



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 GARP

Elisa Liras

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

¿Qué significa GARP?

El acrónimo GARP viene del inglés ***Genetic Algorithm for Rule-set Production*** (algoritmo genético basado en reglas), y fue desarrollado por David Stockwell en el Centro de Supercomputación de San Diego.

¿Y qué es?

GARP es un **algoritmo genético** que crea un modelo de nicho ecológico para una especie. El algoritmo representa las condiciones ambientales la especie sería capaz de mantener su población.

¿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 hace el algoritmo con los datos?

GARP trata, de **forma iterativa**, de encontrar las **correlaciones** entre las **presencias y ausencias** de la especie con los **parámetros ambientales**, utilizando una serie de reglas diferentes. Cada tipo de regla implementa un método diferente para construir los modelos de predicción de la especie. Actualmente hay 4 tipos de reglas implementadas: atómica, regresión logística, envoltura bioclimática y negación de la envoltura bioclimática.

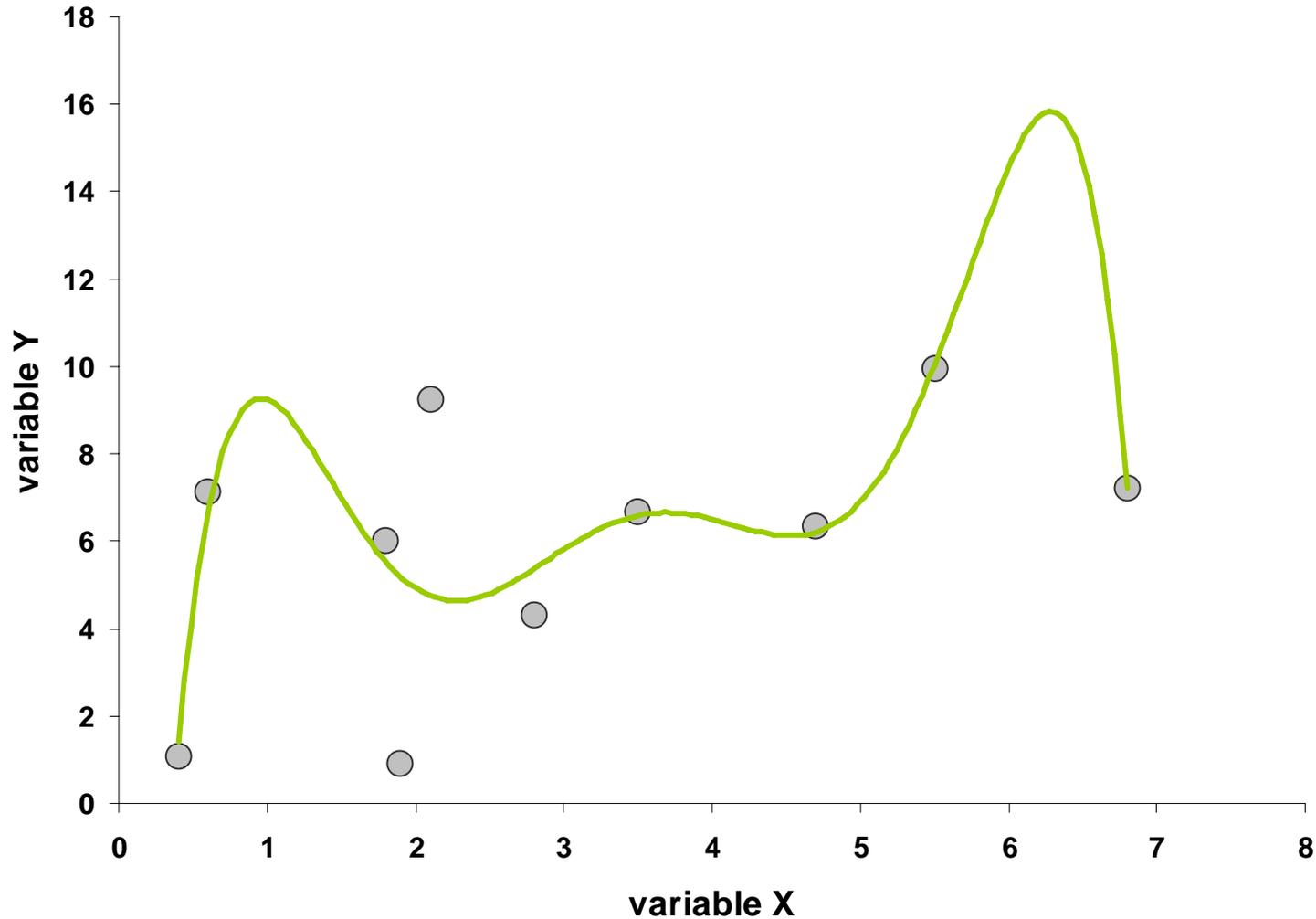
¿Qué es un algoritmo genético?

Los algoritmos genéticos son algoritmos de optimización desarrollados por expertos en inteligencia artificial que usan una metáfora para los conceptos de la genética y la teoría de la evolución de la especie, desde el punto de vista biológico, para definir las estructuras de datos y los procesos.

¿Qué es DesktopGarp?

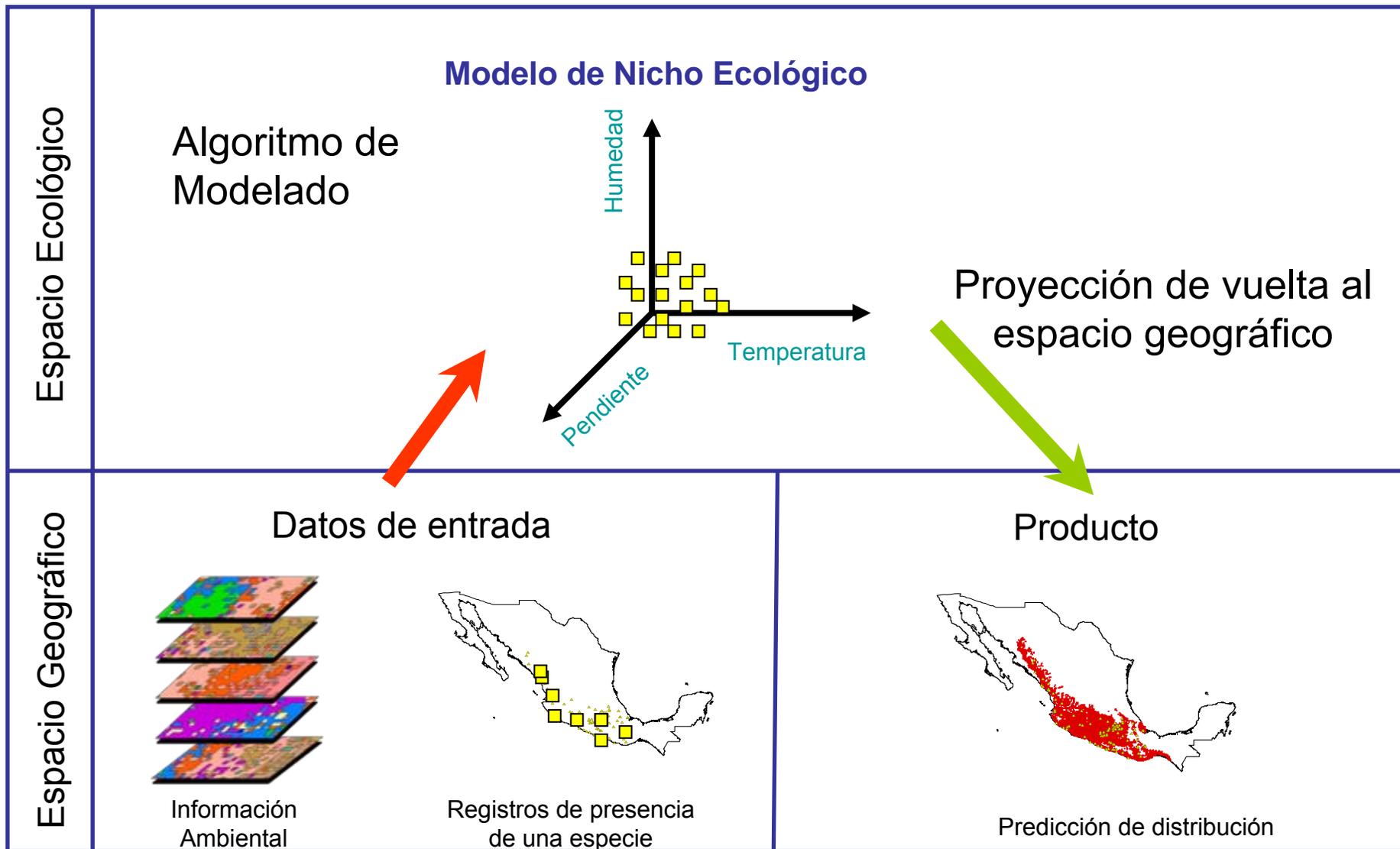
DesktopGarp es una versión adaptada del algoritmo GARP original, que se puede ejecutar en ordenadores personales y en workstations.

Esquema de funcionamiento de un algoritmo genético:

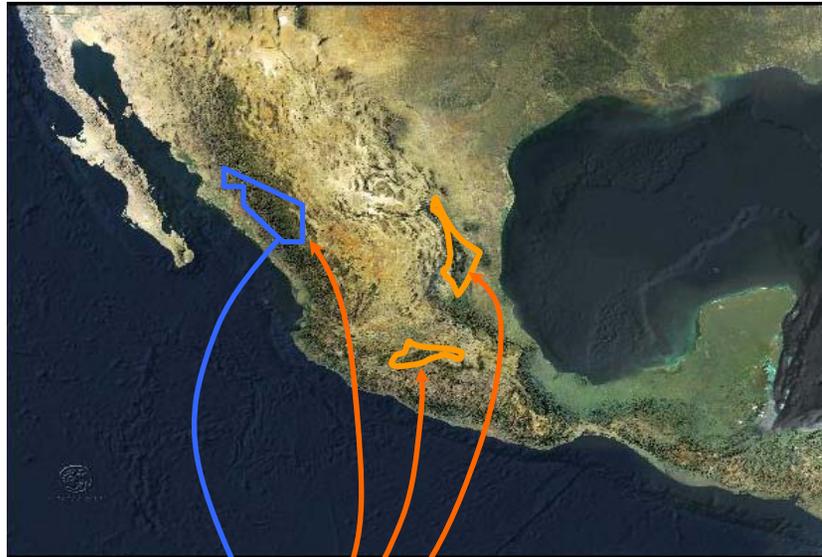


10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

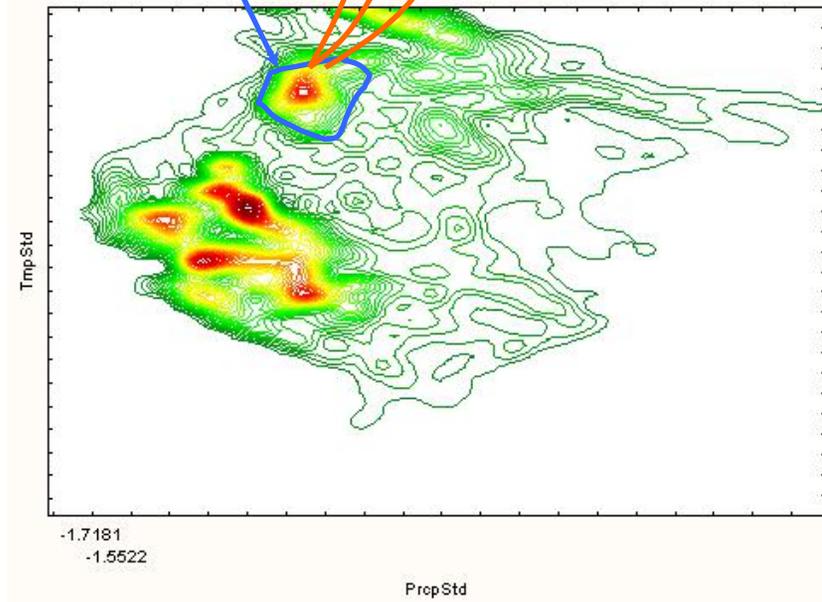
Algoritmo genético para modelos de nicho:



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

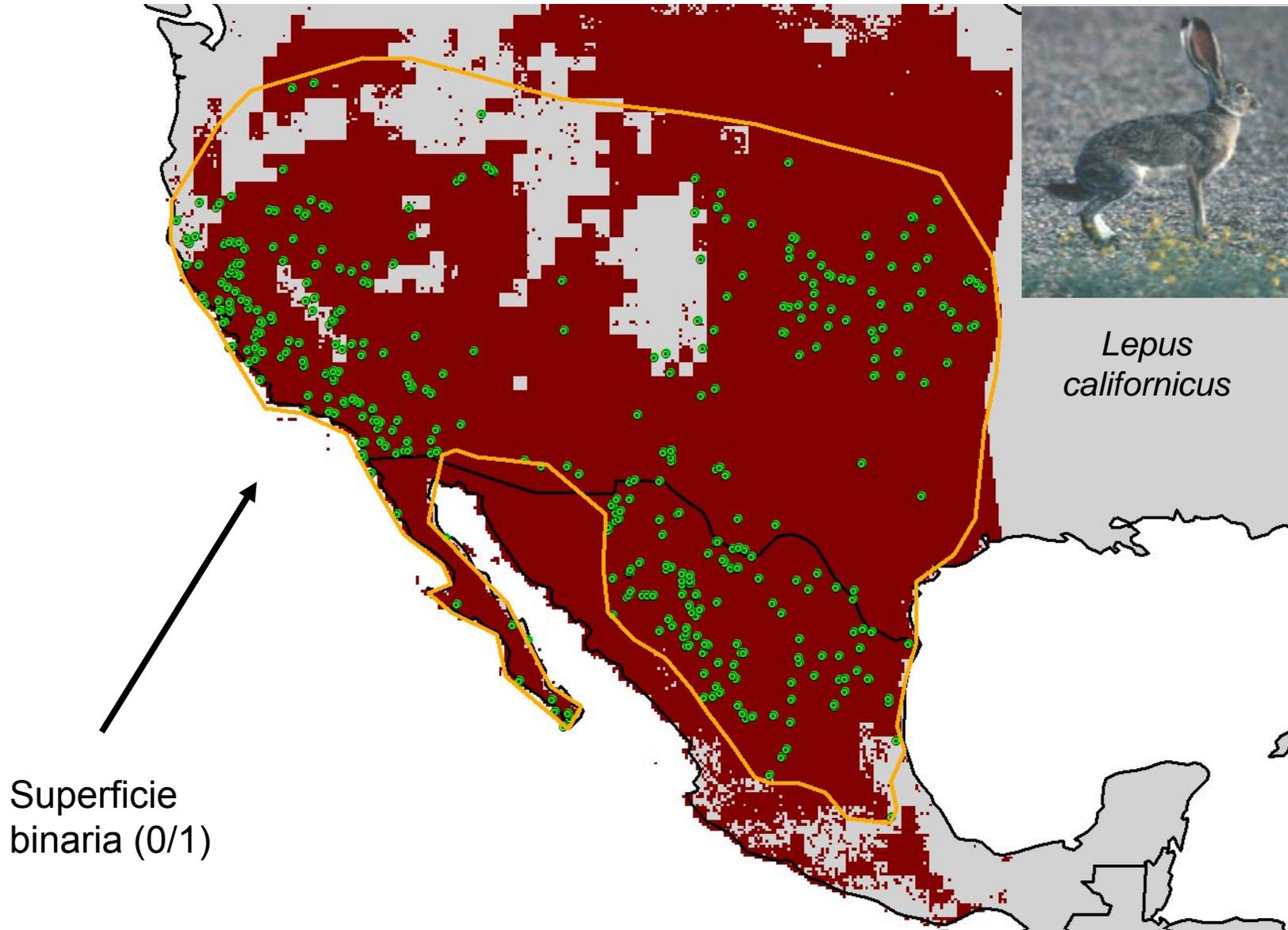


Espacio geográfico
2-dimensiones



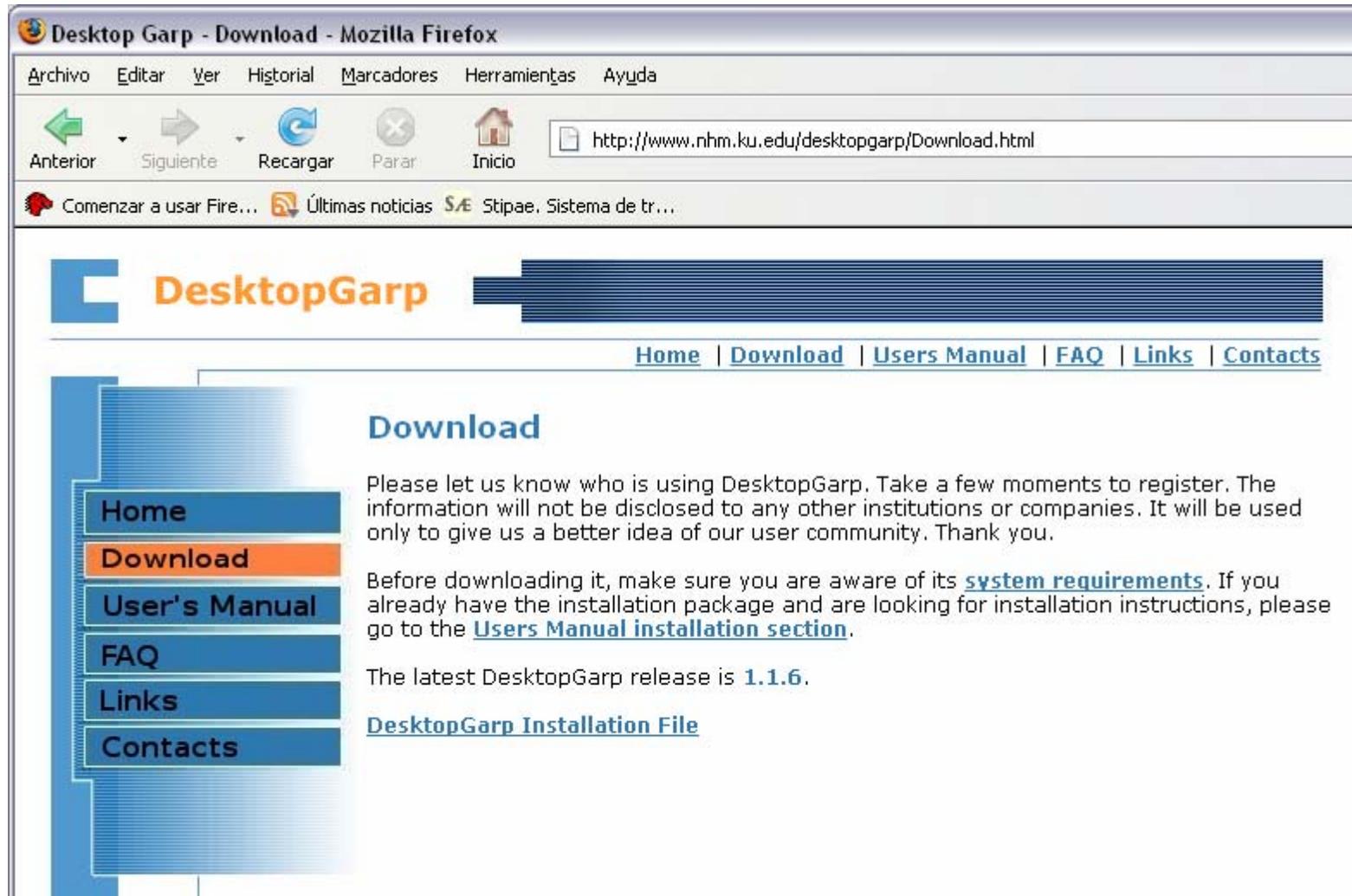
Espacio ecológico
N-dimensiones

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

<http://www.nhm.ku.edu/desktopgarp/Download.html>



The screenshot shows a Mozilla Firefox browser window with the address bar containing the URL <http://www.nhm.ku.edu/desktopgarp/Download.html>. The browser's menu bar includes Archivo, Editar, Ver, Historial, Marcadores, Herramientas, and Ayuda. The navigation bar features buttons for Anterior, Siguiente, Recargar, Parar, and Inicio. The browser's status bar shows "Comenzar a usar Fire..." and "Últimas noticias SAE Stipae. Sistema de tr...".

The website header features the DesktopGarp logo and a navigation menu with links for [Home](#), [Download](#), [Users Manual](#), [FAQ](#), [Links](#), and [Contacts](#). A vertical sidebar on the left contains buttons for Home, Download, User's Manual, FAQ, Links, and Contacts, with the Download button highlighted in orange.

Download

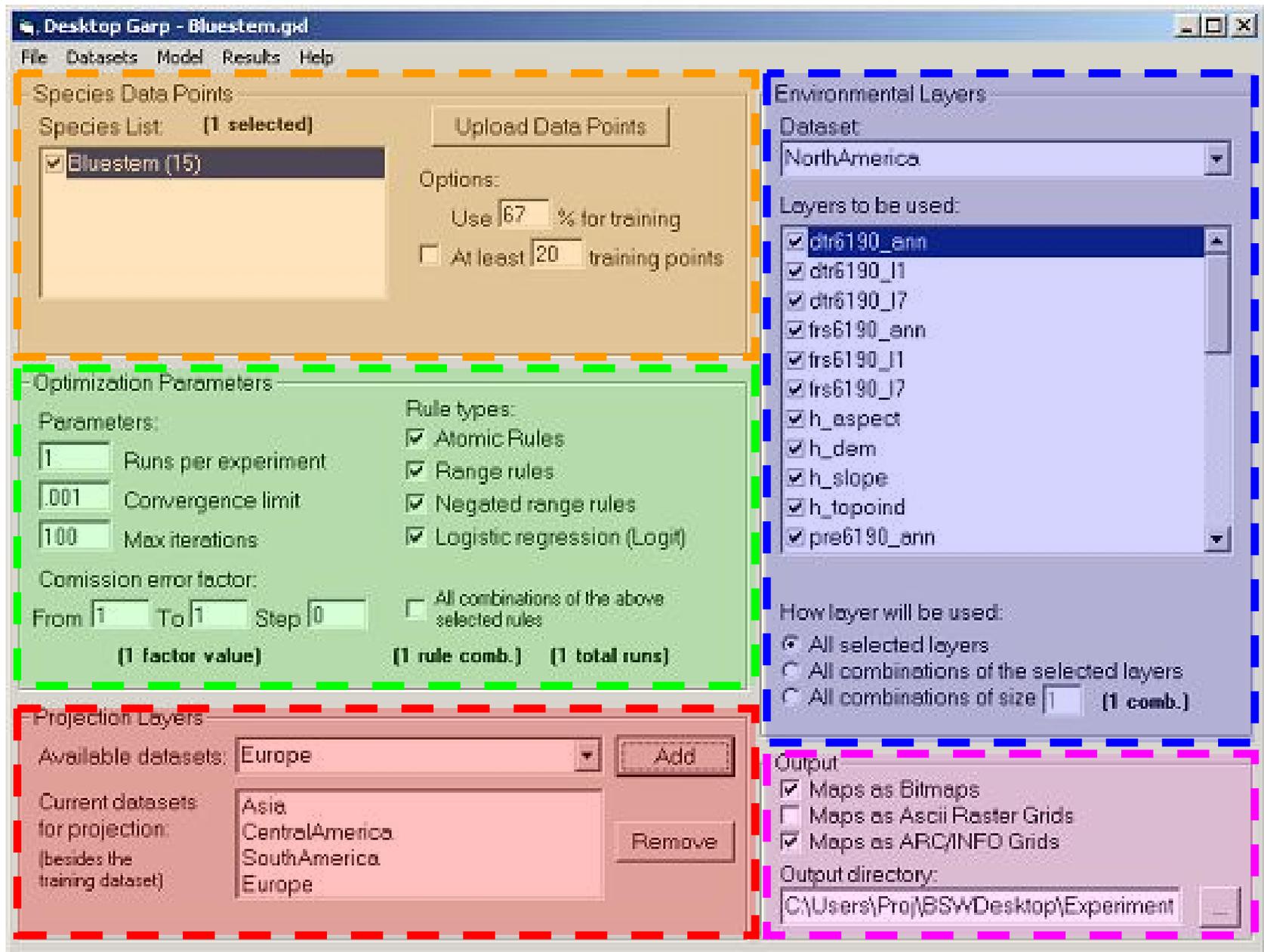
Please let us know who is using DesktopGarp. Take a few moments to register. The information will not be disclosed to any other institutions or companies. It will be used only to give us a better idea of our user community. Thank you.

Before downloading it, make sure you are aware of its [system requirements](#). If you already have the installation package and are looking for installation instructions, please go to the [Users Manual installation section](#).

The latest DesktopGarp release is **1.1.6**.

[DesktopGarp Installation File](#)

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

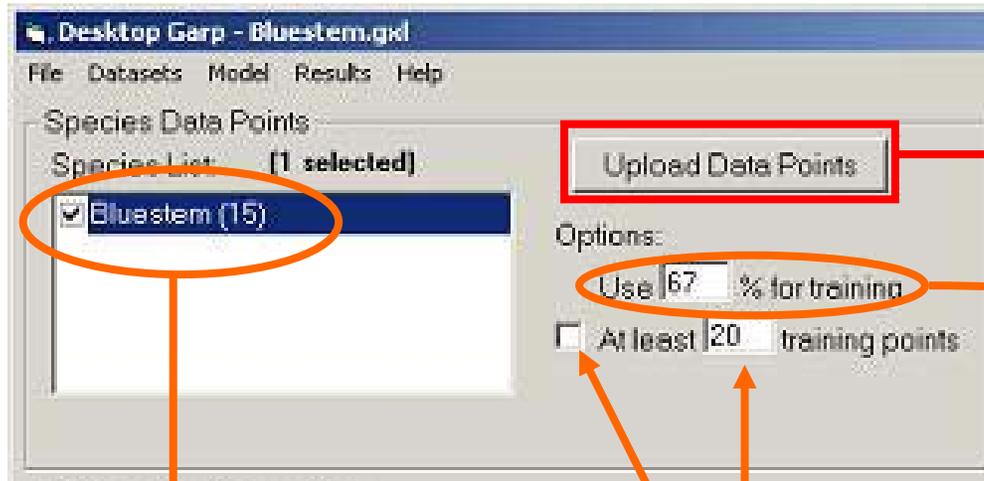


10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

The screenshot displays the GARP software interface with the following sections:

- Species Data Points:** A list with "Bluestem (15)" selected. An "Upload Data Points" button is present. Options include "Use 67 % for training" and "At least 20 training points".
- Optimization Parameters:** Includes input fields for "Runs per experiment" (1), "Convergence limit" (.001), and "Max iterations" (100). It also features "Comission error factor" settings (From 1, To 1, Step 0) and "Rule types" (Atomic Rules, Range rules, Negated range rules, Logistic regression (Logit), and All combinations of the above selected rules).
- Environmental Layers:** A "Dataset" dropdown is set to "NorthAmerica". A list of layers to be used includes "dtr6190_ann", "dtr6190_I1", "dtr6190_I7", "frs6190_ann", "frs6190_I1", "frs6190_I7", "h_aspect", "h_dem", "h_slope", "h_topoind", and "pre6190_ann".
- Projection Layers:** An "Available datasets" dropdown is set to "Europe". "Current datasets for projection" include "Asia", "CentralAmerica", "SouthAmerica", and "Europe".
- Output:** Includes checkboxes for "Maps as Bitmaps", "Maps as Ascii Raster Grids", and "Maps as ARC/INFO Grids". The "Output directory" is set to "C:\Users\Proj\BSW\Desktop\Experiment".

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP



Entrada de los datos de presencia (muestreados)

Puntos a utilizar para "entrenar" el modelo

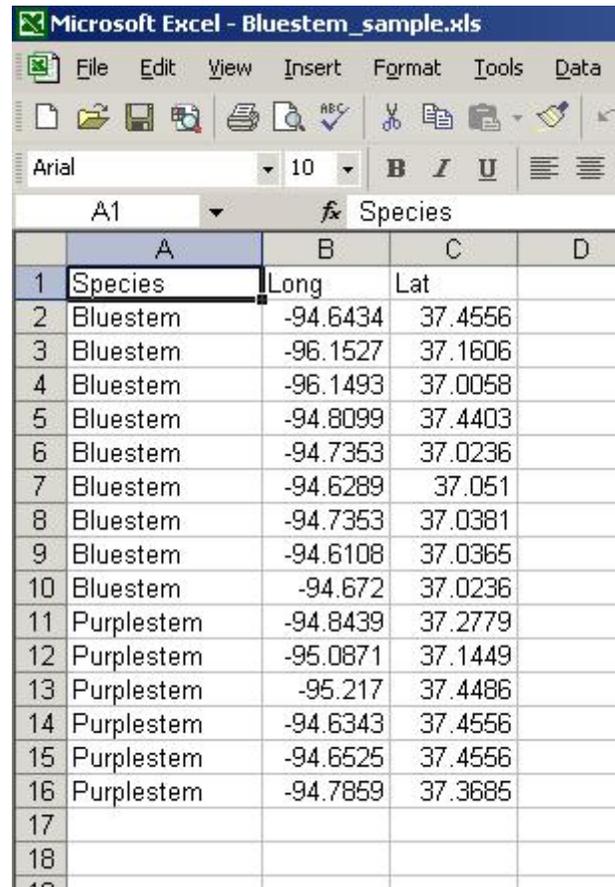
Información de los datos introducidos (y número de puntos para cada especie)

Número mínimo de puntos a utilizar en el test de significancia (si está seleccionada, prevalece sobre el %, que sería 33 en este caso)

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

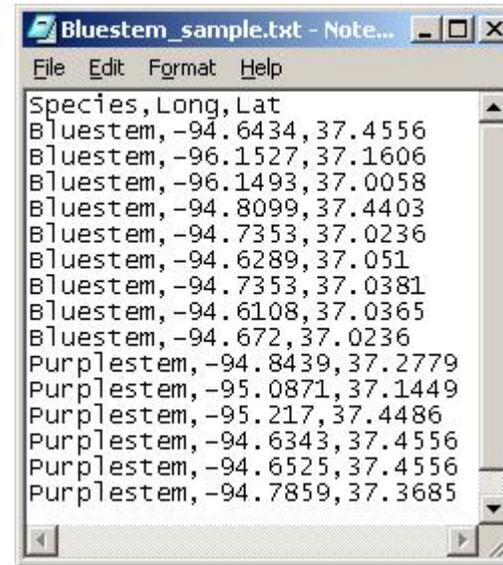
Los datos (puntos de muestreo) se introducen en formato:

MS Excel



	A	B	C	D
1	Species	Long	Lat	
2	Bluestem	-94.6434	37.4556	
3	Bluestem	-96.1527	37.1606	
4	Bluestem	-96.1493	37.0058	
5	Bluestem	-94.8099	37.4403	
6	Bluestem	-94.7353	37.0236	
7	Bluestem	-94.6289	37.051	
8	Bluestem	-94.7353	37.0381	
9	Bluestem	-94.6108	37.0365	
10	Bluestem	-94.672	37.0236	
11	Purplestem	-94.8439	37.2779	
12	Purplestem	-95.0871	37.1449	
13	Purplestem	-95.217	37.4486	
14	Purplestem	-94.6343	37.4556	
15	Purplestem	-94.6525	37.4556	
16	Purplestem	-94.7859	37.3685	
17				
18				
19				

.csv



```
Species,Long,Lat
Bluestem,-94.6434,37.4556
Bluestem,-96.1527,37.1606
Bluestem,-96.1493,37.0058
Bluestem,-94.8099,37.4403
Bluestem,-94.7353,37.0236
Bluestem,-94.6289,37.051
Bluestem,-94.7353,37.0381
Bluestem,-94.6108,37.0365
Bluestem,-94.672,37.0236
Purplestem,-94.8439,37.2779
Purplestem,-95.0871,37.1449
Purplestem,-95.217,37.4486
Purplestem,-94.6343,37.4556
Purplestem,-94.6525,37.4556
Purplestem,-94.7859,37.3685
```

coberturas ArcView



3 columnas:
Species, Long, Lat

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

The screenshot displays the GARP software interface with the following sections:

- Species Data Points:** A 'Species List' with 'Blueslem (15)' selected. An 'Upload Data Points' button is present. Options include 'Use 67 % for training' and 'At least 20 training points'.
- Optimization Parameters:** Includes input fields for 'Runs per experiment' (1), 'Convergence limit' (.001), and 'Max iterations' (100). It also features a 'Comission error factor' section with 'From 1 To 1 Step 0' and a 'Rule types' section with checked options for 'Atomic Rules', 'Range rules', 'Negated range rules', and 'Logistic regression (Logit)'. A summary shows '(1 rule comb.) (1 total runs)'.
- Projection Layers:** Shows 'Available datasets' (Europe) and 'Current datasets for projection' (Asia, CentralAmerica, SouthAmerica, Europe).
- Environmental Layers:** A dashed blue box highlights this section. It includes a 'Dataset' dropdown (NorthAmerica) and a list of 'Layers to be used' with checked items: dtr6190_ann, dtr6190_I1, dtr6190_I7, frs6190_ann, frs6190_I1, frs6190_I7, h_aspect, h_dem, h_slope, h_topoind, and pre6190_ann. Below, 'How layer will be used' is set to 'All selected layers'.
- Output:** Includes checkboxes for 'Maps as Bitmaps', 'Maps as Ascii Raster Grids', and 'Maps as ARC/INFO Grids'. The 'Output directory' is set to 'C:\Users\Proj\BSW\Desktop\Experiment'.

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Menú: "Datasets->Scan directory...".

Dataset: nombre de la carpeta

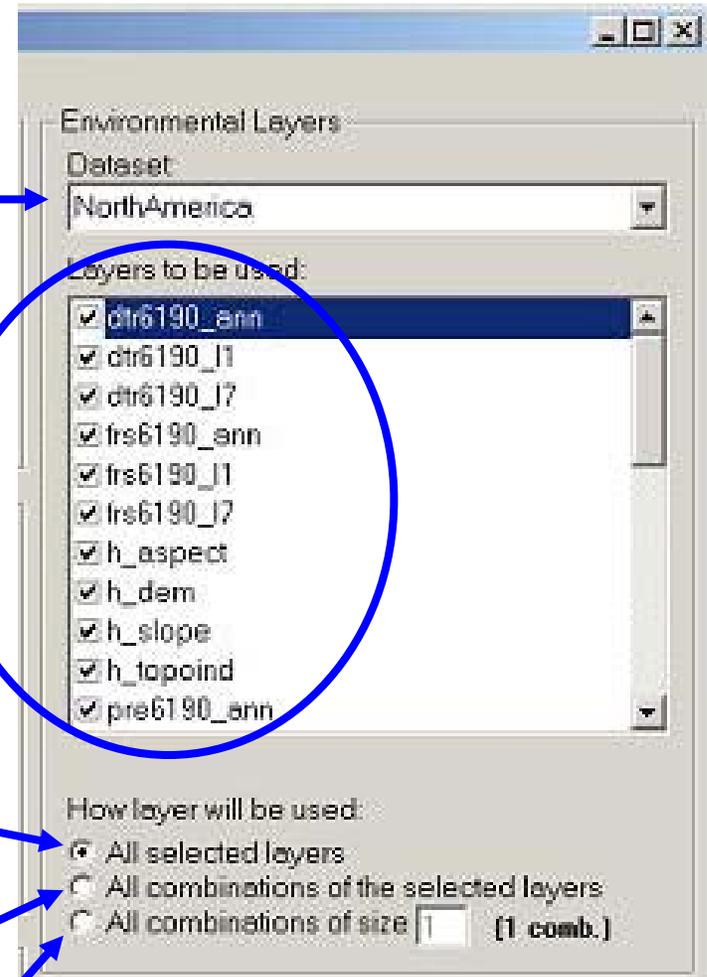
IMP: tiene que haber una cobertura que se llame "MASK"

Lista de coberturas ambientales

fuerza a DesktopGarp a utilizar todas la coberturas en la optimización

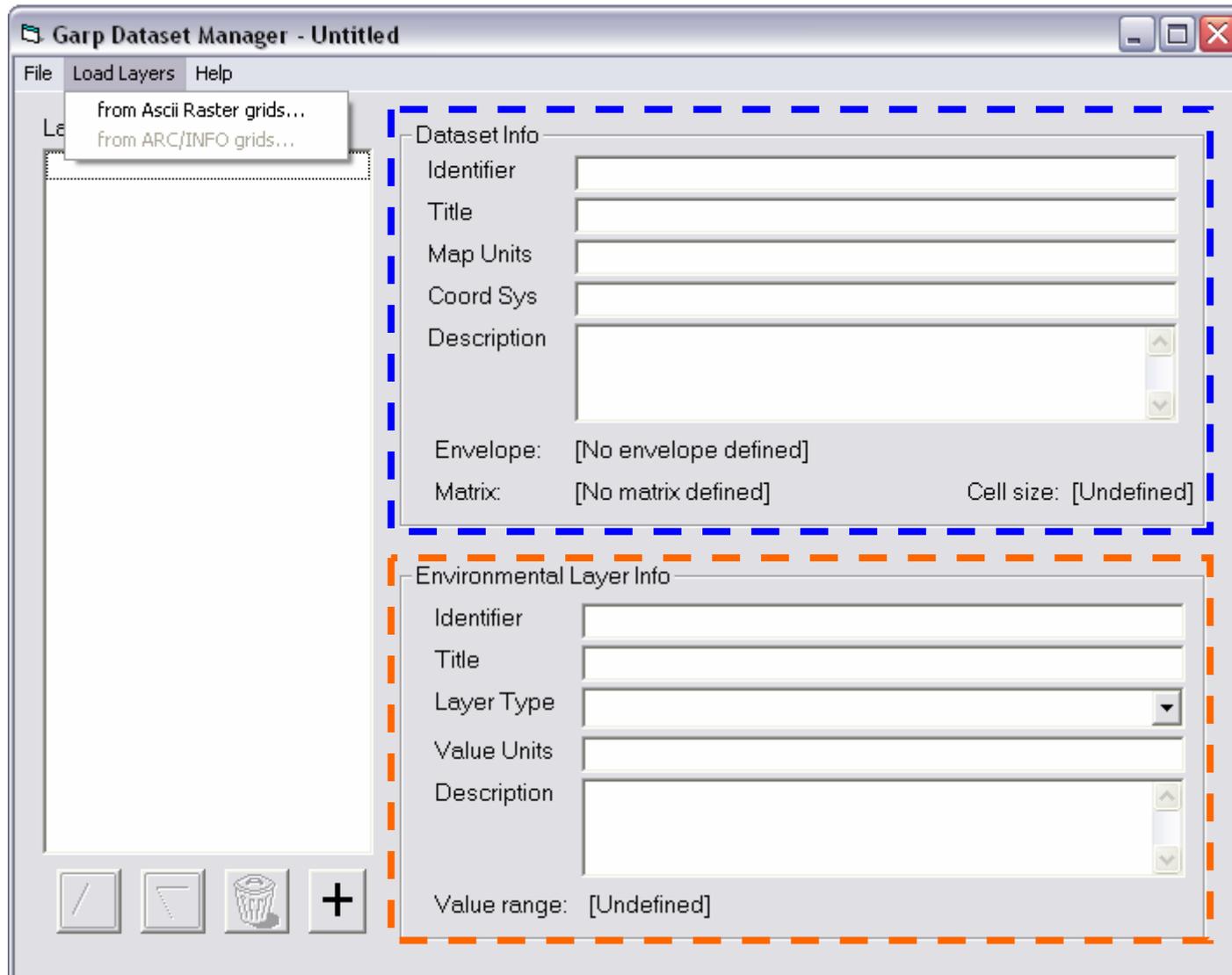
forzará al experimento a ejecutar el modelo una vez por cada una de las combinaciones de coberturas posibles.

limitará el experimento a las combinaciones que contengan exactamente N coberturas.



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

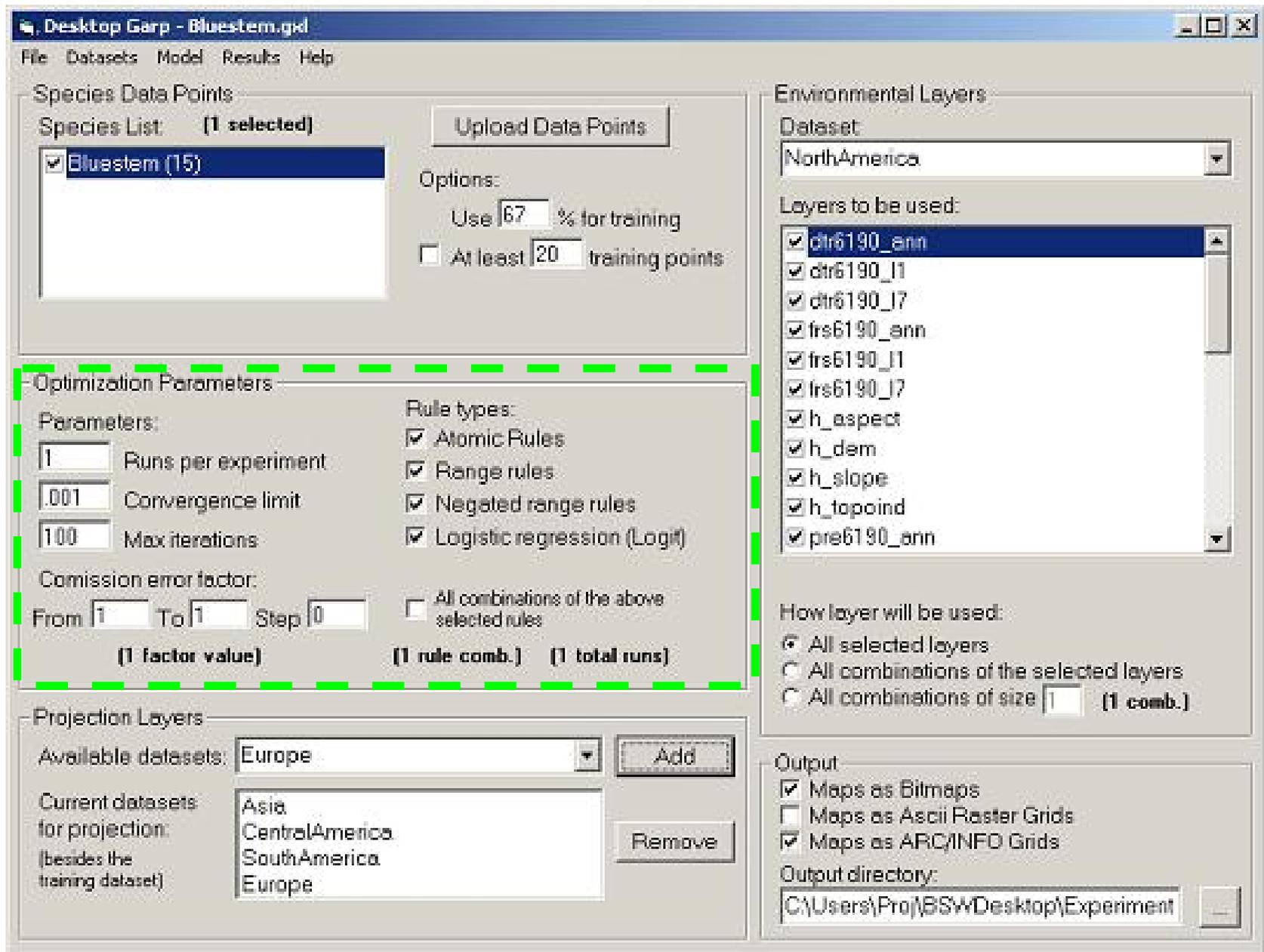
Garp Dataset Manager:



**Metadatos
conjunto
coberturas**

**Metadatos para
cada cobertura**

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

- 1) Número de iteraciones (nº total, ¡no por especie!)
- 2) Límite de convergencia (valores entre 0.01 y 0.10 son adecuados, si es 0 el algoritmo itera hasta que llega al nº máximo de iteraciones)
- 3) Nº máximo de iteraciones (prevalece sobre lim. Convergencia, valores normales 100 - 1000)

Optimization Parameters

Parameters:

1 Runs per experiment

.001 Convergence limit

100 Max iterations

Commission error factor:

From 1 To 1 Step 0

[1 factor value]

Rule types:

Atomic Rules

Range rules

Negated range rules

Logistic regression (Logit)

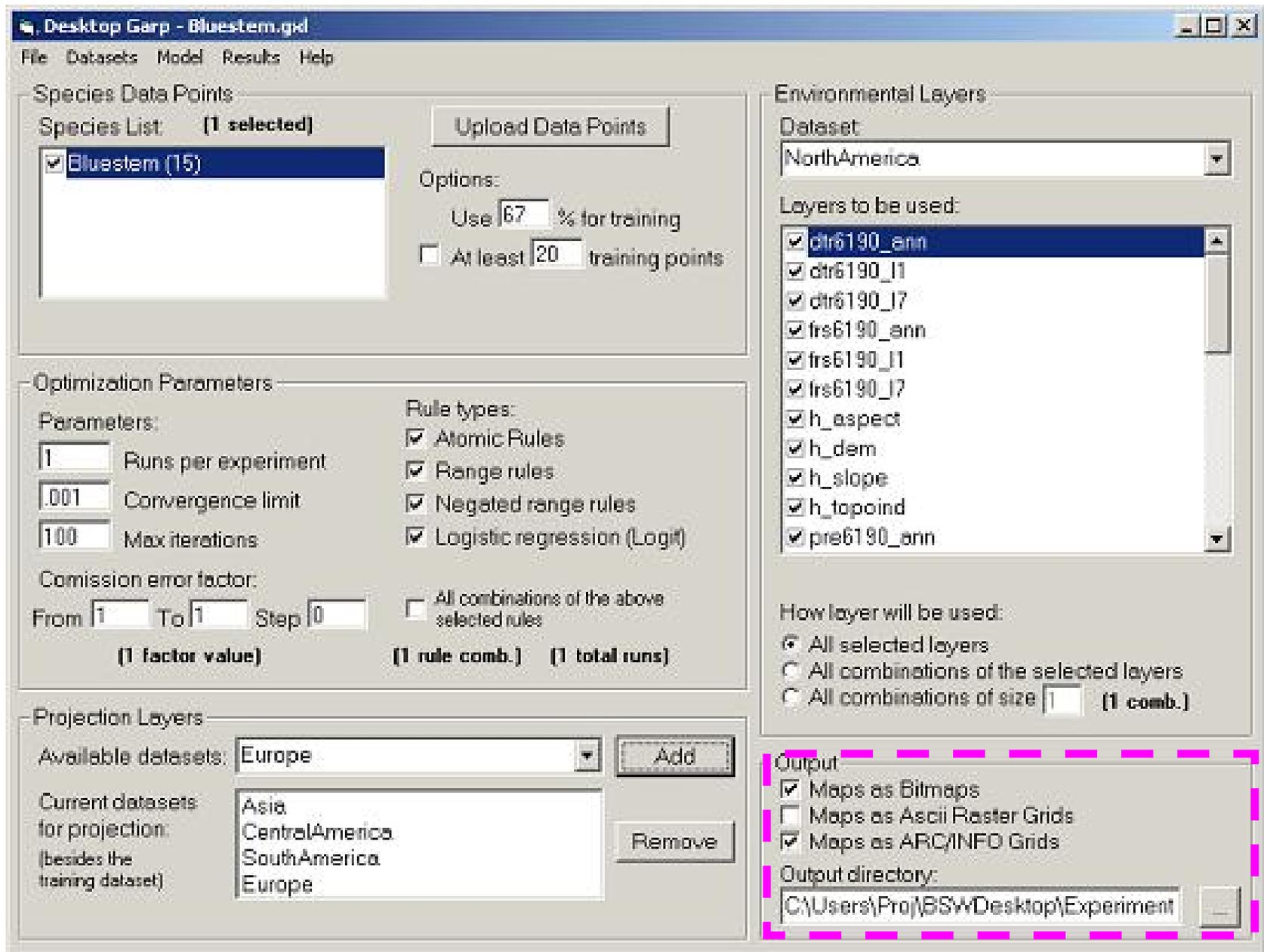
All combinations of the above selected rules

[1 rule comb.] [1 total runs]

Reglas a utilizar

(y si se utilizan todas las combinaciones posibles)

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Todos los resultados se almacenan con un nombre numérico secuencial (cada mapa es el resultado de una de las iteraciones del modelo).

Bitmaps: no permiten hacer postproceso

ARC/INFO Grids: El directorio grid00100, contiene el resultado de las áreas que van de la 100 a la 199. Esto es debido a una limitación de ESRI con las grids permitidas en un directorio.

Tiene que ser una carpeta vacía



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

The screenshot displays the GARP software interface with the following sections:

- Species Data Points:** Includes a 'Species List' with 'Bluestem (15)' selected, an 'Upload Data Points' button, and options for training percentage (67%) and minimum training points (20).
- Optimization Parameters:** Contains input fields for 'Runs per experiment' (1), 'Convergence limit' (.001), and 'Max iterations' (100). It also features a 'Commission error factor' section with 'From' (1), 'To' (1), and 'Step' (0) values. On the right, 'Rule types' are checked: Atomic Rules, Range rules, Negated range rules, and Logistic regression (Logit). A summary shows '(1 rule comb.)' and '(1 total runs)'. An unchecked option for 'All combinations of the above selected rules' is also present.
- Environmental Layers:** Shows a 'Dataset' dropdown set to 'NorthAmerica' and a list of 'Layers to be used' including 'dtr6190_ann', 'dtr6190_I1', 'dtr6190_I7', 'frs6190_ann', 'frs6190_I1', 'frs6190_I7', 'h_aspect', 'h_dem', 'h_slope', 'h_topoind', and 'pre6190_ann'. Below this, 'How layer will be used' is set to 'All selected layers'.
- Projection Layers:** This section is highlighted with a red dashed border. It shows 'Available datasets' with 'Europe' selected and an 'Add' button. The 'Current datasets for projection' list includes 'Asia', 'CentralAmerica', 'SouthAmerica', and 'Europe', with a 'Remove' button.
- Output:** Includes checkboxes for 'Maps as Bitmaps' (checked), 'Maps as Ascii Raster Grids' (unchecked), and 'Maps as ARC/INFO Grids' (checked). The 'Output directory' is set to 'C:\Users\Proj\BSW\Desktop\Experiment'.

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

El hecho de utilizar distintas coberturas en los experimentos es útil a la hora de investigar especies invasivas, cambios climáticos y cambios temporales.

GARP proyecta las reglas del modelo sobre las coberturas que se especifiquen aquí

NO es obligatorio proyectar

ADD permite añadir las coberturas seleccionadas

REMOVE elimina de la lista de proyección la cobertura seleccionada



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

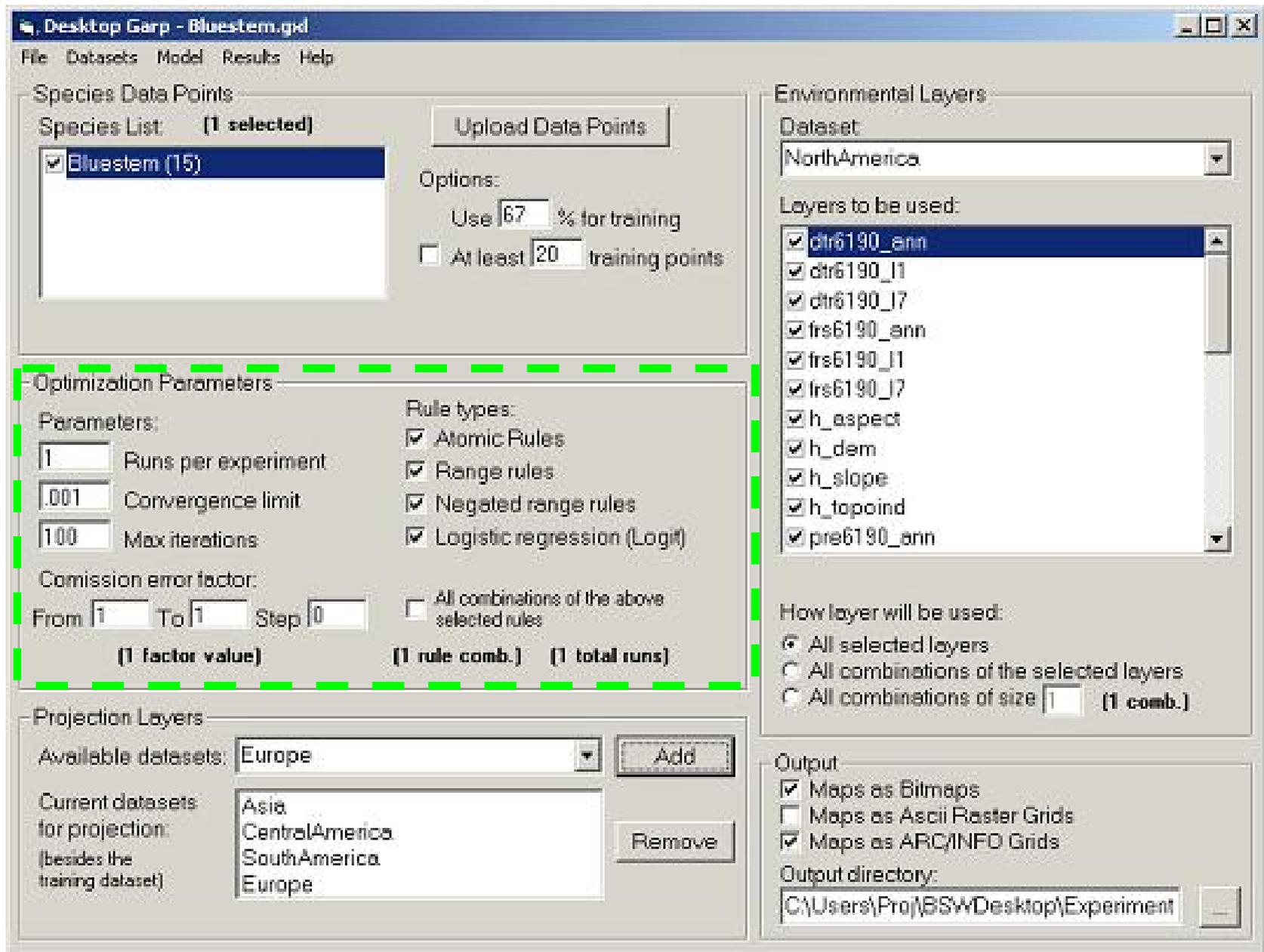
Tabla de Resultados:

Task	Run	Species	Atomic Ru	Range Ru	Negated R	Logit Rule	Iter.	Cov.	CEF	Train Acc	Pr.Pv/Ac	FPr.Ab/Ac	FPr.Pv/Ac	APr.Ab/Ac	APr.Pv/Ac
1	1	1 Anafal	1	1	1	1	1001	0.072469	N/A	0.7772	1173	19	383	770	
2	2	2 Anafal	1	1	1	1	1001	0.073493	N/A	0.7216	1180	40	328	624	
3	3	3 Anafal	1	1	1	1	1001	0.073783	N/A	0.8308	1219	31	392	858	
4	4	4 Anafal	1	1	1	1	1001	0.071702	N/A	0.8108	1197	0	356	830	
5	5	5 Anafal	1	1	1	1	1001	0.072394	N/A	0.8188	1156	0	345	391	
6	6	6 Anafal	1	1	1	1	1001	0.070488	N/A	0.778	1192	22	422	753	
7	7	7 Anafal	1	1	1	1	1001	0.071463	N/A	0.668	1205	0	702	465	
8	8	8 Anafal	1	1	1	1	1001	0.071293	N/A	0.6824	901	50	306	805	
9	9	9 Anafal	1	1	1	1	1001	0.068744	N/A	0.6892	1181	24	602	542	
10	10	10 Anafal	1	1	1	1	1001	0.073028	N/A	0.6588	1191	0	694	456	
11	11	1 Anafal	1	1	1	1	1001	0.026601	N/A	0.794	1226	24	491	759	
12	12	2 Anafal	1	1	1	1	1001	0.027146	N/A	0.7904	1221	29	488	756	
13	13	3 Anafal	1	1	1	1	1001	0.026743	N/A	0.7894	1195	27	401	801	
14	14	4 Anafal	1	1	1	1	1001	0.02654	N/A	0.8176	1219	31	425	825	
15	15	5 Anafal	1	1	1	1	1001	0.027818	N/A	0.776	1250	0	470	690	
16	16	6 Anafal	1	1	1	1	1001	0.025603	N/A	0.712	1250	0	718	530	
17	17	7 Anafal	1	1	1	1	1001	0.025292	N/A	0.73	1227	0	665	698	
18	18	8 Anafal	1	1	1	1	1001	0.024598	N/A	0.8016	1232	0	394	772	
19	19	9 Anafal	1	1	1	1	1001	0.026437	N/A	0.7804	1174	0	378	777	
20	20	10 Anafal	1	1	1	1	1001	0.026927	N/A	0.758	1250	0	425	645	
21	21	1 Anafal	1	1	1	1	1001	0.010211	N/A	0.8916	1173	77	194	1058	
22	22	2 Anafal	1	1	1	1	1001	0.01217	N/A	0.8524	1180	70	289	951	
23	23	3 Anafal	1	1	1	1	1001	0.012321	N/A	0.7872	1150	100	431	818	
24	24	4 Anafal	1	1	1	1	1001	0.010488	N/A	0.8808	1206	44	254	996	
25	25	5 Anafal	1	1	1	1	1001	0.011493	N/A	0.83	1152	98	327	923	
26	26	6 Anafal	1	1	1	1	1001	0.010519	N/A	0.8252	1161	89	348	902	
27	27	7 Anafal	1	1	1	1	1001	0.010548	N/A	0.8356	1201	49	362	888	
28	28	8 Anafal	1	1	1	1	1001	0.011001	N/A	0.842	1141	108	285	954	
29	29	9 Anafal	1	1	1	1	1001	0.010049	N/A	0.8556	1250	0	358	889	
30	30	10 Anafal	1	1	1	1	1001	0.012753	N/A	0.8288	1131	119	314	936	
31	31	1 Anafal	1	1	1	1	558	0.004973	N/A	0.8252	1131	119	315	932	
32	32	2 Anafal	1	1	1	1	580	0.004997	N/A	0.8148	1195	55	408	842	
33	33	3 Anafal	1	1	1	1	558	0.004993	N/A	0.8404	1116	134	365	985	
34	34	4 Anafal	1	1	1	1	594	0.004951	N/A	0.8372	1000	250	157	1093	

Resultado importante:

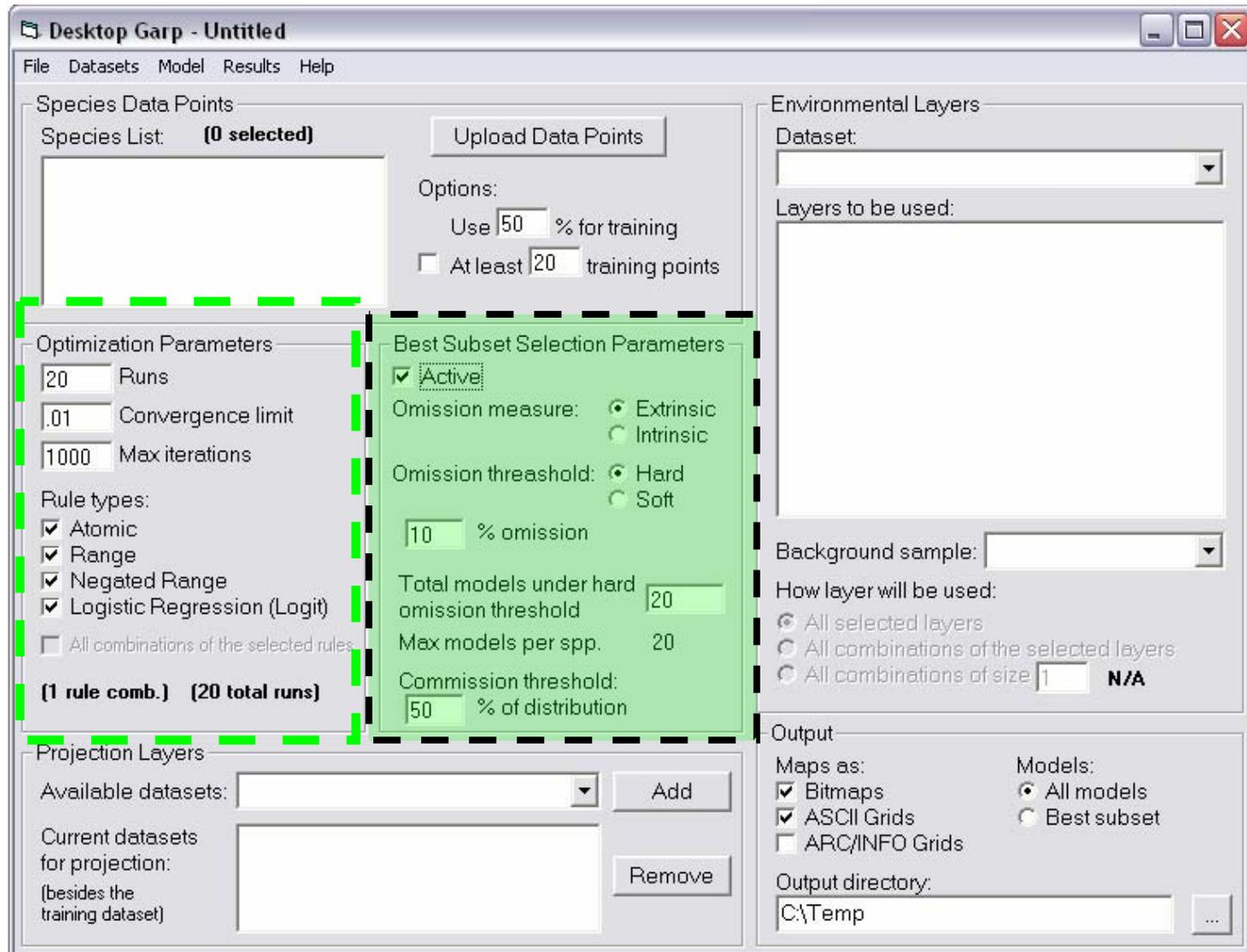
p: Test de Chi-cuadrado. Probabilidad de que una predicción al azar sea similar al resultado, esto es, teniendo el mismo número de puntos correctos que el modelo que ha generado GAP.

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

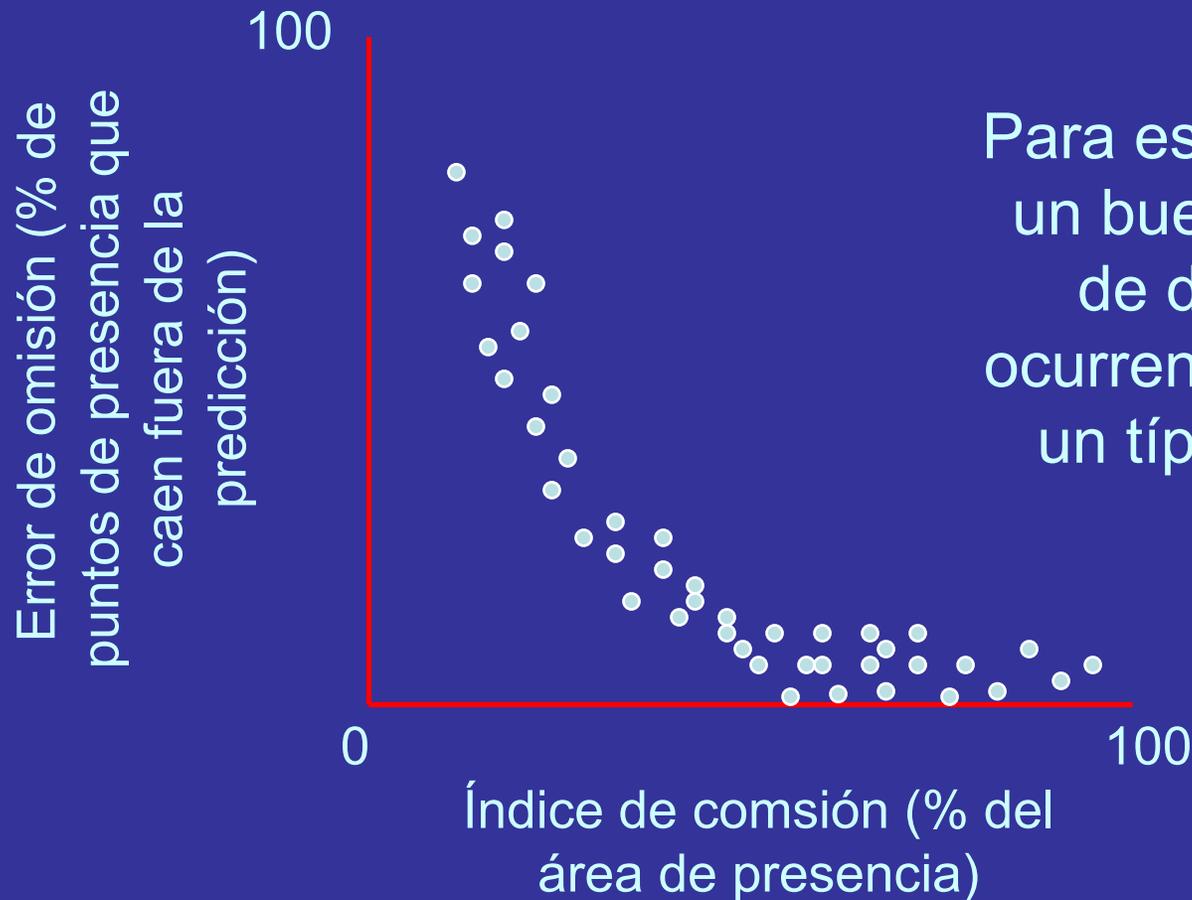


10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

En realidad falta el panel “Best Subset Selection parameter”:

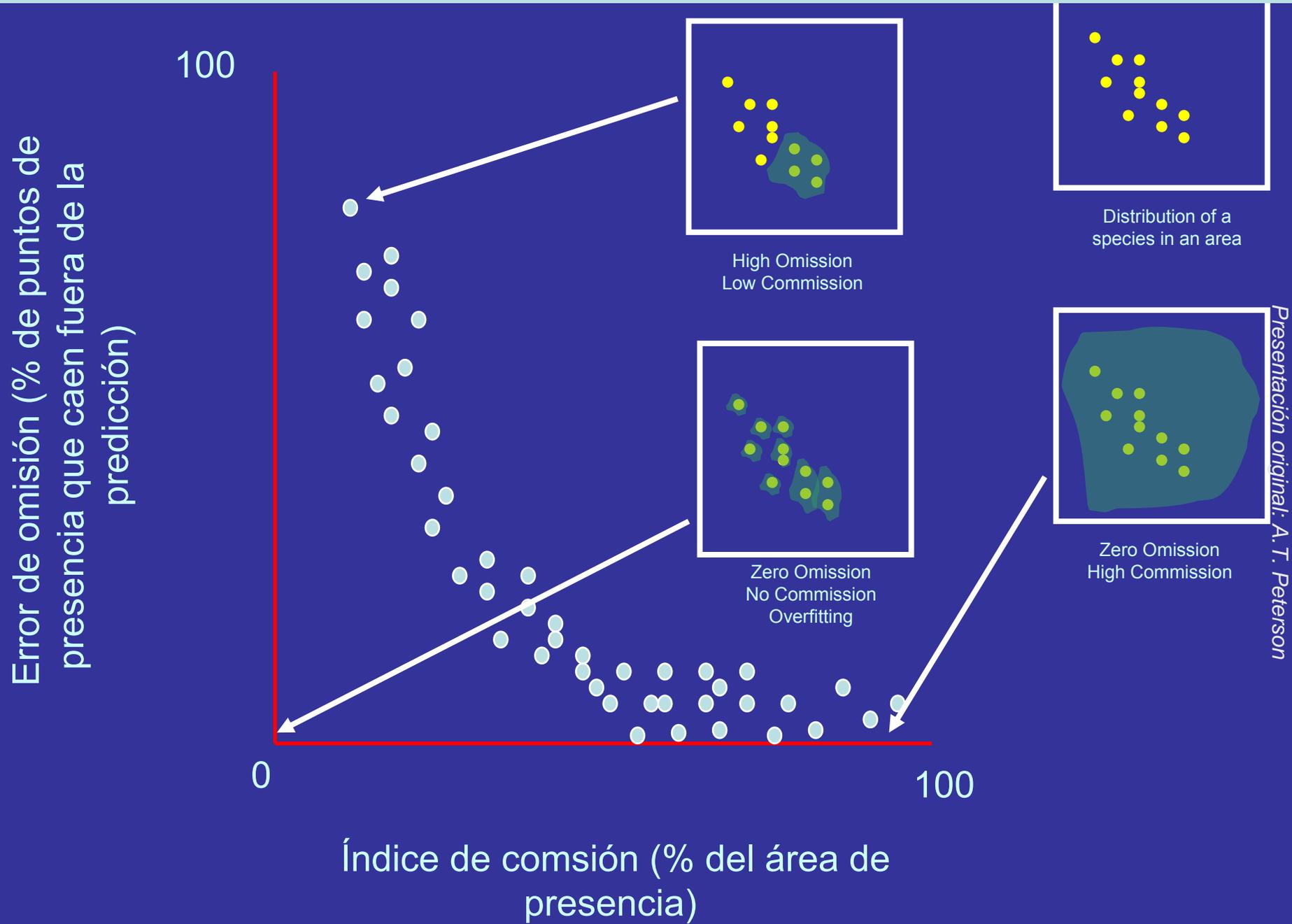


Algunos algoritmos como GARP produce diferentes modelos con los mismos datos de entrada. Si producimos varios modelos, podemos calcular sus errores y graficar la omisión/comisión



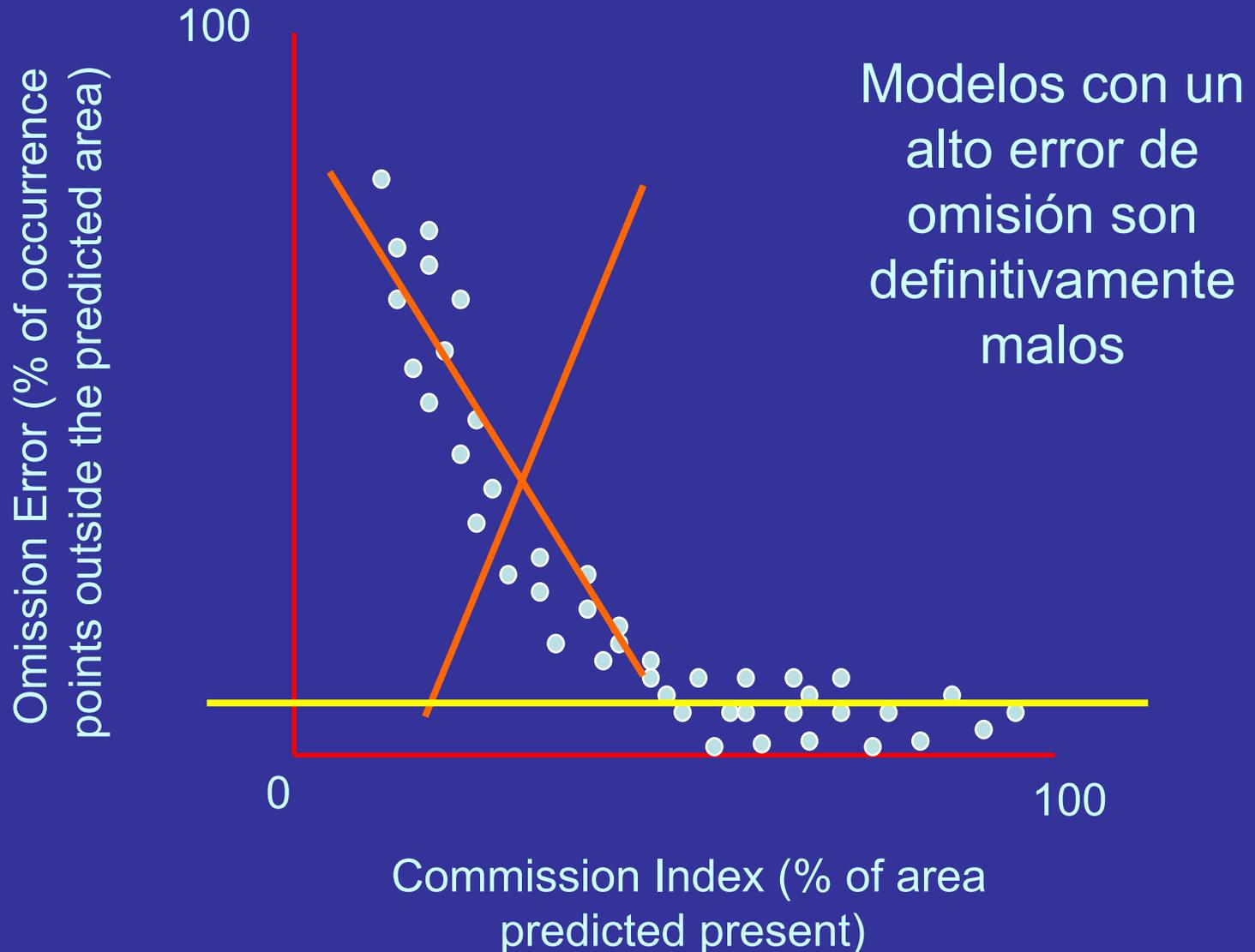
Para especies con un buen número de datos de ocurrencia esta es una típica curva

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP



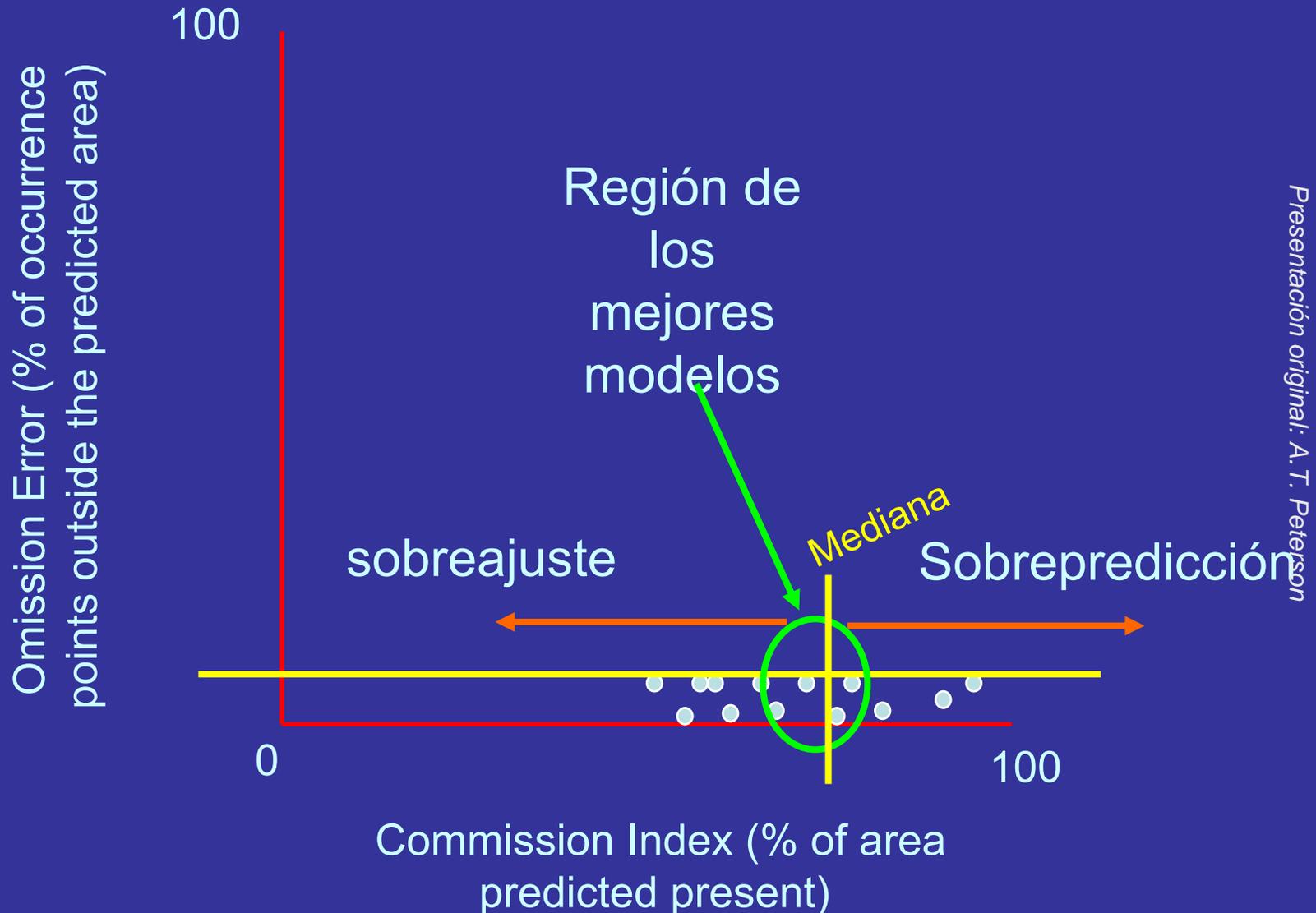
10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

La cuestión ahora es, cual de estos modelos son buenos y cuales malos?



10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

La cuestión ahora es, cual de estos modelos son buenos y cuales malos?



Implementación en Desktop GARP

Teniendo suficientes datos de presencia, se pueden dividir en datos de prueba y de entrenamiento. Cuando este es el caso es conveniente seleccionar *Extrinsic* en la opción *Omission Measure*. Si se utilizan 100% para entrenamiento, se selecciona *Intrinsic*

The screenshot shows the Desktop Garp software interface with the following settings:

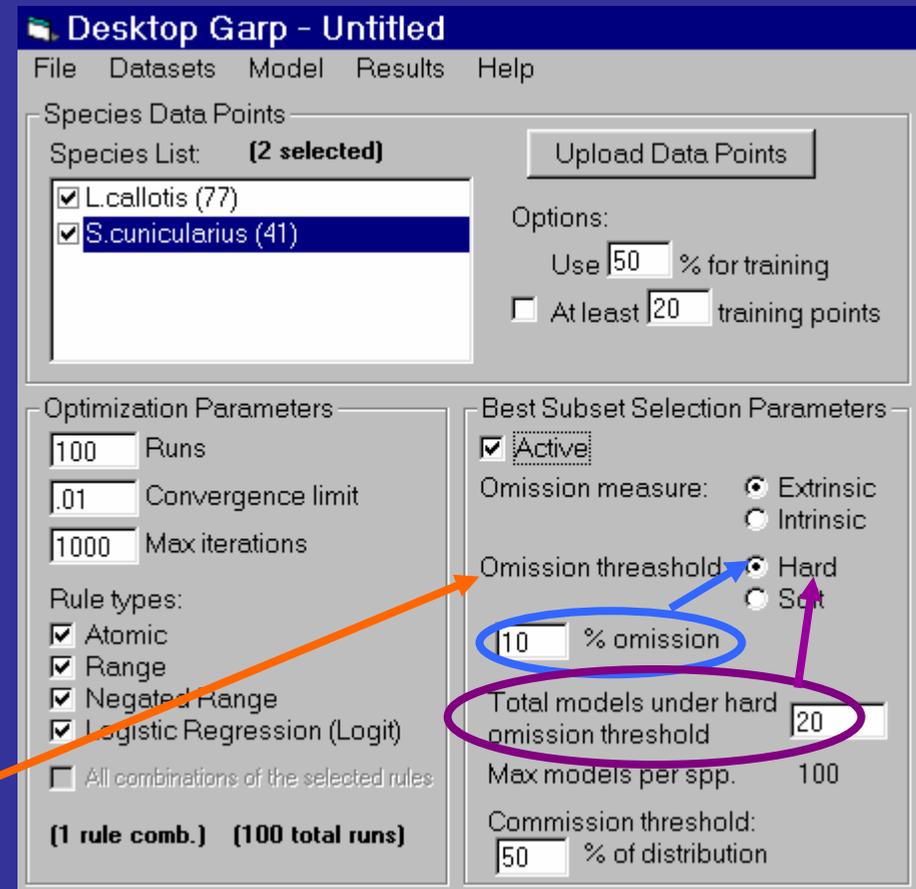
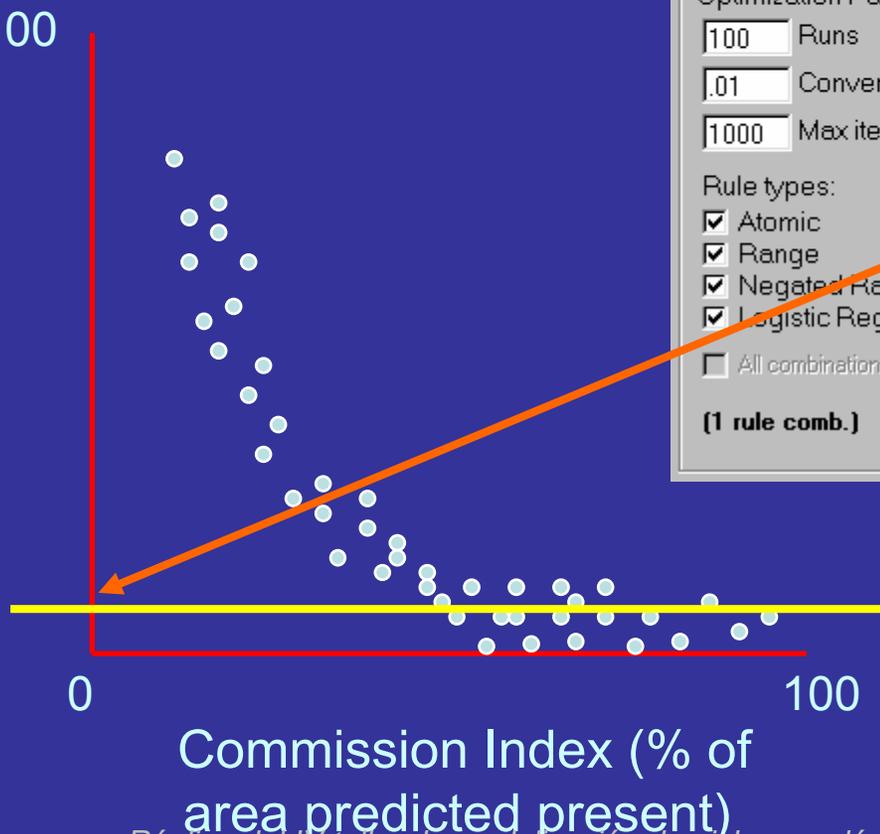
- Species Data Points:** Species List: [2 selected]. Selected species: L.callotis (77) and S.cunicularius (41). An orange arrow points from the 'At least 100 training points' checkbox to the 'Extrinsic' radio button in the Omission measure section.
- Options:** Use 50 % for training. At least 100 training points (checkbox).
- Optimization Parameters:** 100 Runs, .01 Convergence limit, 1000 Max iterations. Rule types: Atomic, Range, Negated Range, Logistic Regression (Logit). (1 rule comb.) (100 total runs).
- Best Subset Selection Parameters:** Active (checkbox). Omission measure: Extrinsic (radio), Intrinsic (radio). Omission threshold: Hard (radio), Soft (radio). 10 % omission. Total models under hard omission threshold: 20. Max models per spp.: 100. Commission threshold: 50 % of distribution.

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Implementación en Desktop GARP

En la sección *Omission threshold*, si se selecciona *Hard*, significa que se usara un valor absoluto en el eje de omission. Se fija ese valor en el % de omission

Omission Error (% of occurrence points outside the predicted area)

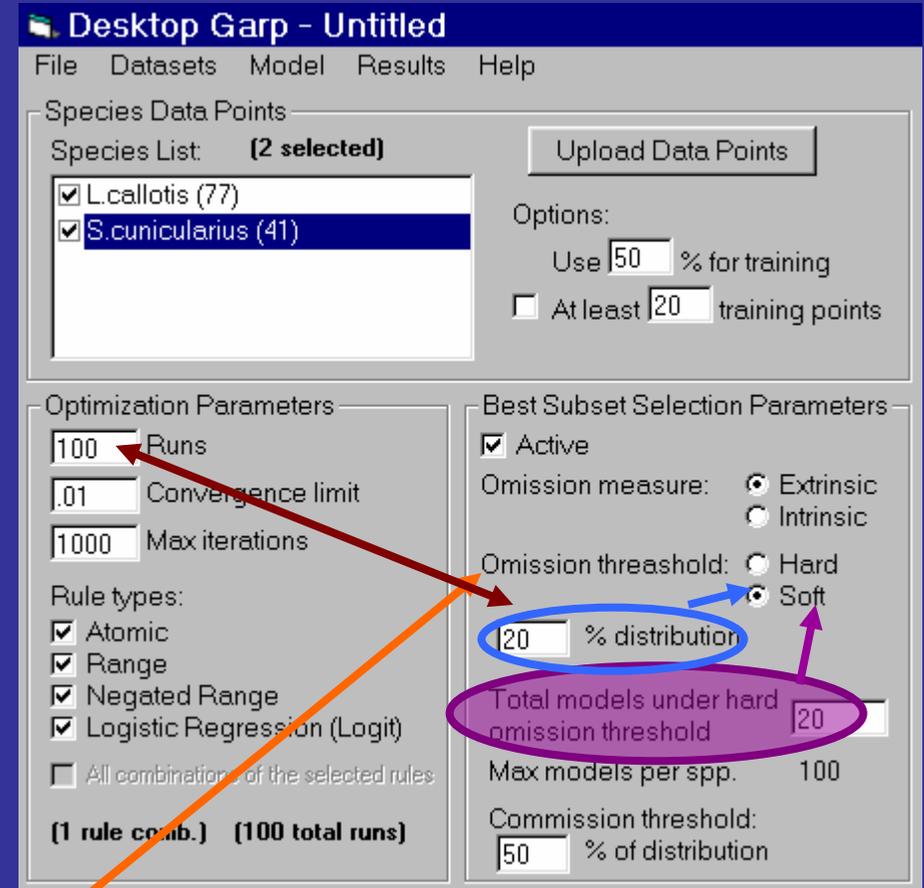
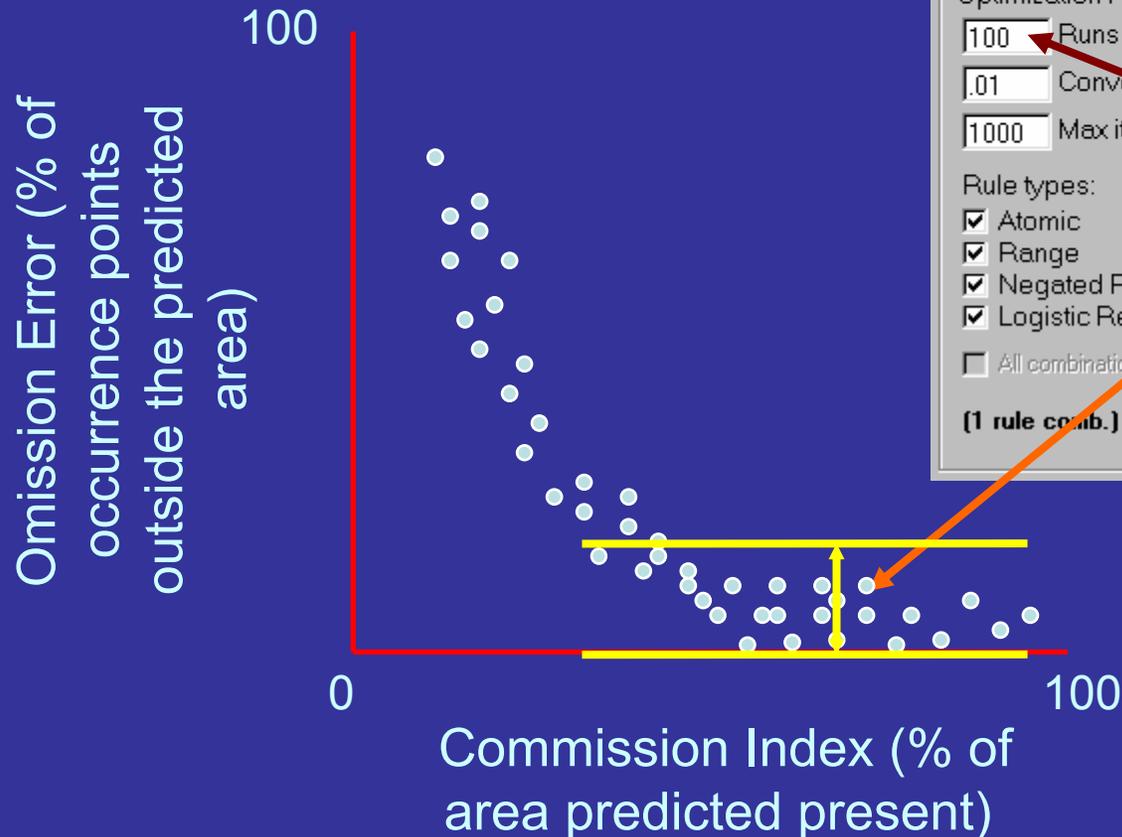


Entonces se selecciona el numero de modelos que se quieren bajo este % de omission threshold

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Implementación en Desktop GARP

Cuando seleccionamos *Soft*, significa que se seleccionarán un cierto número de modelos (en porcentaje), indicado en el % *distribution*, con la menor omisión. Esto es útil cuando se trabaja más de una especie a la vez

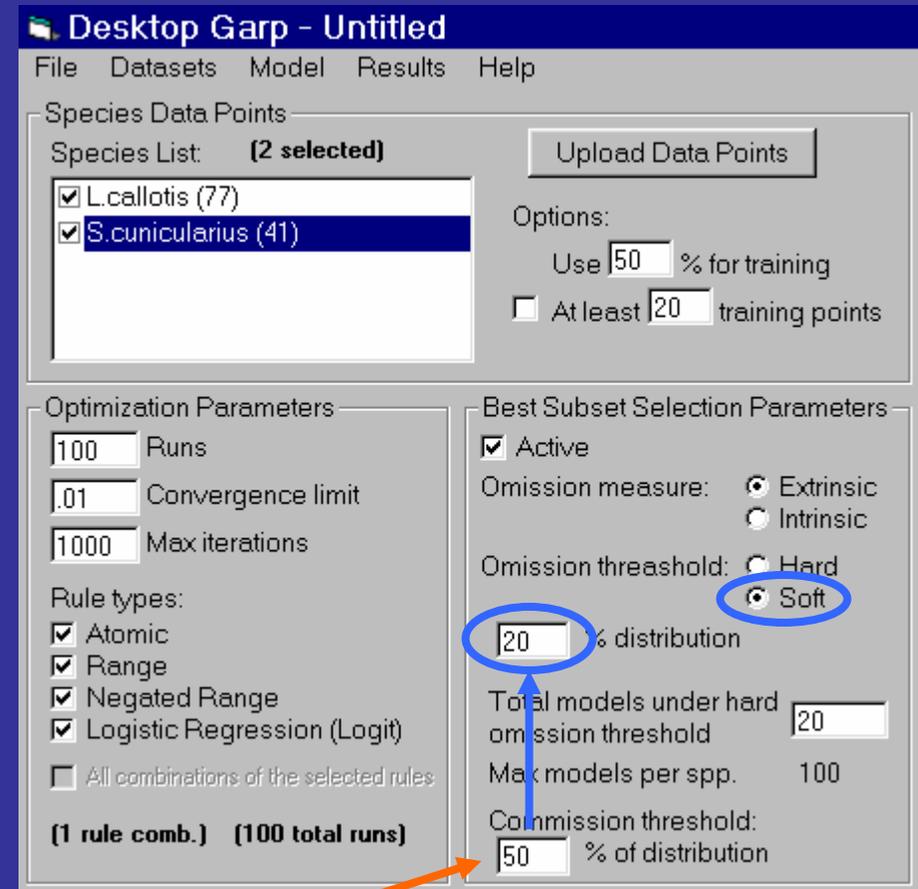
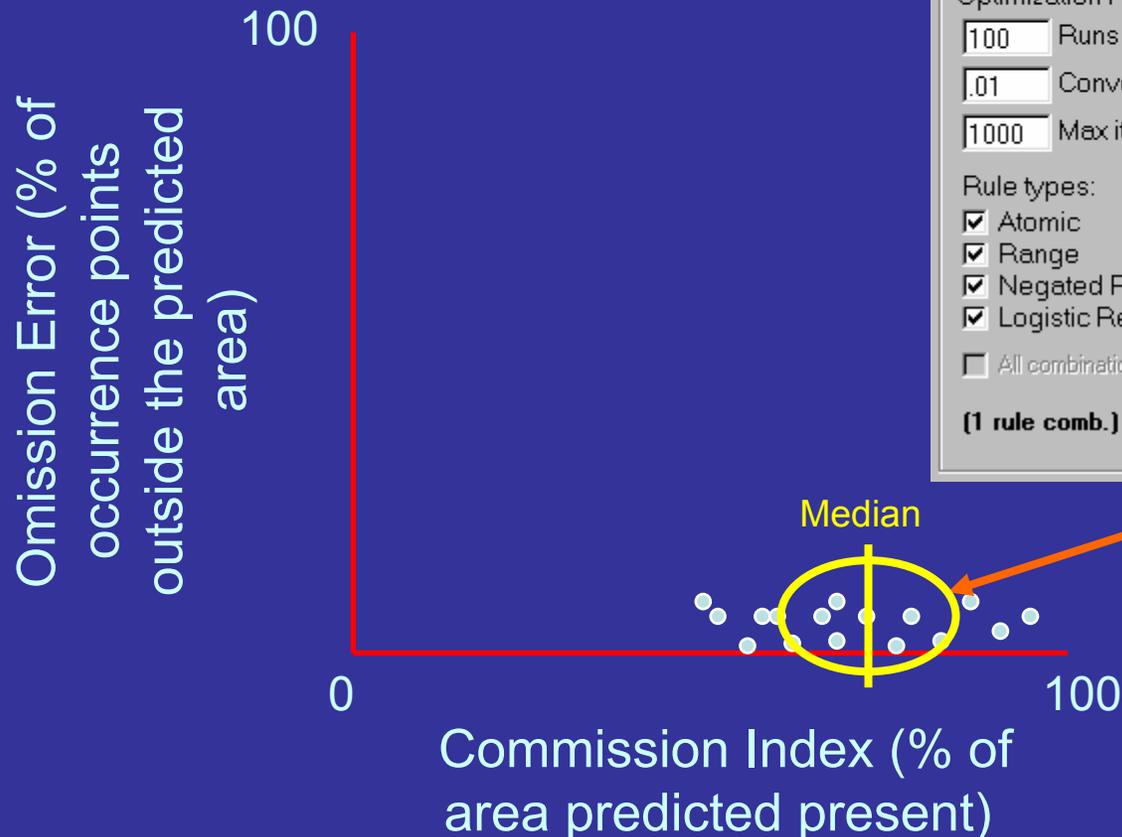


En este caso, el *Total models under hard omission threshold* no aplica

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Implementación en Desktop GARP

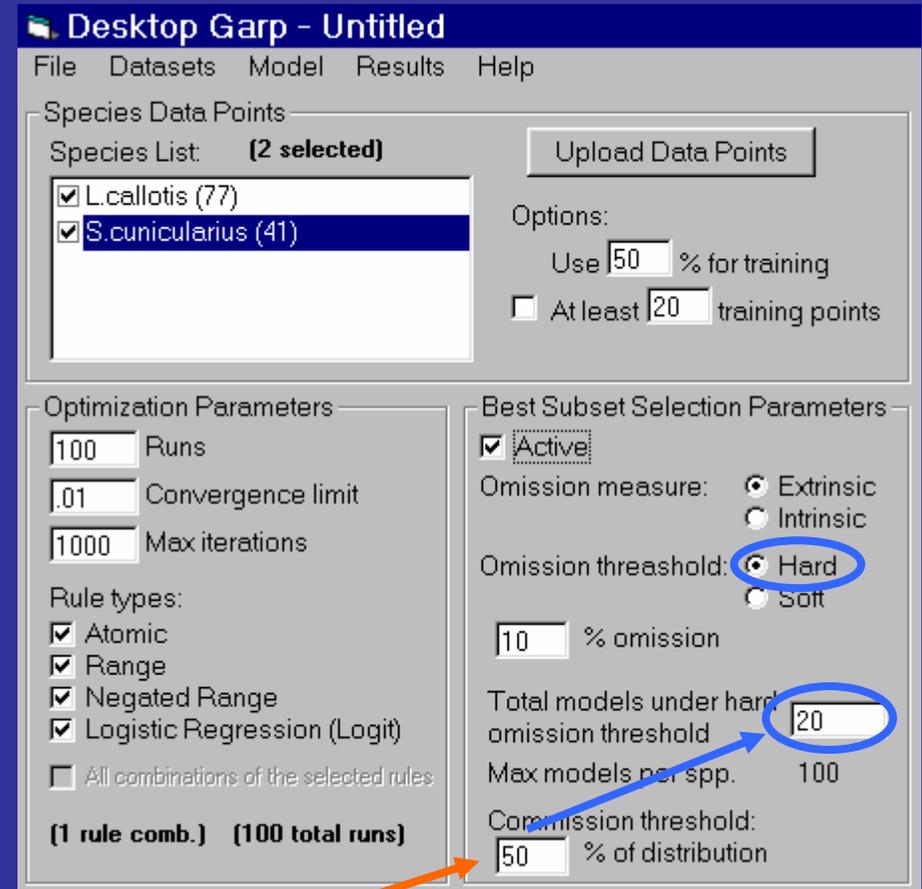
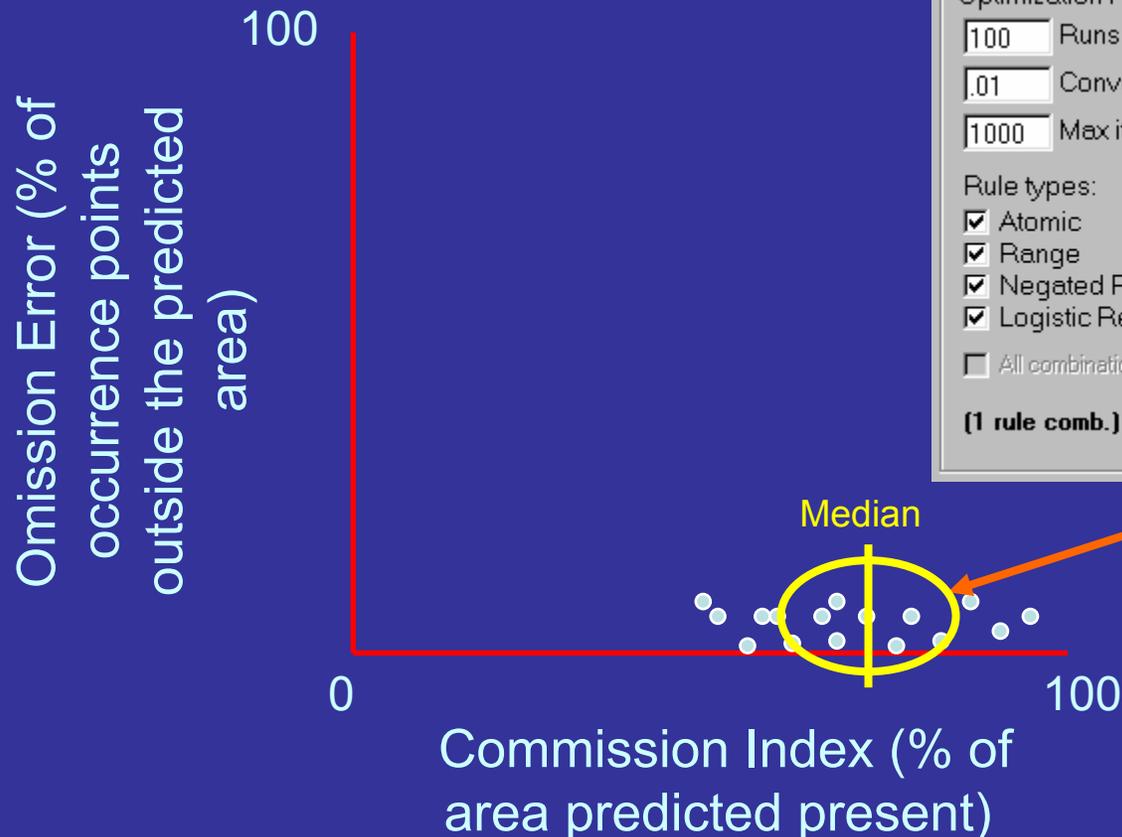
Finalmente, en el *Commission threshold* se indica le numero de modelos (en porcentaje) cercano a la mediana del eje del *Commission Index* que desean seleccionar de los modelos restantes, después del filtro de omisión



Cuando el *Omission threshold* es *Soft*, el valor del *Commission threshold* es relativo al valor del % de *distribution*

10. Funcionamiento, interfaz y formato de los datos (entrada y salida del modelo) en GARP

Implementación en Desktop GARP



Quando el *Omission threshold* esta en *Hard*, el valor del *Commission threshold* es relativo al valor del *Total models under hard omission threshold*

Un par de comentarios finales...

¿Utiliza DesktopGarp datos de ausencia de la especie?

Por el momento no. GARP está optimizado para trabajar sólo con datos de presencia, debido a la naturaleza de las bases de datos de biodiversidad.

Como hay cierta demanda de que se pueda trabajar con datos de ausencia. El manejo de ausencias se está considerando en una versión futura.

¿Entonces cómo trata DesktopGarp los datos de ausencia?

Trata los puntos de ausencia y los no muestreados como conceptos equivalentes. Todos los puntos de no presencia se consideran puntos de ausencia de la especie.

Muchas gracias