

ÚLTIMOS AVANCES EN LA LIBRERÍA R {synoptReg}: UNA CAJA NEGRA DE FÁCIL APLICACIÓN PARA COMPUTAR CLASIFICACIONES SINÓPTICAS

Marc LEMUS-CÁNOVAS^{1,2}

¹*Climatology Group. Departament de Geografia. Universitat de Barcelona.
2Andorra Recerca + Innovació. Av. Rocafort, 21-23, AD600 Sant Julià de Lòria.
Andorra.*

mlemus@ub.edu

RESUMEN

Una librería R es una colección de funciones que tienen un mismo objetivo común. En el caso que nos ocupa, la librería {synoptReg} tiene por objetivo: 1) permitir un acceso fluido y sencillo a datos atmosféricos de reanálisis; 2) el cómputo de una clasificación sinóptica objetiva; 3) cartografiar una variable climática o ambiental a partir de los patrones sinópticos clasificados en el punto anterior; 4) finalmente, obtener una regionalización categórica de tipo climático/ambiental mediante los patrones espaciales de la variable climática/ambiental cartografiada.

Una de las virtudes que tiene el lenguaje R es que se trata de software libre, lo que facilita su acceso y uso a gran escala, tal y como demuestra el aumento de librerías desarrolladas en este lenguaje durante los últimos años. Lógicamente, la librería synoptReg no ha estado exenta de actualizaciones desde sus inicios a finales del año 2018 hasta la actualidad. Uno de los cambios más significativos que ha sufrido la librería ha sido el incremento de los métodos disponibles de clasificación sinóptica. Actualmente, es posible aplicar el análisis de componentes principales en una matriz en modo T o S, así como aplicar otras clasificaciones tales como el método automático de Lamb -conocido también como Jenkinson y Collison-. Por último, el método más recientemente incorporado es el de los mapas auto-organizados (en inglés, self-organizing maps; SOM), el cual está basado en el aprendizaje automático mediante redes neuronales.

Palabras clave: clasificación sinóptica, tipos de circulación, R, synoptReg.

ABSTRACT

An R package is a collection of functions that have the same common objective. In this case, {synoptReg} aims to: 1) allow fluid and simple access to atmospheric reanalysis data; 2) compute an objective synoptic classification; 3) map a climatic or environmental variable from the synoptic patterns classified in the previous point; 4) finally, obtain a categorical regionalisation of climatic/environmental type by means of the spatial patterns of the mapped climatic/environmental variable.

One of the strengths of the R language is that it is free software, which facilitates its access and use, as demonstrated by the increase in the number of libraries developed in this language in recent years. Logically, the synoptReg library has not been exempt from updates from its beginnings at the end of 2018 to the present day. One of the

most significant changes has been the increase in the available synoptic classification methods. It is now possible to apply principal component analysis on a matrix in T or S mode, as well as to apply other classifications such as Lamb's automatic method - also known as Jenkinson and Collison-. Finally, the most recently introduced method is the self-organising maps (SOM), which is based on machine learning using neural networks.

Key words: synoptic classification, circulation types, R, synoptReg

1. INTRODUCCIÓN

Una de las virtudes que tiene el lenguaje R es que se trata de software libre, lo que facilita su acceso y uso a gran escala, tal y como demuestra el aumento de librerías desarrolladas en este lenguaje durante los últimos años. La librería `synoptReg` (Lemus-Canovas et al., 2019) tiene por objetivo: 1) permitir un acceso fluido y sencillo a datos atmosféricos de reanálisis; 2) el cómputo de una clasificación sinóptica objetiva; 3) la cartografía de una variable climática o ambiental a partir de los patrones sinópticos clasificados en el punto anterior; y finalmente, 4) obtener una regionalización categórica de tipo climático/ambiental mediante las cartografías obtenidas en el punto 3. Lógicamente, la librería `synoptReg` no ha estado exenta de actualizaciones desde la aparición de la primera versión a finales del año 2018 en el *Comprehensive R Archive Network* (CRAN).

Por lo general, los cambios introducidos han ido en la línea de facilitar su uso con el fin de que el usuario no tenga que preocuparse por la adquisición previa de los datos o la construcción del método de clasificación, entre otros. De esta forma, la librería `synoptReg` complementa las tareas realizadas en la acción europea COST733, que culminó con el desarrollo de la librería `cost733cat` (Philipp et al., 2010), en la cual se programaron multitud de métodos para el cómputo de clasificaciones sinópticas usando programación en Bash (Unix).

En esta comunicación se presentarán las últimas novedades de la librería, haciendo hincapié en los métodos de clasificación sinóptica disponibles hasta la fecha.

2. MÉTODOS IMPLEMENTADOS EN LA LIBRERÍA

2.1. Estructura de la librería

Uno de los cambios más significativos que ha sufrido la librería ha sido el incremento de los métodos disponibles para el cálculo de clasificaciones sinópticas. Inicialmente, tan solo existía la posibilidad de realizar una clasificación mediante el análisis de componentes principales (ACP) sobre una matriz modo S, con la posterior conformación de los tipos de circulación mediante el método de clustering *k-means* siguiendo el procedimiento propuesto por Esteban et al. (2005). Actualmente, también es posible realizar el ACP para una matriz en modo T –la transpuesta de la S–, que no precisa de algoritmos de clustering para la conformación de los tipos de circulación. También se ha habilitado la opción de realizar una clasificación sinóptica mediante el método automático de Lamb –también conocido como Jenkinson y Collison–. Por último, el método más recientemente implementado es el de los mapas

autoorganizados (en inglés, self-organizing maps; SOM), el cual está basado en el aprendizaje automático mediante redes neuronales.

Desde la propia librería es posible realizar la descarga de datos del reanálisis atmosférico NCEP/NCAR (Kalnay et al., 1996) gracias a la función `download_ncep()`. Para poder ejecutar las clasificaciones sinópticas mediante el ACP y los SOM, es necesario pasar previamente la función `tidy_nc()`. Esta función formatea los datos para poderlos ejecutar en las funciones de clasificación anteriormente mencionadas y, además, permite filtrar temporal y espacialmente nuestra base de datos atmosférica. Una vez realizado el preproceso de los datos de entrada –a excepción de si se quiere ejecutar la clasificación automática de Lamb (`lamb_clas`), para la cual no es necesario tal preproceso–, se puede computar cualquier función relativa a los métodos de clasificación. La ejecución de estas funciones –`synoptclas()` para ACP, `lamb_clas()` para Lamb y `som_clas()` para SOM– permite asociar cada día a un tipo de circulación determinado, así como los valores promedios de nuestra variable atmosférica por tipo de circulación y coordenadas, permitiendo así generar los mapas sinópticos de la clasificación.

A parte del cómputo de clasificaciones sinópticas, la librería está pensada para realizar el enlace de la circulación atmosférica con la superficie, pudiendo así cartografiar los valores de la variable climática o ambiental (precipitación, temperatura, NO₂, O₃, etc.) de interés en relación con cada patrón sinóptico con la función `ct2env()`. Por último, cabe la posibilidad de realizar una regionalización categórica de esta variable climática o ambiental para extraer sus principales patrones espaciales mediante la función `regionalization()`.

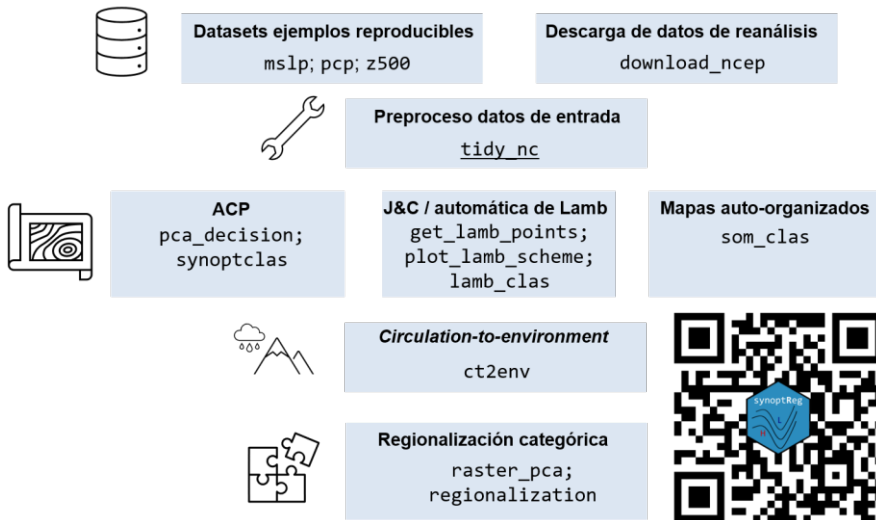


Fig. 1: Estructura de la librería synoptReg

2.2. Automática de Lamb o Jenkinson and Collison

La clasificación automática de Lamb o de Jenkinson and Collison es una clasificación sinóptica objetiva o automática, cuyos tipos se obtienen siguiendo un algoritmo

compuesto por unas reglas definidas, consultables en Trigo y DaCamara (2000). De este modo no interviene la subjetividad del operador, siendo así sus resultados reproducibles y transferibles. Esta clasificación contiene 27 tipos obtenidos a partir de la presión atmosférica en superficie en un retículo de 16 o más puntos que englobe el área objeto de estudio (Fig. 2). En la librería synoptReg se dispone de la función `get_lamb_points()` para obtener los 16 puntos de malla a partir de un centroide dado, mientras que la función `plot_lamb_scheme()` cartografía estos 16 puntos en un mapa. Finalmente, la función `lamb_clas()` es la que se encarga de realizar el cálculo de la clasificación a partir de estos 16 puntos.

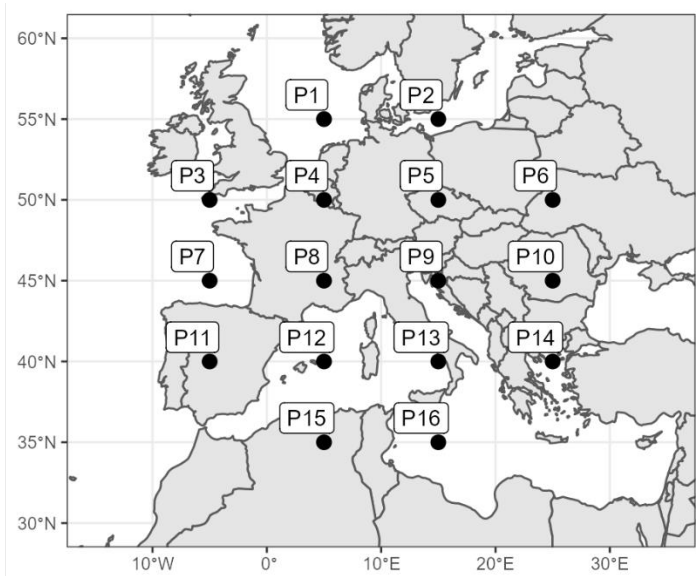


Fig. 2: Output de la ejecución de la función `plot_lamb_scheme()`

2.3. Análisis de Componentes Principales

2.3.1. Modo S

La clasificación en modo S es aquella en la que los puntos de la rejilla (lon-lat) son las variables y las observaciones (filas), los días. Así, la aplicación del ACP sobre este tipo de matriz establece las relaciones lineales entre las series temporales de todos los puntos de rejilla. En este sentido, las puntuaciones obtenidas muestran el grado de representatividad de cada día para cada uno de los componentes principales (CP). Sin embargo, no hay que confundir que estos patrones espaciales en ningún caso son tipos de circulación atmosférica, ya que la variabilidad captada por este proceso es baja (Hewitson y Crane, 1992; Huth, 1996), y hace referencia a los patrones de teleconexión. Por esta razón, necesitamos utilizar tales puntuaciones como coordenadas en el espacio multivariante y computar un método de clustering que acabe creando grupos entre aquellos días más parecidos entre sí. Además, se aplica la rotación VARIMAX a los CP retenidos para redistribuir su varianza explicada. Es probable que la varianza de los CP después de la rotación no siga un orden

consecutivo. Por ejemplo, el CP1 no tiene por qué ser el que acumule la mayor varianza explicada.

Antes de la rotación, primero hay que seleccionar el número de CPs que se usarán en la clasificación. Para ello se emplea la función `pca_decision()` que devuelve los valores de varianza explicada por cada CP así como un scree plot para facilitar la selección del número óptimo de componentes a retener. La rotación de los componentes se aplicará posteriormente cuando se ejecute la función `synoptclas()` que dará lugar a la clasificación final.

Como se comentaba anteriormente, para la consecución de la clasificación sinóptica mediante un ACP en modo S es necesario recurrir a un método de clustering. En el paquete `synoptReg` el método de agrupación implementado es el propuesto por Esteban et al. (2005), basado en las puntuaciones extremas. Esteban et al. (2005) sugieren que las puntuaciones extremas (positivas y negativas) de cada CP (>2 o <-2) pueden ser utilizadas como las coordenadas multivariantes de cada centro de clúster para luego aplicar el algoritmo *k-means*. En este sentido, al utilizar las puntuaciones positivas y negativas, los grupos de cada componente se duplican –siempre y cuando para cada grupo se alcancen estas puntuaciones extremas–. Así, si retenemos 4 CPs, nuestra clasificación final puede alcanzar hasta 8 tipos de circulación atmosférica.

2.3.2. Modo T

La estructura de la matriz en modo T establece que las variables son los días y las observaciones son los puntos de rejilla. En este caso, el ACP establece las relaciones ortogonales entre cada mapa diario de la variable atmosférica. En este caso, el ACP asocia directamente los días más parecidos entre sí para agruparlos en tipos de circulación, por lo que para este modo no es necesario utilizar ningún método de clustering. `synoptReg` permite escoger entre tres variantes (tabla 1) dentro del modo T basadas en las:

- **Correlaciones positivas y negativas (PN)** de cada uno de los días frente a cada CP. El día 1 puede asociarse a la CP X+ o a la CP X-, dependiendo de si la correlación es positiva o negativa. Esta última opción, permite obtener el doble de clases de los componentes seleccionados. Si se seleccionan 4 CPs, el resultado final puede alcanzar hasta un máximo de 8 tipos de circulación.
- **Correlaciones positivas máximas (MAX)** de cada día frente a cada CP. El día 1 se asociará a la CP con mayor correlación positiva;
- **Correlaciones absolutas (ABS)** de cada día contra cada CP. El día 1 se asociará al CP con mayor correlación absoluta.

Fecha	CP1	CP2	CP3	PN	MAX	ABS
2000-01-01	0.7	0.3	-0.4	1+	1	1
2000-01-02	0.5	0.4	-0.6	3-	1	3
2000-01-03	0.2	0.5	-0.7	3-	2	3
2000-01-04	-0.3	0.5	-0.4	2+	2	2
2000-01-05	-0.6	0.4	-0.1	1-	2	1
...

Tabla 1: Métodos de asignación de los tipos sinópticos cuando el análisis de componentes principales (ACP) se aplica a una matriz en modo T. Los valores

obtenidos son valores de correlación de cada día con cada componente principal (CP)

2.4. Mapas auto-organizados (en desarrollo)

La técnica SOM fue desarrollada por Kohonen (2001) y fue utilizada en varias aplicaciones, incluyendo el estudio de eventos de precipitación extrema (Gibson et al. 2017; Pinto et al. 2018). Una revisión de las aplicaciones de SOM en climatología sinóptica se puede encontrar en Hewitson y Crane (2002). Los SOM emplean un algoritmo de red neuronal para establecer patrones espaciales generalizados -nodos- que abarcan el rango de condiciones presentes en los datos de entrenamiento. Cada nodo se caracteriza por un vector de referencia, que se actualiza cada vez que se presentan nuevos datos (campos diarios) a la red neuronal. Los nodos se organizan en una matriz bidimensional de forma que los nodos similares se sitúan cerca unos de otros, lo que facilita la interpretación de los TC representativos y los procesos físicos subyacentes (Hewitson y Crane, 2002).

La librería synoptReg dispone de la función som_clas() que permite definir el número de dimensiones e iteraciones para ejecutar esta metodología.

3. CASO PRÁCTICO: CLASIFICACIÓN SINÓPTICA DE LOS EVENTOS TORRENCIALES EN EL OBSERVATORIO DEL EBRO-TORTOSA (1948-2022).

Como demostración de las posibilidades que ofrece synoptReg, a continuación, se enlaza un pequeño tutorial para clasificar sinópticamente los eventos de precipitación abundante en el Observatori del Ebre-Tortosa (Tarragona) durante el periodo 1948-2022: https://lemuscanovas.github.io/es/post/20220412/tutorial_acc2022/

En este tutorial se detalla paso a paso como ejecutar las distintas funciones que existen actualmente en la librería y que fácilmente se puede usar para cualquier otra estación o región. Uno de los principales resultados que se pueden obtener de la ejecución de las funciones de la librería, los mapas sinópticos, se presentan en la figura 3.

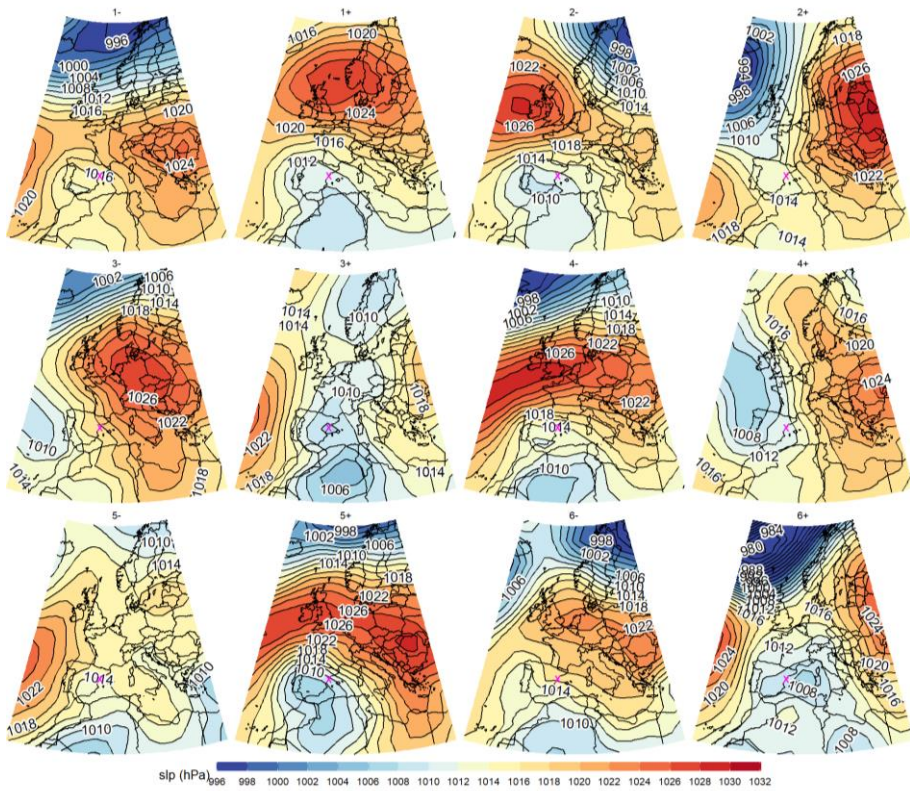


Fig. 3: Mapas de presión en superficie asociados a la clasificación sinóptica de eventos de precipitación abundante en el Observatori de l'Ebre- Tortosa (1948-2022).

4. CONCLUSIONES Y TRABAJO FUTURO

La librería synoptReg contiene un conjunto de funciones que permiten el cómputo de una clasificación sinóptica para cualquier región del planeta. Como se ha descrito en esta comunicación, hasta la fecha, esta librería contiene tres metodologías de clasificación distintas con la posibilidad de descargar los datos necesarios para el cálculo de la clasificación. Sin embargo, se plantean futuros desarrollos que permitan la opción de descargar datos del reanálisis ERA-5, así como la creación de una aplicación web que facilite a cualquier usuario sin conocimientos de programación en R, poder ejecutar cualquier tipo de clasificación disponible en la librería.

AGRADECIMIENTOS

La presente investigación se ha realizado en el marco del Grupo de Climatología de la Universidad de Barcelona (2017 SGR 1362, Generalitat de Catalunya), y del proyecto EXE (PID2020-116860RB-C21).

REFERENCIAS

- Esteban, P., Jones, P. D., Martín-Vide, J., & Mases, M. (2005). Atmospheric circulation patterns related to heavy snowfall days in Andorra, Pyrenees. *International Journal of Climatology*, 25(3), 319-329. Doi: <https://doi.org/10.1002/joc.1103>
- Gibson, P. B., Perkins-Kirkpatrick, S. E., Uotila, P., Pepler, A. S., & Alexander, L. V. (2017). On the use of self-organizing maps for studying climate extremes. *Journal of Geophysical Research: Atmospheres*, 122(7), 3891-3903. Doi: <https://doi.org/10.1002/2016JD026256>
- Hewitson, B. C., & Crane, R. G. (1992). Large-scale atmospheric controls on local precipitation in tropical Mexico. *Geophysical research letters*, 19(18), 1835-1838.
- Huth, R. (1996). An intercomparison of computer-assisted circulation classification methods. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 16(8), 893-922.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... & Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3), 437-472.
- Kohonen, T. (2001). Learning vector quantization. In *Self-organizing maps* (pp. 245-261). Springer, Berlin, Heidelberg. Doi: https://doi.org/10.1007/978-3-642-56927-2_6
- Lemus-Canovas, M., Lopez-Bustins, J. A., Martín-Vide, J., & Royé, D. (2019). synoptReg: An R package for computing a synoptic climate classification and a spatial regionalization of environmental data. *Environmental modelling & software*, 118, 114-119. Doi: <https://doi.org/10.1016/j.envsoft.2019.04.006>
- Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., ... & Tymvios, F. S. (2010). Cost733cat—A database of weather and circulation type classifications. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(9-12), 360-373. Doi: <https://doi.org/10.1016/j.pce.2009.12.010>
- Pinto, I., Jack, C., & Hewitson, B. (2018). Process-based model evaluation and projections over southern Africa from coordinated regional climate downscaling experiment and coupled model intercomparison project phase 5 models. *International Journal of Climatology*, 38(11), 4251-4261. Doi: <https://doi.org/10.1002/joc.5666>
- Trigo, R. M., & DaCamara, C. C. (2000). Circulation weather types and their influence on the precipitation regime in Portugal. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20(13), 1559-1581.