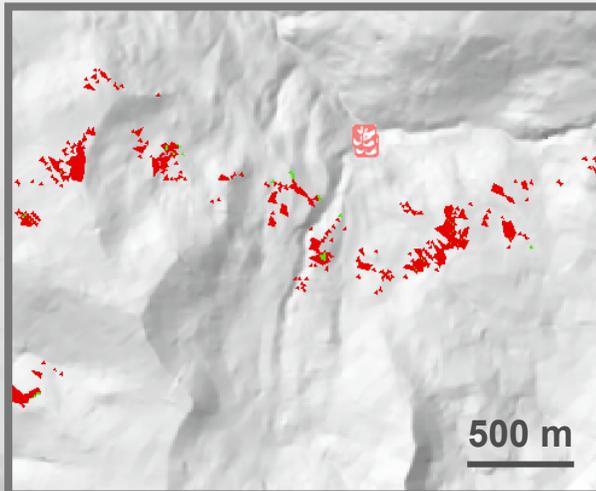
Datos de partida



BIOCLIM



- Los lugares de presencia no son considerados como adecuados ...¿¿??



BIOMAPPER

Fundamentos teóricos

- Se basa en el análisis factorial de nicho ecológico (ENFA)
- Resume todas las variables ambientales en sólo unas pocas que no están correlacionadas entre sí
- Tiene en cuenta sólo los datos de presencia de la especie
- Compara la distribución actual de la especie con “todo” el medio disponible
- Los conceptos clave son: **marginalidad** y **especialización**

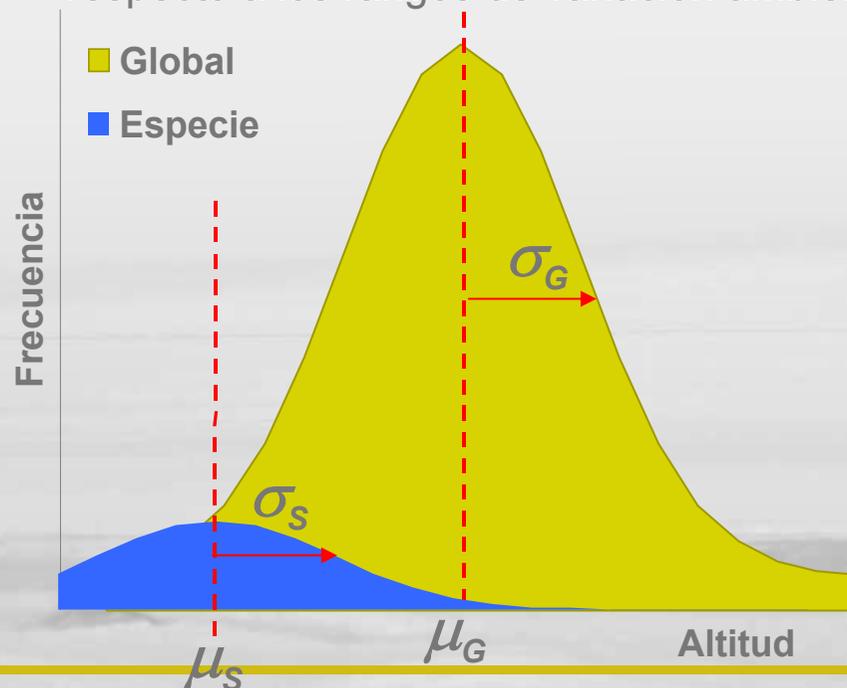
Alexandre Hirzel
Lab. of Conservation Biology
University of Lausanne, Switzerland
<http://www2.unil.ch/biomapper>



BIOMAPPER

Fundamentos teóricos

- El nicho de la especie es un subconjunto del ambiente global
- Se analizan las diferencias entre el espectro ambiental de las áreas con presencia constatada, y el de toda la región.
- El factor de **marginalidad** muestra lo alejado que está el óptimo de la especie de la media del hábitat de la región. Es la recta que une los centroides de la distribución en toda la región y en los lugares con presencia constatada.
- El factor de **especialización** muestra lo especializada que está la especie con respecto a los rangos de variación ambiental presentes en la región.

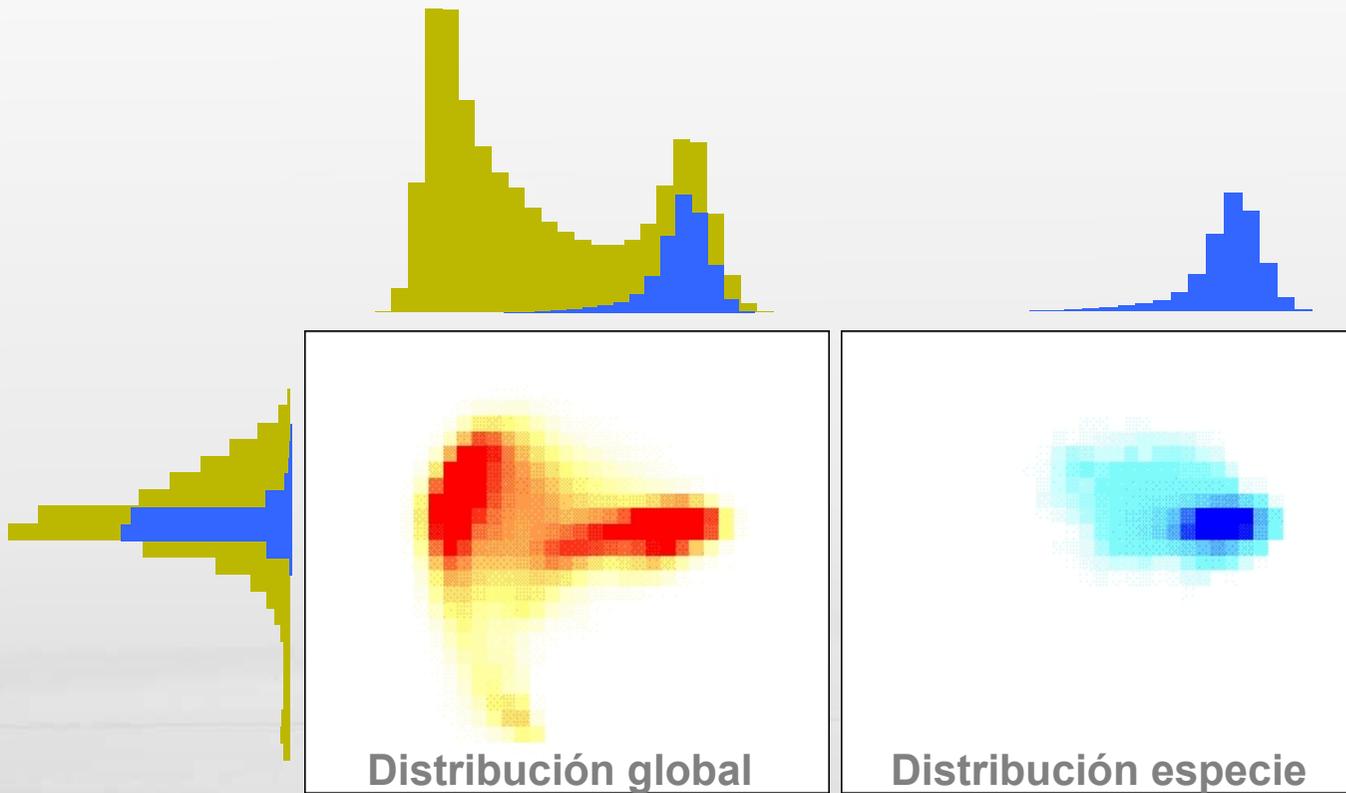


$$\text{Marginalidad} = \frac{|\mu_G - \mu_s|}{1.96 \cdot \sigma_G}$$

$$\text{Especialización} = \frac{\sigma_G}{\sigma_s}$$

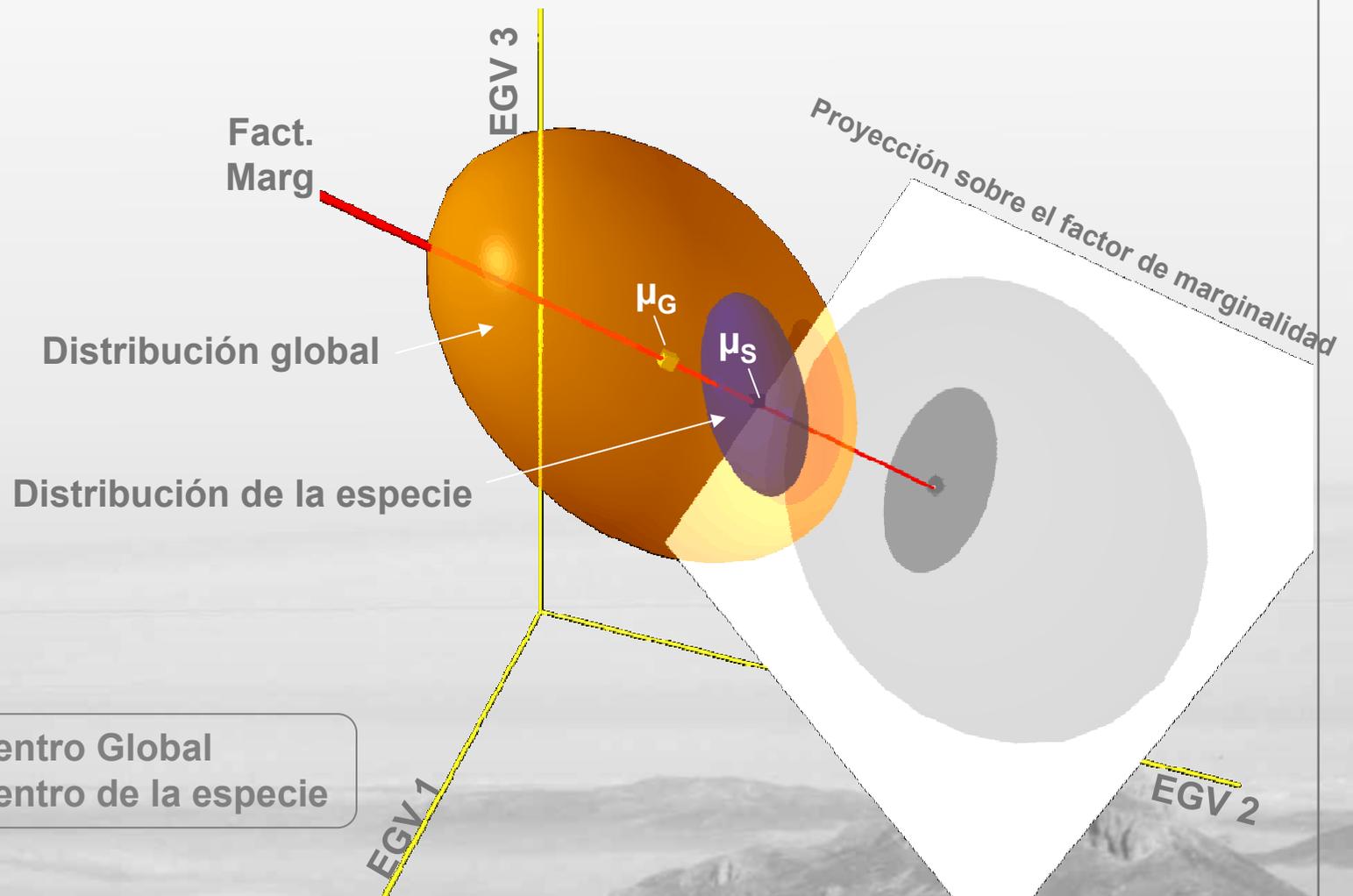
BIOMAPPER

■ Si ahora incorporamos otra variable ...



BIOMAPPER

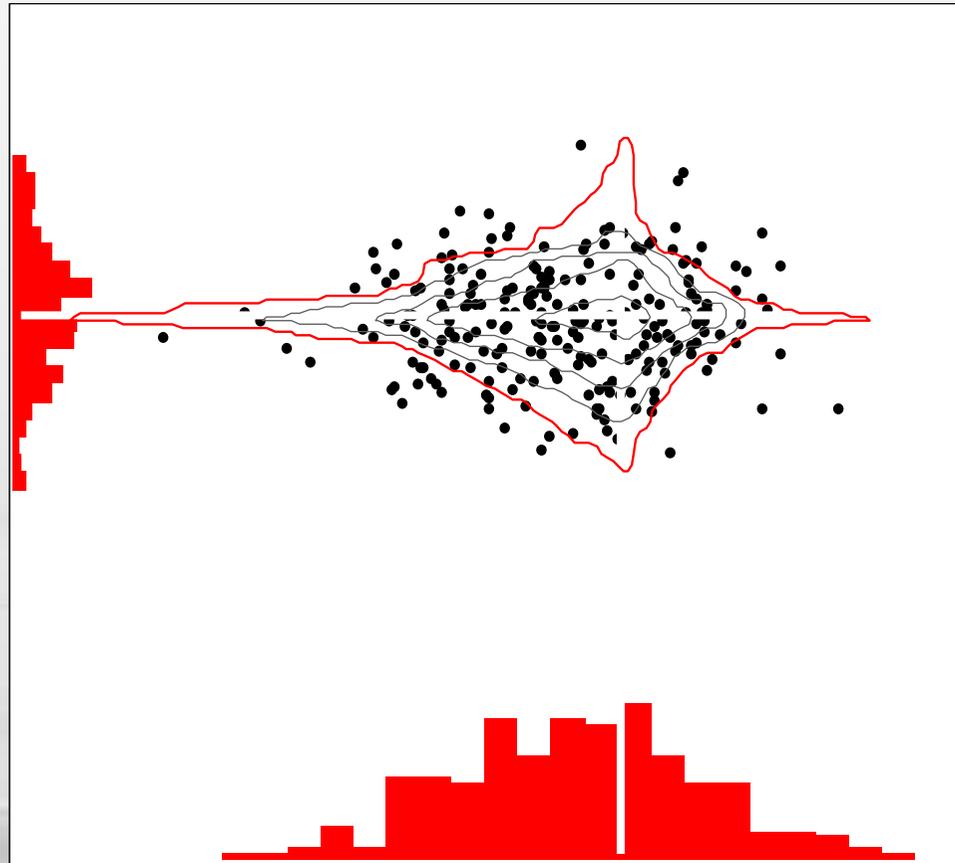
■ Y otra más...



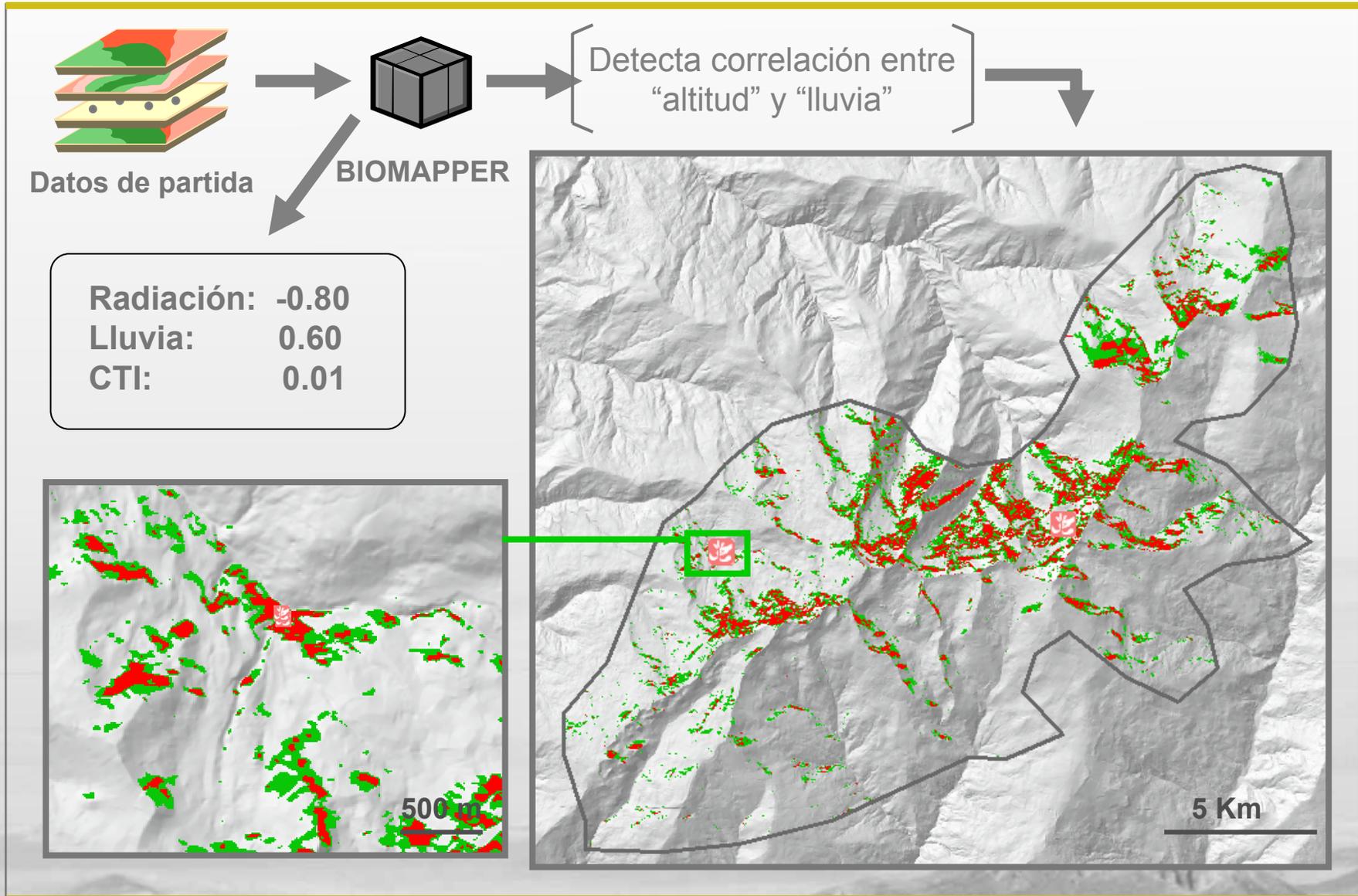
BIOMAPPER

Fundamentos teóricos

- Se define una **envolvente** mediante la frecuencia y la mediana, asumiendo una distribución unimodal y simétrica de las variables
- Cuando toda la variabilidad presente en el espectro ambiental se recoge en los ejes, se obtiene un listado de los diferentes factores en orden decreciente de explicación.



BIOMAPPER



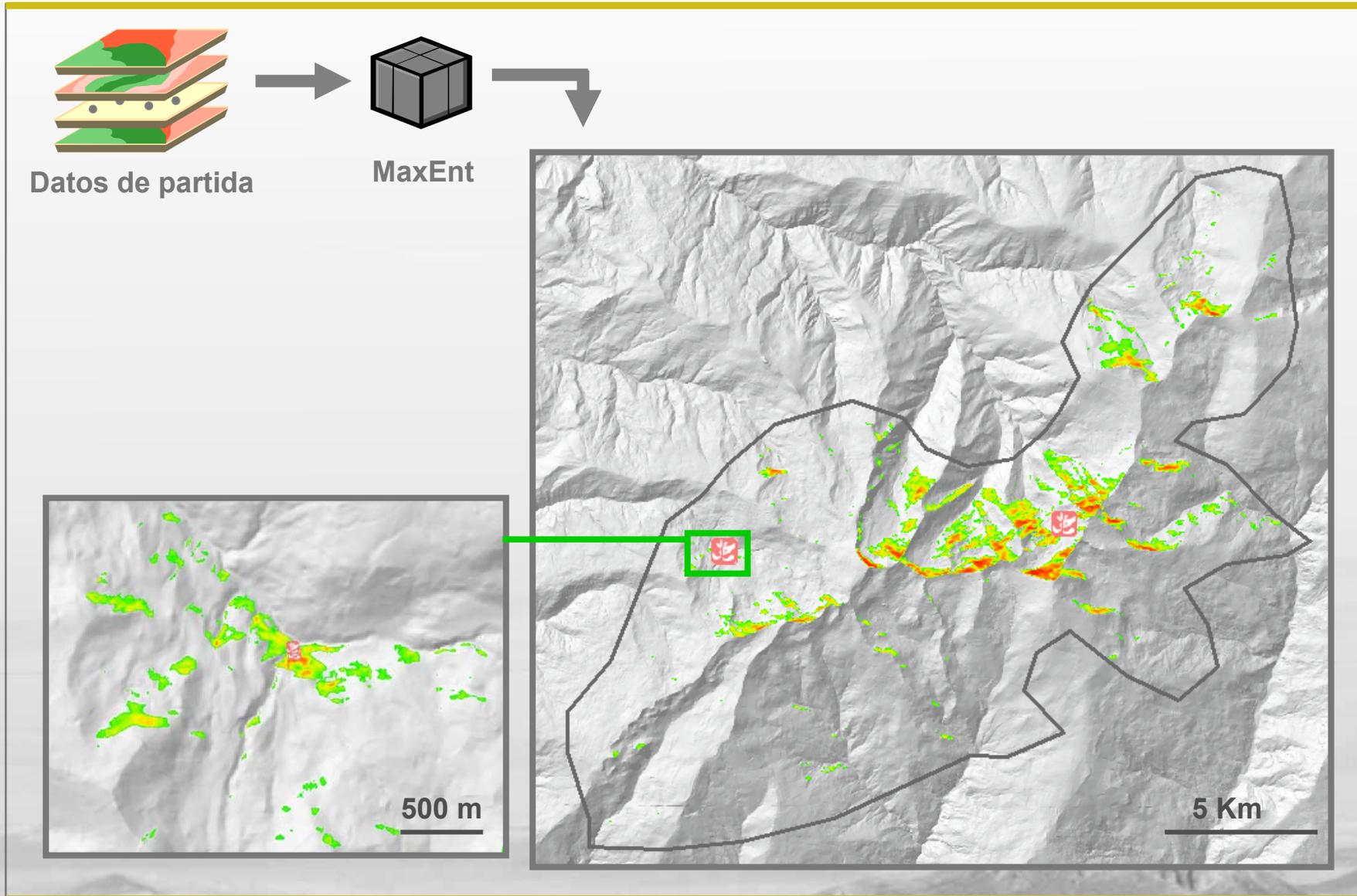
MaxEnt

Fundamentos teóricos

- Método genérico para hacer estimaciones o inferencias a partir de información incompleta.
- Cuando queremos aproximarnos a una distribución de probabilidad desconocida, nos preguntamos cuál será la mejor aproximación. Una respuesta general a esto es: **“La mejor aproximación es asegurarse de que la aproximación satisface todas las restricciones de la distribución desconocida, y que sujeta a esas restricciones, la distribución debería de tener la mínima entropía”**
- Se basa sólo en datos de presencia
- Se implementa teniendo en cuenta procesos iterativos que permiten que el modelo “aprenda”.
- Consume muchos recursos informáticos a la hora de ejecutar los modelos
- Es un algoritmo relativamente nuevo, por lo que está en fase de experimentación.

Modelos algorítmicos

MaxEnt



GARP

Fundamentos teóricos

- GARP: **G**enetic **A**gorithm for **R**ule set **P**rediction
- Un algoritmo genético es un método de buscar la solución a un problema que asemeja dicha búsqueda a la evolución biológica.
- Se establece una analogía entre el conjunto de soluciones del problema, y el conjunto de individuos de una población natural. Se codifica la información de cada solución en una cadena binaria llamada cromosoma. Los símbolos que forman la cadena, se denominan genes.
- Los cromosomas evolucionan (buscando la solución al problema) a través de iteraciones (generaciones). En cada generación los cromosomas son evaluados usando alguna medida de aptitud.

GARP

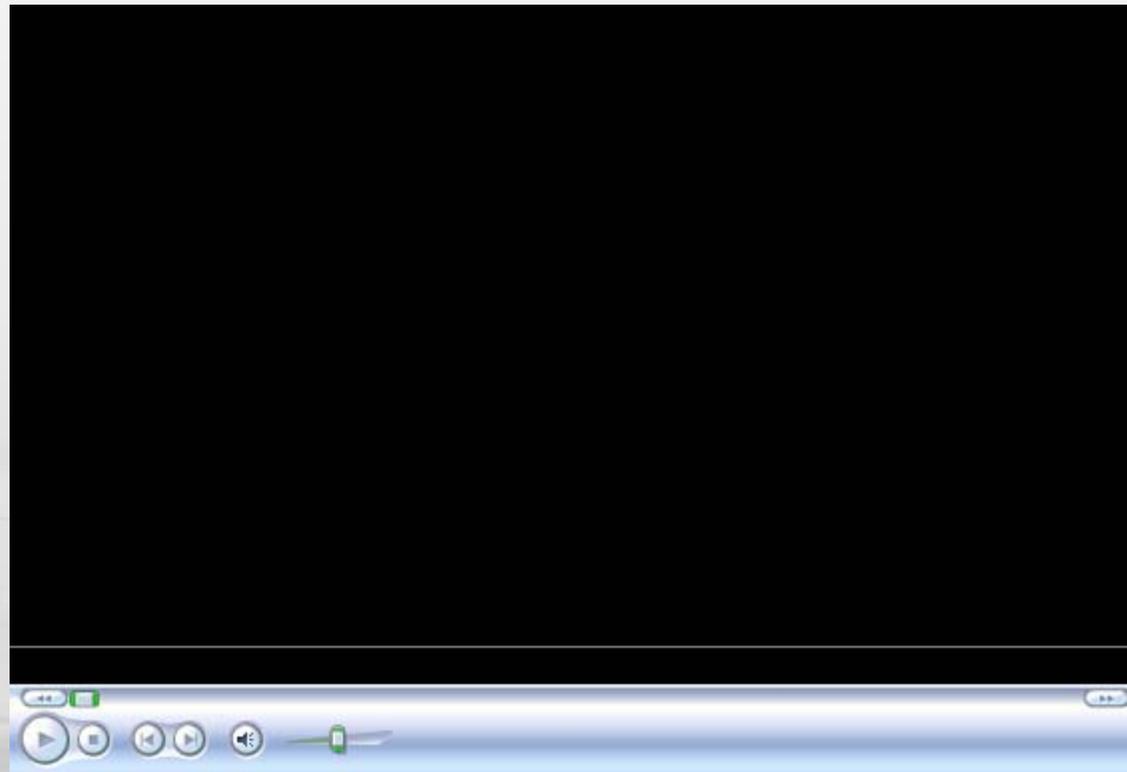
■ Fundamentos teóricos: Funcionamiento de un AG básico

- **Generación de la población inicial:** Consta de cromosomas, que representan las posibles soluciones del problema.
- **Función de aptitud:** A cada cromosoma se le aplica una función de aptitud para saber cómo de buena es la solución que está codificando.
- **Elección de cromosomas:** una vez que se sabe la aptitud de cada cromosoma, se procede a elegir aquellos que serán cruzados en la siguiente generación. Los más aptos tiene más probabilidad de reproducirse.
- **Cruzamiento:** es el principal operador genético, representando la reproducción sexual. Opera sobre dos cromosomas a la vez para genera dos descendientes donde se combinan las características de los padres.
- **Iteración:** el algoritmo se detiene cuando alcanza la solución óptima. Pero ésta generalmente se desconoce, por lo que se deben de usar otros métodos de detección: número máximo de iteraciones o detención cuando no haya cambios.

GARP

■ Fundamentos teóricos: Funcionamiento de un AG básico

Este humanoide tiene que mantenerse de pie. Sus músculos y articulaciones están controlados por una red neuronal que aprende gracias un **algoritmo genético**. En cada iteración se van conservando aquellos movimientos que contribuyen a la estabilidad.



GARP

Fundamentos teóricos

- GARP genera una serie de reglas
- La primera iteración genera la primera regla y evalúa los errores de omisión y comisión del modelo.
- Las siguientes iteraciones generan más reglas. Estas reglas son incluidas o excluidas del conjunto según el grado de ajuste del modelo.
- El algoritmo se para cuando ya no se pueden crear mejores modelos o se alcanza el número máximo de iteraciones.
- El conjunto final de reglas consiste en una serie de declaraciones del tipo “si ... entonces”

GARP

Fundamentos teóricos

Ejemplo de las reglas de GARP

5 r 0.50 1.00 29.52 0.22 0.620

```
IF - clim_01*0.25 - clim_02*0.01 + clim_05*0.17 - clim_20*0.23 + lith*0.26 + clim_06*0.21 +  
clim_07*0.36 + clim_08*0.13 + clim_09*0.03 + clim_10*0.34 + clim_11*0.26 - clim_12*0.34 -  
clim_13*0.30 + clim_14*0.06 - clim_16*0.04 + clim_17*0.07 - clim_18*0.45 + clim_19*0.08 +  
clim_21*0.37 - clim_22*0.46 - clim_24*0.15 - clim_25*0.17 - clim_26*0.38 + clim_27*0.46 +  
dem*0.04 + topopos*0.02 + wetind*0.41 - lat*0.01 - long*0.28 - evc*0.10 + slope*0.03 -  
geol*0.09 - gully*0.26 + aspeast*0.04 - aspsouth*0.50 THEN species=ABSENT  
19 d 0.50 0.85 16.58 0.14 0.023
```

```
IF clim_20=[14.9,15.5] AND clim_07=[20.3,25.3] AND clim_11=[4.5,9.2] AND clim_13=[105,178]u  
AND clim_17=[184,262] AND clim_19=[334,485] AND clim_21=[24.5,25.4] AND clim_22=[5.9,6.6]u  
AND clim_24=[7.5,13.5] AND topopos=[-82,48] AND wetind=[-2.2,23.1] AND lat=[-38.06,-37.10]u  
AND long=[145.31,146.31] AND slope=[1.1,17.0] AND geol=[ 1, 5] THEN species=ABSENT
```

0 m 0.49 0.95 41.47 0.50 0.013

```
IF clim_01=[6.8,12.7] AND clim_02=[8.8,10.8] AND clim_05=[20.2,25.1]  
AND clim_20=[14.9,15.3] AND lith=[ 3, 6] AND clim_06=[-2.4,3.4] AND clim_07=[21.2,23.0] AND  
AND clim_17=[205,282] AND clim_18=[213,308] AND clim_19=[288,519]u  
AND clim_21=[24.5,25.2] AND clim_22=[6.0,6.5] AND clim_24=[9.4,13.5] AND clim_25=[21.2,21.4]u  
AND clim_26=[22.9,23.9] AND clim_27=[7.3,7.5] AND dem=[437,1387] AND topopos=[-70,109]u  
AND wetind=[1.4,14.7] AND lat=[-37.93,-37.67] AND long=[145.62,146.36] AND evc=[11,30] AND  
slope=[-4.4,32.7] AND geol=[ 3, 5] AND gully=[ 0, 1] AND aspeast=[-0.95,0.86] AND aspsouth=[-  
0.91,0.93] THEN species=PRESENT
```

GARP

■ Ventajas

- Fácil de implementar
- Relativamente robusto en casos en los que hay pocas muestras iniciales

■ Desventajas

- Es una verdadera caja negra
- Los mapas no son determinísticos
- Tiende a la generar errores por comisión: lugares no aptos considerados como tales.
- Los resultados son binarios, no continuos
- Consume muchos recursos informáticos

Modelos algorítmicos

GARP

