

Uso R y GRASS para la obtención de mapas de disponibilidad de hábitat mediante análisis factorial de nicho.

J.A. Palazón Ferrando ⁽¹⁾ y F. Alonso Sarría ⁽²⁾

⁽¹⁾Dpto. Ecología e Hidrología. Facultad de Biología, Universidad de Murcia, Campus Universitario de Espinardo, 3101 Murcia, palazon@um.es.

⁽²⁾ Departamento de Geografía. Facultad de Letras, Universidad de Murcia, Santo Cristo, 1, 30100 Murcia, alonsarp@um.es.

RESUMEN

La obtención de mapas de disponibilidad de hábitat es posible gracias a diversas aplicaciones, entre ellas cabe destacar el programa "Biomapper", actualmente en su versión 3.1.5 (<http://www2.unil.ch/biomapper>) o la librería "adehabitat", desarrollada para utilizar en el programa R (<http://www.r-project.org>), que se encuentra en su versión 1.2-1.

En ambas soluciones el problema para la realización de los análisis radica en la necesidad de llevar la información de la cartografía digital al programa. Esto conduce a diversos problemas relacionados con el manejo de grandes volúmenes de información y de la necesidad de la duplicación de la información con otros formatos.

Para evitar esta situación la solución que se plantea se basa en el uso de dos programas de licencia GPL, es decir, de distribución libre. Uno de ellos es GRASS un sistema de información geográfica (<http://grass.itc.it>) con una gran potencia y capacidad para la modelización de procesos naturales y que será la herramienta SIG. El otro es R un potente programa-lenguaje estadístico. Ambos han sido conectados por una sencilla librería que realiza el análisis a partir de un fichero que contiene información descriptiva sobre el análisis, un fichero con la relación de capas del SIG que han de ser utilizadas y un fichero que contiene las coordenadas donde se ha localizado a la especie. Así, los tratamientos de los datos espaciales son realizados por GRASS y el análisis estadístico por R. El resultado final proporciona un mapa en el SIG. La librería está constituida por una función de R que realiza el análisis con la ayuda de unos script en bash, encargados de extraer y devolver la información de GRASS.

Palabras clave: ENFA, disponibilidad de hábitat, HSM, distribución de especies, GRASS, lenguaje R, GPL, software libre.

ABSTRACT

It is possible to obtain habitat suitability maps using several applications like "Biomapper" v. 3.1.5 (<http://www2.unil.ch/biomapper>) or the "adehabitat" library v. 1.2.1, developed to be used within R program (<http://www.R-project.org>).

Both approaches have the same problem, the need of the geospatial information to be imported to the program. That means several difficulties related with the management of large information volumes and the need of a duplication of the information using several different formats.

A solution is proposed in order to overcome such problems. It is based on the use of two GPL programs (Free and Open Source Systems). One of them is GRASS, a Geographical Information System (<http://grass.itc.it>) with a large capacity to model environmental processes, it is used as a GIS tool. The other is R, a powerful statistic program-language. Both have been linked by a simple library that run the analysis from several files containing, in the first one, descriptive information about the problem, in the second one the GIS layers to be used, and a final file with the coordinates where the species has been located. The library contains a R function for develop the analysis and several bash scripts for extract and return de information to GRASS.

This way, the spatial analysis is performed by GRASS and the statistical modelling by R. A raster map is obtained as a final product.

Key words: ENFA, habitat suitability, HSM, species distribution, GRASS, language R, GPL, free software

INTRODUCCIÓN

La distribución y abundancia de los organismos es un problema central en ecología. El concepto de nicho considera un conjunto de variables ambientales que definen un volumen n -dimensional en el que las especies pueden persistir[1] Los problemas derivados de la integración de la componente espacial en modelos y análisis de los datos ecológicos es una constante de la tarea del ecólogo. Por ello la incorporación de los SIG como herramienta habitual en el trabajo del ecólogo es creciente a pesar de una cierta "distancia" entre ambos, debido posiblemente a la dificultad en el manejo de estos sistemas y la falta de formación de los profesionales de la ecología en tecnologías de la información.

Sin embargo, hay numerosísimas aplicaciones y metodologías que implican necesariamente el uso de información espacial. En mucho casos estas aplicaciones recurren a la importación de datos procedentes de los SIG. En el caso que nos ocupa, los métodos para la obtención de mapas de disponibilidad de hábitat (*habitat map suitability*)[2] pueden utilizarse dos herramientas: **Biomapper**[3] y la librería **adehabitat**[4]. En los dos casos son herramientas disponibles en la red y siguen este principio, la primera usa ficheros procedentes de IDRISI[5], de forma directa, y la segunda de Arc-INFO[6], con una conversión obligatoria.

Desde un punto de vista formal la mayor diferencia está en que el primero es un programa ejecutable para WinXX, código cerrado, y el segundo es una librería de R, abierta. El presente trabajo que se presenta utiliza las ventajas del segundo, apoyándose en él, adecuando su código, y optimiza el uso del SIG. Para conseguirlo se ha utilizado una combinación entre **GRASS**[7] como SIG, **R**[8] como herramienta estadística y **bash** [9] un lenguaje para las tareas de interconexión y programación en GRASS. Así, se lleva a la práctica uno de los principios que movían el desarrollo de GRASS: "*GRASS is a part of a garden*" [10]; y que mueve la principal ventaja del uso del software libre conectando diversas aplicaciones para un fin.

Se ha huido del desarrollo de una aplicación específica ya que R proporciona un entorno de gran potencia, con un gran número de librerías que crece a gran velocidad y que facilita la programación así como la interacción entre usuarios y programadores. En la figura 1 se muestra la curva de crecimiento de número de librerías disponibles en cada nueva versión (y) y el número de versión (x); se trata de un modelo de crecimiento exponencial con el modelo $y=e^{2.06597x+1.95780}$ (con $\alpha<0.001$ y con $R^2_{adj}=0.992$) que predice una duplicación del número de librerías en cada versión. El propio proyecto ADE se inicia como una aplicación cerrada para el sistema operativo de Macintosh; posteriormente se exporta a WinXX y finalmente se porta como librería a R[11]. Inicialmente una librería llamada **ade4**[12], posteriormente se incluye otra más específica **adehabitat**[4] y finalmente se ha añadido una interfaz gráfica **ade4TkGUI**[13]. Esta última incorporación supone una ayuda para el usuario en la elaboración de los análisis y la elección de las opciones más adecuadas. En nuestro caso también se busca la facilidad de trabajo. En este caso como se discutirá más adelante el primer paso es con la creación de ficheros de trabajo de contenido sencillo.

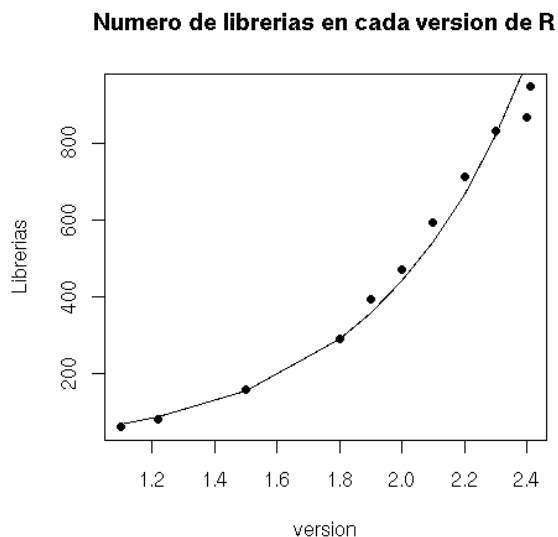


Figura 1: Crecimiento del número de paquetes de R con el número de la versión.

EL ANÁLISIS FACTORIAL DE NICHOS ECOLÓGICO

El proyecto tiene como objetivo el desarrollo de una función que permita el cálculo de un **mapa de disponibilidad de hábitat** (HSM, en adelante) de forma sencilla y sin la duplicación de la información, evitando la carga en memoria de los datos innecesarios para el cálculo. Para obtener los resultados se recurre a un análisis

factorial del nicho ecológico (ENFA: ecological niche factor analysis), desarrollado por Hirzel[2]. En esencia la obtención de un HSM se basa en la información correspondiente a los datos de presencia de una especie en el territorio y la información espacial en área estudiada de las variables que pueden conformar el hábitat de la especie. El método selecciona por un procedimiento de análisis factorial la información relevante del hábitat mediante la obtención de la **marginalidad** y la **especificidad**. La primera describe como se aleja el comportamiento promedio de las celdillas ocupadas por la especie del comportamiento del área de estudio. La segunda mide cuan estrecho es el rango de la variable para la especie. A partir de estos valores se construye el HSM.

El principal problema de cálculo radica en el cálculo de la matriz de correlaciones entre los mapas utilizados. Es preciso calcular $n*(n+1)/2-n$ coeficientes para obtener la matriz que se ha de diagonalizar. Además la carga en memoria de la información de los mapas hace inalcanzables ciertos objetivos (el detalle en los mapas, aumento geométrico de la necesidades de memoria con el aumento de la resolución, y el mapas, que aumentan exponencialmente el tiempo de cálculo).

En resumen, si el procedimiento está implementado en la función, sólo necesitamos indicar donde están los datos de presencia de la especie y que mapas del SIG ha de utilizarse en el análisis. Los resultados son básicamente la tabla de marginalidad/especificidad de las variables y los valores propios de la matriz diagonalizada.

LA FUNCIÓN ENFA.GRASS.R

El trabajo ha proporcionado una función de R que aprovecha la capacidad de conectar con programas externos. Mediante la función **system()** se solicita a GRASS que realice las tareas precisas destinadas a obtener los resultados intermedios que R utilizará para los distintos cálculos. En realidad a la función de ENFA.GRASS.R no es el único programa, se acompaña de varios **scripts** escritos con lenguaje bash, destinados a la automatización de diversas tareas en GRASS. Así, por ejemplo, para obtener los datos de la región actual definida en GRASS se utiliza la expresión:

```
as.matrix(as.numeric(
  system("sh r.describe.sh",intern=T)))->describeregion
```

que crea la matrix describeregión con la ayuda del script **r.describe.sh** proporcionando a R los datos deseados. Otro script, **r.correlacion.sh** que se ha escrito ha de obtener la matriz de correlaciones de los mapas. La necesidad de este se fundamenta en la limitación que GRASS tiene para el número de mapas a correlacionar.

Breve manual

El uso de ENFA.GRASS.R es realmente sencillo, no es necesario ser un usuario de GRASS ni de R para poder realizar un análisis. Bastará el acceso un servidor donde se tenga una base de datos en GRASS y los programas pertinentes y seguir un sencillo protocolo. Se describe a continuación los pasos desde la instalación hasta la obtención de los resultados.

Instalación

La función, los scripts, algunas bases de datos en GRASS y ficheros de trabajo de ejemplo pueden descargarse desde <http://fobos.bio.um.es/palazon/ENFA.GRASS.R.tgz> un fichero tar comprimido. Contiene tres subdirectorios: 1) **gis**, con el DATABASE de GRASS para los ejemplos; 2) **ejemplos**, con los directorio de las especies de ejemplo); y 3) **programa**, con la función y los scripts necesarios para realizar el análisis

Datos para el análisis

Tal como se refleja en los ejemplos, resulta conveniente crear un directorio por cada análisis previsto, en él se situarán los ficheros de datos.

Se necesitan dos fichero de datos:

- El primero contiene las coordenadas en las que se ha detectado la presencia del organismo.
- El segundo contiene la relación nombres de lo mapas que han de ser utilizados en el análisis y que están disponibles en el MAPSET de GRASS.

Procedimiento

Para utilizar la función es necesario seguir el siguiente esquema, que se ilustra en la figura 2:

- Iniciar un terminal con bash (trabando en un entorno GNU/Linux suele ser la opción por defecto en las distintas distribuciones) Fig. 2: a.
- Iniciar en la terminal anterior el programa GRASS (versión 6.x) Fig. 2: a y b.
- Iniciar desde GRASS el programa R (versión 2.x) Fig. 2: c.
- Cargar la función mediante la función source("/path/ENFA.GRASS.R") desde su ubicación (que abreviamos como /path/) Fig. 2.d.
- Realizar el análisis invocando a la función Fig. 2.d.

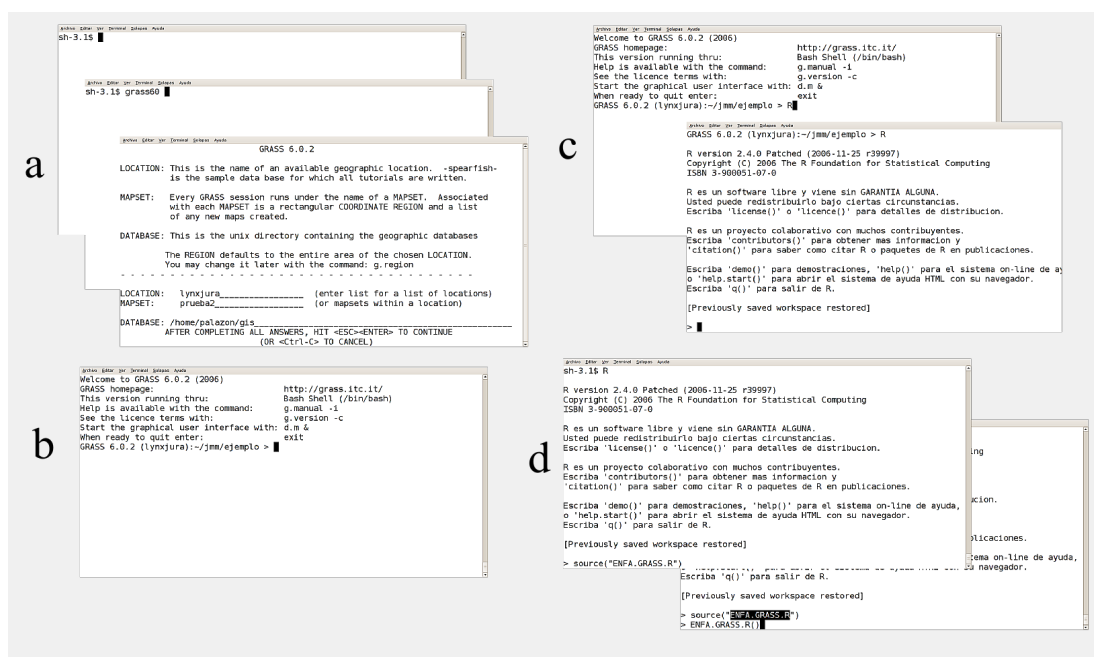


Figura 2: Esquema del proceso de la realización de un ENFA mediante la función ENFA.GRASS.R (aclaraciones en el texto).

Resultados

Los resultados tienen dos componentes, por un lado la descripción de la marginalidad y la especificidad que se proporcionan en un fichero de texto en el directorio de trabajo. Por otro, el HSM de la especie que se ha incorporado en la base de datos de GRASS, en el mapset activo.

En la figura 3, se presentan los resultados de un ENFA realizado con los datos de lince de un proyecto de seguimiento en Francia[14], datos disponibles en el paquete adehabitat, y que se incorporan en el fichero de instalación. Para comparar los resultados se presenta los obtenidos con la propia librería y mediante ENFA.GRASS.R.

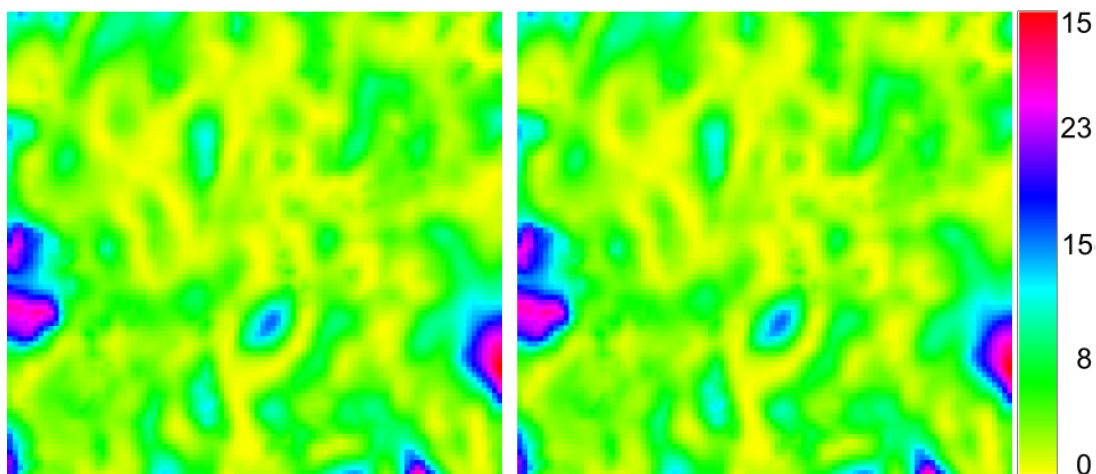


Figura 3: HSM de los datos de lince obtenido: con la función ENFA.GRASS.R (izda) y con las funciones de la librería adehabitat (dcha). La escala gráfica indica la disponibilidad de hábitat.

Los resultados varían muy ligeramente, debido a los redondeos que GRASS realiza al calcular la matriz de correlaciones. Sin embargo, esto no parece afectar, tal como muestra la figura 3, en ninguna de las pruebas realizadas. En este caso la diferencia media entre los dos mapas es de 1.48×10^{-5} , con una desviación típica de 3.94×10^{-4} , con un rango de 4.22327×10^{-3} . Mientras los valores de mapa varían entre 0.024 y 30.455.

Estado actual del proyecto y tarea pendiente

La versión actual de la función y los script son plenamente funcionales. Se trata de la versión 0.2. Todavía está en una fase muy inicial y la documentación es escasa, tanto del manejo de la función, como en las fuentes de los programas.

Dado que un aspecto importante es la documentación la tarea que sigue, conseguido el pleno funcionamiento de la función, es desarrollar una documentación que permita en adelante presentar a ENFA.GRASS.R como una librería de R.

Desafortunadamente queda un problema que ha de solucionarse para tener una librería plenamente operativa. Determinar una ubicación de los scripts y que estos se copien en ella de forma automática en la instalación de la librería.

Otra tarea es conectar la función con **wxgrass**[16] de manera que aún pueda ser más sencillo su uso.

Como un resultado adicional interesante se ha desarrollado un lenguaje muy simplificado, por el momento, que permite a los usuarios de R explotar las capacidades de GRASS de una forma sencilla, inicialmente con objetivo docente[16].

CONCLUSIONES

Tal como se esperaba la integración de los dos sistemas ha proporcionado el resultado deseado, es posible aumentar el número de mapas y la resolución sin más problema que el derivado del tiempo de ejecución.

Se ha conseguido un buen resultado que ha permitido plantear el uso de GRASS desde funciones de R, que permite un buen nivel de tratamiento estadístico y la completa funcionalidad de GRASS.

El uso de un protocolo basado en sencillos ficheros de información permite una automatización del análisis que tiene gran interés para analizar conjuntos de especies, o modificaciones en la lista, resolución, u otras opciones que el investigador considere oportunas.

AGRADECIMIENTOS

Debe agradecerse ante todo a la comunidad de desarrolladores de software libre, en particular a los de GRASS y R, el esfuerzo, el trabajo y el tesón que permiten que día a día la comunidad de *GRASSeros* y *R-ros* crezca en cantidad y calidad. Además, el agradecimiento debe extenderse a sus familias que en muchas ocasiones sufren la ausencia de sus seres queridos.

REFERENCIAS

1. HUTCHINSON, G.E. (1957) Concluding Remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, 22:415-427.
2. HIRZEL, A. H.; HAUSSER, J.; CHESSEL, D. y PERRIN, N. (2002), Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036.
3. Hirzel, A. (2006), Biomapper A GIS-toolkit to model ecological niche and habitat suitability. (<http://www2.unil.ch/biomapper/index.html>)
4. CALENGE, C. (2007), The adehabitat Package. Title Analysis of habitat selection by animals. January 17, 2007. Version 1.5-2. (<http://cran.univ-lyon1.fr/doc/packages/adehabitat.pdf>)
5. Clark Labs (2007), Página principal del proyecto IDRISI. Worcester, Universidad de Clark. (<http://www.clarklabs.org>)
6. ESRI (2007) . Página principal de la empresa. Redlands. (<http://www.esri.com/>)
7. NETELER, M. y MITASOVA, H. (2002), Open Source GIS. A GRASS GIS Boston/Dodrecht/London, Approach Kluwer. 434 pags.

8. R DEVELOPMENT CORE TEAM (2006), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. (<http://www.R-project.org>)
9. WALLINGTON, J. P. (2002), *GNU Bash Reference Manual, for Bash, Version 2.05b*. (<http://www.gnu.org/software/bash/manual/bashref.html>)
10. WESTERVELT, J. (1991), *Introduction to GRASS 4*, GRASS Information Center, U.S. Army. CERL,. Champaign, Illinois, U.S. July 1991.
11. CHESSEL, D.; DUFOUR, A.-B. y THIOULOUSE, J. (2004), The ade4 package-I- One-table methods. *R News* 4:5-10.
12. THIOULOUSE, J.; CHESSEL, D.; DOLÉDEC, S. y OIIVIER, J.M. (1996), ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* 7:75-83.
13. THIOULOUSE, J. Y DRAY, S. (2006), *The ade4TkGUI Package. Title ade4 Tcl/Tk Graphical User Interface*. September 29, 2006. Version 0.1-4 (<http://pbil.univ-lyon1.fr/ADE-4/>)
14. VANDEL, J.M. (2001), *Répartition du Lynx (Lynx lynx) en France (Massif Alpin, Jurassien et Vosgien). Méthodologie d'étude et statut actuel*. École Pratique des Haute Etudes de Montpellier II: Dissertation.
15. ALONSO SARRÍA, F. (2006), *wxGRASS. Una interfaz gráfica de usuario para GRASS desarrollada con python y wxPython*. XII Congreso Nacional de Tecnologías de la Información Geográfica.
16. PALAZÓN FERRANDO, J.A. y CALVO SENDÍN, J.F. (2007), *Ecología Metodológica y Cuantitativa. Sesión práctica 11. Modelos ecológicos con sistemas de información geográfica*. Universidad de Murcia (<http://fobos.bio.um.es/palazon/emc/practicassesion11/index/>)