

## REGIONALIZACIÓN DE LOS DÍAS SECOS EN ARGENTINA. UN ENFOQUE METODOLÓGICO.

María L. Bettolli<sup>1,2</sup>, Juan A. Rivera<sup>1,2</sup>, Olga C. Penalba<sup>1</sup>

<sup>1</sup> Departamento de Ciencias de la Atmósfera y los Océanos, FCEN-UBA.

<sup>2</sup> Consejo Nacional de Investigaciones Científicas y Técnicas.  
Buenos Aires, Argentina

(Manuscrito recibido el 9 de junio de 2009, en su versión final el 11 de noviembre de 2010)

### RESUMEN

La regionalización de diferentes variables climáticas ha sido llevada a cabo en diversas zonas del mundo, dado que para distintos propósitos es conveniente dividir espacialmente la climatología de una variable en un número de áreas cuasi-homogéneas. El objetivo de este trabajo es la obtención de una regionalización objetiva de las distintas variabilidades temporales de las cantidades de días secos en la República Argentina durante el trimestre de verano. Con el fin de lograr regiones con similar variabilidad temporal en las cantidades de días secos se exploraron dos métodos: Análisis de Componentes Principales (ACP) y el algoritmo de agrupamiento no-jerárquico de k-means. En base a una evaluación de los patrones espaciales, la regionalización obtenida mediante el método de k-means aplicado sobre las componentes principales rotadas más importantes, es más apropiada en comparación con la determinada por el método de ACP rotadas. Mediante este método, el territorio nacional presenta seis regiones principales: las regiones Centro-Este; Noreste; Centro y Sur Bonaerense; Noroeste; Centro-Oeste y Patagónica, las cuales son climáticamente coherentes.

*Palabras clave:* regionalización; días secos; sequías; análisis de componentes principales; k-means.

### REGIONALIZATION OF DRY DAYS IN ARGENTINA. A METHODOLOGICAL APPROACH.

### ABSTRACT

Objective regionalizations of different climatic variables have been performed in many regions of the World. For different purposes is convenient to make a spatial regionalization to find quasi-homogeneous climatic regions. The main objective of this work is to identify spatially homogeneous regions of dry days in Argentina with different temporal variabilities during summer season. In order to achieve regions with similar temporal variability in the amount of dry days, two methods are explored: principal component analysis (PCA) and k-means nonhierarchical cluster method. By means of a spatial patterns examination, the regionalization derived by k-means on the most

important rotated principal components extracted is more adequate in comparison with the proposed by PCA method. Through this methodology, the country is divided in six main regions: Northeast region; Central-East region; Central and South Buenos Aires region; Northwest region; Central-West region; and Patagonia region, which are climatically consistent.

*Key words:* regionalization; dry days; droughts; principal components analysis; k-means.

## 1. INTRODUCCIÓN

La regionalización objetiva de diferentes variables climáticas (típicamente temperatura y precipitación) ha sido llevada a cabo en diversas regiones del mundo. Esto se debe a que para distintos propósitos resulta conveniente dividir espacialmente la climatología de una variable en un número de áreas cuasi-homogéneas respecto a su variabilidad temporal. El valor de una regionalización espacial, asociada a variables vinculadas con la precipitación radica en su posible aplicación en cuestiones relacionadas con el manejo de recursos hídricos (Lana y otros 2001). Habitualmente, el comportamiento espacial de las sequías resulta muy complejo, siendo común diferencias territoriales significativas en las condiciones de sequedad, incluso entre espacios próximos (Cuadrat Prats y Vicente-Serrano 2004). Diversos estudios a nivel mundial utilizan métodos de regionalización con el fin de obtener zonas cuyas condiciones climáticas presenten eventos de sequía semejantes, obtenidas mediante algún índice adecuado. Los trabajos de Soulé (1990) y Balling y Goodrich (2007) encuentran regiones homogéneas basadas en el Índice de Severidad de Sequía de Palmer (ISSP) para los Estados Unidos mediante la aplicación del análisis de componentes principales (ACP). Los trabajos de Lana y otros (2001); Vicente-Serrano y otros (2004) y Raziei y otros (2008) aplican ACP a series de excesos y déficit de precipitación, monitoreadas mediante el Índice de Precipitación Estandarizado (IPE) en España e Irán, obteniendo regiones homogéneas respecto a las condiciones de sequía.

En vista de las pérdidas ocasionadas por los períodos de sequías prolongadas resulta crucial el desarrollo de planes de seguimiento, prevención y alerta temprana de las sequías para la mitigación del riesgo. Estos planes requieren la identificación de áreas homogéneas respecto al comportamiento temporal de alguna variable representativa de

condiciones secas para optimizar sus propósitos y permitir un uso más eficaz de los recursos. Una de estas variables puede ser, por ejemplo, la cantidad de días sin precipitación (días secos). Esta variable puede tomarse como una primera aproximación para el estudio de condiciones de sequía, aunque mediante la misma no se represente su intensidad. Sin embargo, no existe una definición universal de sequía, excepto el concepto general de la falta de precipitación (Soulé 1990).

Entre los trabajos que utilizan alguna herramienta de regionalización en la región de estudio pueden mencionarse a Compagnucci y Vargas (1985), Coronato y Bisigato (1998), Compagnucci y Araneo (2005) y Barreira y Compagnucci (2007) quienes utilizan el método de ACP aplicado sobre las variables presión, temperatura, caudales y anomalías de hielo marino respectivamente. Sin embargo, no existen trabajos científicos en la región de estudio que busquen áreas homogéneas utilizando como variable de estudio algún índice relacionado con períodos secos.

A partir de lo expuesto, este trabajo plantea realizar una regionalización tomando como variable de estudio la cantidad de días secos, centrandó el análisis en el trimestre de verano. La elección de este trimestre en particular radica en las características aleatorias que presenta la precipitación a lo largo de gran parte del territorio nacional, las cuales responden a la actividad convectiva que se desarrolla en verano. El objetivo de este trabajo es identificar un método de agrupamiento adecuado, que permita obtener una regionalización lo más objetiva posible y que posibilite un posterior análisis desde el punto de vista sinóptico. Al tener en cuenta los aspectos de la precipitación de verano, el hecho de considerar este trimestre permitirá poner a prueba los métodos de agrupamiento, para luego proyectar un análisis a nivel estacional. Un acabado conocimiento de las regiones homogéneas asociadas a la cantidad de

días sin precipitación a lo largo de la República Argentina puede ser útil para definir índices que permitan el monitoreo de condiciones secas a nivel regional.

## 2. DATOS

La base de datos utilizada consiste en datos diarios de precipitación procedentes de un total de 62 estaciones que corresponden al período comprendido entre 1970 y 2005, las cuales fueron provistas por el Servicio Meteorológico Nacional (SMN). Esta base de datos cuenta con series que poseen menos del 10% de sus datos faltantes, los cuales fueron sometidos a un proceso de rellenado dado que la metodología a utilizarse en este trabajo no admite matrices de entrada con dato faltante. Si bien un verano con más de 10% de dato faltante podría considerarse como verano faltante, hay casos en los cuales la mayor parte de ese porcentaje de datos faltantes provino de un mes en particular, con lo cual se optó por rellenar el mes en cuestión con su promedio climatológico de días secos. No hubo necesidad de rellenar un verano durante más de dos años consecutivos. En los casos en los cuales este relleno mensual no fuera posible, se relleno el verano completo. Los detalles de los procedimientos de control de calidad pueden encontrarse en Rivera (2009).

La ubicación geográfica de las estaciones seleccionadas se muestra en la Figura 1. Se consideró que la cobertura espacial de estas estaciones resultó satisfactoria para el tipo de análisis que se plantea realizar. Sin embargo, hubo regiones que presentaron zonas con cobertura espacial escasa, como es el caso de la región Patagónica, las áreas cordilleranas de las provincias del Noroeste y la región de Chaco, las cuales son indicadas de manera esquemática en la Figura 1. A pesar de la escasez de datos en estas regiones, se plantea obtener una regionalización que considere todo el país.

La variable de estudio es la cantidad de días secos, los cuales fueron considerados como el día donde la precipitación acumulada fue nula. La cantidad de días secos fue calculada para el trimestre de verano, compuesto por los meses de diciembre, enero y febrero. Estas series de cantidades de días secos para cada estación meteorológica fueron

sometidas a un proceso de remoción de tendencia lineal (Wilks, 2006).

## 3. ASPECTOS METODOLÓGICOS

El análisis de grupos (cluster analysis) comprende el agrupamiento de variables u observaciones similares. Esta herramienta permite el agrupamiento de estaciones meteorológicas en regiones climáticamente homogéneas, basándose en un determinado parámetro meteorológico, que en este caso son las cantidades de días secos para el trimestre de verano. Se detallan otras aplicaciones del análisis de grupos en el campo de la meteorología en el trabajo de Gong y Richman (1995) y en Wilks (2006).

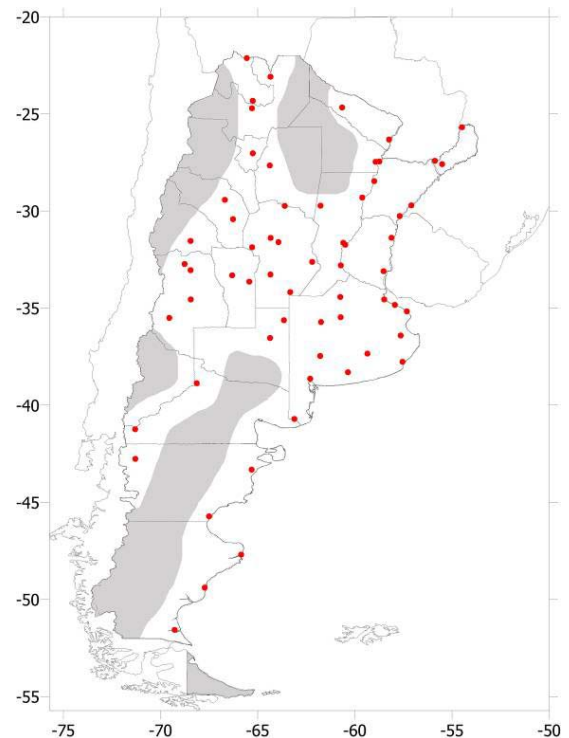


Figura 1. Ubicación de las 62 estaciones que conforman la base de datos utilizada. Las áreas grisadas indican de manera esquemática regiones sin datos.

Existen tres tipos de algoritmos de agrupamiento, los cuales se describen a continuación:

A) Los algoritmos de agrupamiento jerárquico, que se caracterizan por la construcción de una estructura de árbol de jerarquías. Ejemplos de este

tipo de algoritmos son los métodos de single linkage, complete linkage y average linkage. Otro método de este tipo utilizado comúnmente en el área climatológica es el método de Ward, el cual fue aplicado en los trabajos de Sumner (1996), Baeriswyl y Rebetz (1997), y Chambers (2001) para regionalizar la precipitación en Gales, Suiza y Australia respectivamente.

B) Los algoritmos de agrupamiento no-jerárquico permiten un reasignamiento de los miembros de cada grupo en cada etapa. Un ejemplo de este tipo de algoritmos es el método de k-means (Anderberg 1973), uno de los métodos no-jerárquicos más utilizados. Este método fue aplicado en el trabajo de Kahya y otros (2007) a fin de obtener los patrones espaciales más importantes asociados a los caudales en Turquía.

C) Análisis de componentes principales (ACP). Numerosos autores consideran a esta metodología no solo como una mera herramienta de reducción de datos, sino también como una alternativa para el análisis de grupos. En un ACP (en modo-S y matriz de correlaciones), los factores de peso de las componentes pueden ser graficados para mostrar los patrones espaciales característicos de la variable investigada. Estos factores reflejan las correlaciones entre las variables utilizadas y las componentes principales extraídas (Jayawardene y otros 2005). En base a establecer un umbral determinado para estos factores, se pueden obtener regiones homogéneas.

A partir de lo descrito anteriormente y teniendo en cuenta el objetivo de este trabajo, con el fin de encontrar la “mejor” regionalización de la cantidad de días secos se exploraron los métodos de análisis de componentes principales y k-means. Además se propone evaluar una combinación entre ambos métodos. Se eligieron estas metodologías debido a que en general, los métodos no-jerárquicos brindaron mejores resultados que los jerárquicos en el campo meteorológico (Gong y Richman 1995). Por otro lado, estos autores encontraron que los métodos basados en componentes principales rotadas son los más precisos.

### 3.1. Análisis de Componentes Principales

A través del ACP se puede simplificar la información original, a partir de la representación de los mismos objetos en un menor número de

variables (dimensiones), las cuales van a retener la mayoría de la información del espacio original. De este modo se minimiza cualquier pérdida de información, posibilitando un mejor entendimiento e interpretación de las estructuras de los datos. Esta metodología fue utilizada en modo-S, la cual permite el agrupamiento de estaciones con variaciones temporales similares en la cantidad de días secos. Las principales características del ACP pueden encontrarse en forma más extensa y con sus respectivas formulaciones matemáticas en Green (1978), Jolliffe (1986) y Richman (1986). Para aplicar esta metodología se tuvieron en cuenta los siguientes aspectos que pueden modificar los resultados:

#### *i. Distribución espacial de los datos*

Errores potenciales pueden surgir al utilizar una base de datos cuya distribución espacial no sea homogénea al realizar un ACP en modo-S. Esta distribución irregular de datos influencia los factores de peso, haciendo que los mismos reflejen la varianza de las áreas con mayor densidad de puntos (Karl y otros 1982). Para este trabajo se optó por eliminar las estaciones que aportaran información redundante, de forma tal de obtener una distribución espacial de estaciones más homogénea.

#### *ii. Matriz de similitud*

Se utilizó la matriz de correlaciones, la cual es comúnmente aplicada cuando diferentes variables tienen diferentes varianzas (Jayawardene y otros 2005). El uso de esta matriz inicialmente le da el mismo peso a todas las variables de la matriz de entrada (que en este caso son las estaciones meteorológicas utilizadas). Esto se debe a que todas se encuentran estandarizadas o tipificadas, es decir, poseen media nula y desvío estándar unitario. En el caso de trabajar con la cantidad de días secos, que es una variable asociada a la estacionalidad de la precipitación, permite comparar regiones con distintos regímenes de precipitación.

#### *iii. Distribución de probabilidad.*

El ACP no demanda explícitamente que la variable utilizada se encuentre normalmente distribuida para operar correctamente (Fovell y Fovell 1993; Kalayci y Kahya 2006). En la práctica, ACP

funciona de manera robusta frente a datos de entrada que poseen desviaciones moderadas respecto a la distribución normal (Comrie y Glenn 1998). No obstante, el coeficiente de correlación de Pearson utilizado en la matriz de entrada de ACP puede ser afectado por la no-normalidad de los datos (White y otros 1991; Kalayci y Kahya 2006). En vista de esta problemática, numerosos estudios aplican diversas transformaciones a las variables utilizadas de forma tal de reducir la asimetría de las mismas y obtener variables cuya distribución de probabilidad se aproxime a una distribución normal (Drosdowsky 1993; Romero y otros 1999; Phillips y Denning 2007; entre otros).

Es por esto que, para la continuidad del trabajo, se evaluaron las distribuciones de probabilidad para las estaciones meteorológicas utilizadas mediante el test de Chi-cuadrado, el cual fue considerado significativo al 5% (Wilks 2006). A través de este análisis se obtuvo que únicamente tres estaciones en todo el país presentaron una distribución de probabilidades cuyo ajuste a una distribución normal no es significativo. Por lo tanto, se decidió no transformar la variable y realizar el cálculo del ACP sobre los datos originales.

#### *iv. Criterios de corte*

En este trabajo se analizaron conjuntamente los resultados de tres criterios a fin de separar las componentes con señales más claras y dejar de lado las componentes que representen ruido.

El “scree test” (Cattell 1966), se basa en graficar la varianza explicada por cada CP en función del orden en el que fueron extraídas. Una vez obtenida una curva, se procede a buscar el “codo” o “quiebre” en la misma. El criterio de Kaiser (1958) propone retener las CP cuyos autovalores sean mayores que 1, lo cual sugiere la retención de aquellas componentes que expliquen una fracción de varianza mayor que la explicada por cada variable original. Por último se estableció un corte en un determinado valor de porcentaje de varianza total a fin de poder separar las estructuras que aporten solo ruido. Inicialmente se estableció considerar las componentes que acumularan un porcentaje de varianza de 75%. Para cada criterio se evaluaron las regiones resultantes a través de la inspección de los factores de peso de cada una de las componentes retenidas, a fin de determinar la

correcta elección de las mismas, y se buscó un acuerdo entre los tres criterios.

#### *v. Rotación de las componentes principales.*

Cuando el principal objetivo del ACP es la interpretación física en vez de la reducción de variables, es conveniente rotar una sub-muestra de los autovectores obtenidos inicialmente en un nuevo conjunto de vectores, a fin de mejorar la interpretación de los resultados (Baeriswyl y Rebetez 1997; Comrie y Glenn 1998; Phillips y Denning 2007). En este trabajo se evaluó la rotación Varimax (Kaiser 1958), la cual fue utilizada para transformar los vectores asociados al ACP en una estructura simple. En el caso de existir una estructura simple, ésta se manifiesta con componentes con mayores valores de factores de peso en algunas variables y factores de peso casi nulos en el resto. En el análisis en modo-S esto resulta en una regionalización de las variables espacialmente distribuidas (Drosdowsky 1993).

#### *vi. Determinación de las regiones homogéneas.*

Una vez rotadas y graficados los factores de peso de cada una las componentes principales retenidas, resta determinar un criterio que permita decidir que regiones resultan significativas. Estas regiones se obtuvieron en base de un criterio que consideró apropiado tomar como umbral el factor de peso de  $r = |0.4|$ , propuesto por Richman y Lamb (1985). El hecho de considerar este valor como umbral, si bien fue conveniente, no deja de ser arbitrario. Se utilizó la letra  $r$  para denominar los factores de peso dado que estos son las correlaciones entre las componentes principales y las variables utilizadas, que en este caso fueron las series temporales de cantidades de días secos para cada estación meteorológica. Podrían encontrarse casos para los cuales haya estaciones que queden sin clasificación por no llegar al valor de coeficiente elegido. Quizás todas las estaciones puedan ser clasificadas eligiendo el umbral de  $r = |0.3|$ , o algunas pocas siendo el umbral de  $r = |0.5|$ , lo cual hace que este criterio no resulte objetivo.

Dado que al considerar este criterio la distribución espacial de los factores de peso mostró que las regiones resultantes presentan intersecciones, para comparar de manera adecuada las soluciones provenientes del ACP rotadas y del método de k-means es necesario aplicar alguna forma de

separación de las regiones, sin dejar de contemplar la física contenida dentro cada una de ellas. Es por eso que se utilizó el criterio del máximo factor de peso para poder llevar a cabo esta separación. Si una estación se encuentra en la intersección de dos regiones distintas, asociadas cada una a una componente distinta, se adjudica esa estación a la región con la cual posea un mayor valor de factor de peso. En este caso la interpretación de las áreas homogéneas resulta ser más clara que al considerar el umbral de  $r = |0.4|$  y posibilita la comparación de los resultados obtenidos a través de ambas metodologías. De todos modos se consideró el umbral de  $r = |0.4|$  como condición necesaria para que las estaciones pertenezcan a una región homogénea.

### 3.2. Método de K-means

El método de k-means es un método de agrupamiento no-jerárquico, cuyo objetivo es encontrar la partición óptima para dividir un número de objetos en k grupos. Este procedimiento mueve los objetos de un grupo a otro con el objetivo de minimizar la varianza dentro de cada grupo y maximizar la varianza entre grupos. En primera instancia se establece el valor de k, ya sea mediante una técnica específica o en base a una determinación subjetiva. Una vez determinado este valor, entre las opciones de inicialización del algoritmo se eligió que las observaciones iniciales maximicen las distancias iniciales entre grupos. Luego se procedió al cálculo del algoritmo tal como se describe en Wilks (2006). Las siguientes consideraciones metodológicas fueron tenidas en cuenta al momento de su aplicación:

#### *i. Medida de distancia*

Previo a la aplicación del algoritmo de cálculo de k-means, es necesario establecer una medida de similaridad o disimilaridad para caracterizar las relaciones entre los objetos. Se optó por utilizar la distancia Euclídea, que es la distancia geométrica entre dos objetos  $i$  y  $j$  en el espacio  $k$ -dimensional de vectores, dado que esta medida de distancia genera soluciones más precisas, comparada con la correlación inversa (Gong y Richman 1995).

#### *ii. Cantidad de grupos*

Un problema adicional asociado con el uso del análisis de grupos es la selección de un número apropiado de grupos. Existe un gran número de técnicas gráficas y estadísticas a partir de las cuales puede seleccionarse un número adecuado de grupos (Sumner 1996). El trabajo de Milligan y Cooper (1985) evaluó 30 reglas para determinar un número óptimo de clusters a considerar. Los autores obtuvieron que el test denominado “pseudo-F” de Calinski y Harabasz (1974) presentó un mejor desempeño en comparación con el resto de los criterios evaluados. Por lo tanto en este trabajo se utilizó el estadístico pseudo-F, el cual está dado por la siguiente fórmula:

$$pseudo - F = \frac{A (n - k)}{W (k - 1)}$$

donde A y W son las variabilidades entre y dentro de cada cluster, respectivamente, n es el número de objetos, y k es el número de clusters. A fin de determinar la cantidad de grupos sugerida por el estadístico, se realizó el cálculo del mismo para una cierta cantidad de valores de k, que en el caso de este trabajo varió de 2 a 15 grupos. Para cada valor de k se calcularon los parámetros A y W y de esta forma se determinaron los valores del estadístico, obteniendo una curva con 14 valores. Siguiendo la recomendación de Romero y otros (1999), se buscaron los picos locales en el estadístico pseudo-F a fin de encontrar el número de grupos óptimo. Mediante el análisis de la distribución espacial que presentaron las regiones resultantes para cada valor de k significativo, se determinó la cantidad de grupos a considerar.

### 3.3. Combinación entre ACP y K-means

Para este trabajo se utilizó la base de componentes principales retenidas rotadas como matriz de entrada al método de k-means, obtenida según se explicó anteriormente. La aplicación de algoritmos de agrupamiento (jerárquicos o no-jerárquicos) sobre las CP rotadas brinda una metodología alternativa y fue evaluada en comparación con los resultados obtenidos del ACP y del método de k-means. La combinación entre el método de k-means y el ACP ha sido utilizada anteriormente en

los trabajos de Romero y otros (1999) para obtener áreas homogéneas de la precipitación diaria en España; y Demirel y otros (2007) para obtener regiones homogéneas en cuanto a los caudales mínimos en Turquía. Cabe destacar que en las tres metodologías propuestas para lograr la regionalización se tuvo en cuenta la climatología de la precipitación y la cantidad media de días secos para el trimestre de verano (Rivera 2009). Esta consideración se utilizó como criterio para definir y caracterizar las regiones encontradas con el fin de lograr regiones coherentes no sólo desde el punto de vista de la homogeneidad en la variabilidad temporal.

## 4. REGIONALIZACIÓN

### *i. Análisis de componentes principales.*

Para el trimestre de verano, la regla de Kaiser propone la retención de las primeras 14 componentes, acumulando un total de 85.8% de la varianza (Tabla I), pero esta cantidad de componentes resulta excesiva e imposibilita un análisis objetivo de las mismas. Al elegir el nivel de varianza correspondiente al 75%, las 10 componentes resultantes siguieron siendo excesivas, produciendo patrones espaciales de difícil interpretación. Por lo tanto, para este trimestre se realizó el siguiente procedimiento: en base a los resultados del scree test (Figura 2) la elección más clara respecto al número de componentes corresponde al quiebre que presenta la gráfica en la tercer componente. Se consideró que estas 3 componentes rotadas no son suficientes para describir los modos de variabilidad presentes durante el trimestre de verano. Pero el resultado del scree test puede ser tomado como un número base de componentes a retener. Por lo tanto, el número mínimo de componentes que son consideradas para la rotación corresponde a 3. En el siguiente paso, se rotaron las primeras 4 componentes y se examinaron los mapas de los factores de peso resultantes. Este proceso se repitió con las sucesivas cantidades de componentes hasta acumular 10 componentes. Como resultado se obtuvo que las regiones más estables se encontraron reteniendo 8 componentes.

De esta manera, el hecho de retener las primeras 8 componentes corresponde a considerar como umbral un valor mínimo de varianza acumulada de

70% (Tabla I). Se considera que el porcentaje de varianza explicado, si bien no alcanza el valor de 75% prefijado, es adecuado para el análisis y permite separar de forma conveniente las componentes más importantes.

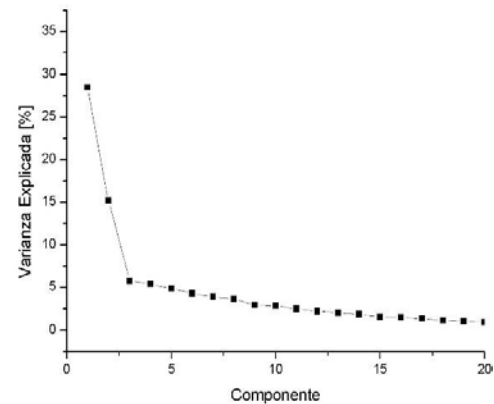


Figura 2. Scree test de las CP. Se muestran solo las primeras 20 componentes.

| CP | Autovalor | Varianza exp. (%) | Varianza acum. (%) |
|----|-----------|-------------------|--------------------|
| 1  | 17.36     | 28.47             | 28.47              |
| 2  | 9.26      | 15.18             | 43.65              |
| 3  | 3.52      | 5.77              | 49.42              |
| 4  | 3.29      | 5.40              | 54.82              |
| 5  | 2.97      | 4.87              | 59.68              |
| 6  | 2.61      | 4.28              | 63.96              |
| 7  | 2.37      | 3.89              | 67.85              |
| 8  | 2.20      | 3.60              | 71.45              |
| 9  | 1.79      | 2.93              | 74.38              |
| 10 | 1.74      | 2.85              | 77.23              |
| 11 | 1.50      | 2.43              | 79.65              |
| 12 | 1.36      | 2.19              | 81.85              |
| 13 | 1.26      | 2.02              | 83.87              |
| 14 | 1.15      | 1.86              | 85.73              |

Tabla I. Autovalores y porcentajes de varianza explicados y acumulados para las primeras 14 CP.

Para estas 8 componentes retenidas (no rotadas) se graficaron sus factores de peso con el propósito de visualizar los patrones espaciales obtenidos para cada una de ellas (Figura 3). En la Figura 3a se solaparon las regiones obtenidas mediante el contorno de  $r = |0.4|$  para cada una de las componentes, mientras que en la Figura 3b se

presentan los resultados obtenidos luego de la aplicación del criterio del máximo factor de peso a cada una de las estaciones. El número asignado a cada región corresponde al orden de cada componente extraída y las letras corresponden a las diferentes sub-regiones, asignadas a regiones homogéneas ubicadas en regiones geográficas

diferentes. En el caso de regiones contiguas se propuso la división teniendo en cuenta la climatología de la precipitación y de la cantidad media de días secos para el trimestre de verano (Rivera 2009). Se observa que la interpretación de las áreas homogéneas resulta ser más clara que en la Figura 3a.

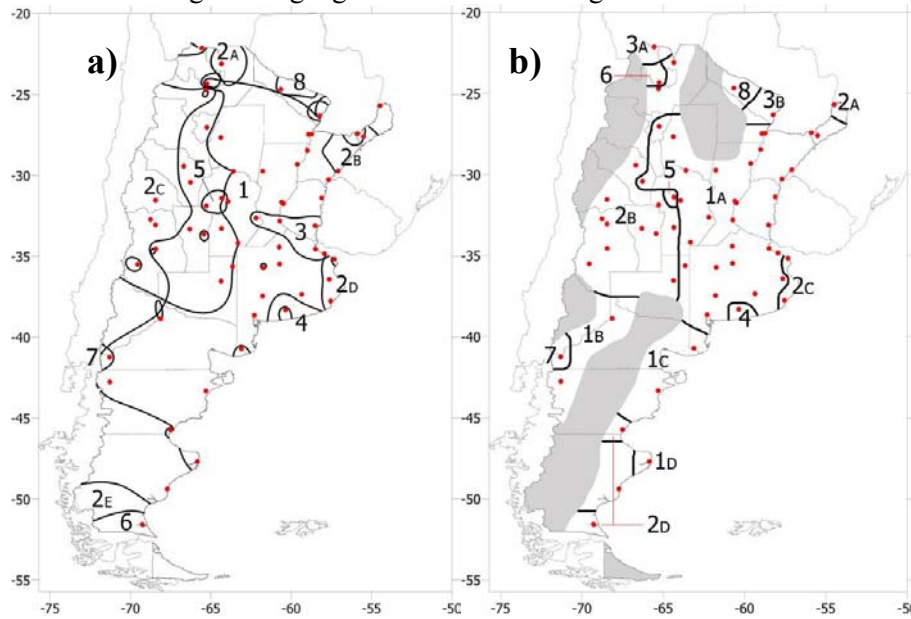


Figura 3. Regiones obtenidas mediante la aplicación de ACP previo a la rotación. El número asignado a cada región corresponde al orden de cada componente. a) Considerando como límite el contorno de  $r = |0.4|$ , b) Considerando como límite la regla del máximo factor de peso. Las áreas grisadas corresponden a regiones sin datos

El patrón espacial asociado a la componente 1 ocupa gran parte del territorio nacional (Figura 3b), abarcando la porción Centro-Este y el Norte de la Patagonia; mientras que el patrón asociado a la componente 2 se centra sobre la región Centro-Oeste del país. Dado que las 6 componentes restantes representan áreas homogéneas pequeñas, se considera que esta regionalización no es adecuada para el análisis. Es por ello que se procede a evaluar la regionalización obtenida luego de la rotación de las 8 componentes principales a través del método Varimax. La Figura 4 muestra los factores de peso de cada una de las componentes rotadas, las cuales representan diferentes modos geográficos de variabilidad.

La combinación de las regiones consideradas en la Figura 4 se presenta en la Figura 5, donde se puede observar que las primeras tres componentes son las que abarcan mayores cantidades de estaciones meteorológicas (Figura 5b). Se observó que las regiones obtenidas presentan muchas divisiones, en

la mayoría de los casos asociadas a regiones geográficamente distantes. Esto se verificó en las regiones 2, 5 y 6. En particular las sub-regiones 6B y 6C se separaron de acuerdo a los regímenes de precipitación presentes en ambas sub-regiones (Paruelo y otros 1998). La existencia de esta gran cantidad de sub-regiones puede responder a que en verano las precipitaciones convectivas dominan gran parte del territorio nacional y su distribución espacial es errática. Estos patrones espaciales de convección errática resultan en un incremento en la heterogeneidad de las divisiones climáticas. Teniendo en cuenta que la mayoría de las sub-regiones abarca una porción pequeña del país, en algunos casos con solo una estación meteorológica, se pone en duda la representatividad de las mismas. El hecho de considerar que un número elevado de regiones no brinda una solución correcta desde el punto de vista climático radica en que si la solución cuenta con muchas regiones, la misma contendrá demasiados detalles que imposibilitarán un análisis



Regionalización de los días secos...

desde el punto de vista de la climatología sinóptica que afecte estas áreas homogéneas. Por lo tanto otros factores de escala más pequeña, como efectos

asociados a la mesoescala, deberían ser considerados para su interpretación, análisis que esta fuera de los objetivos de este trabajo.

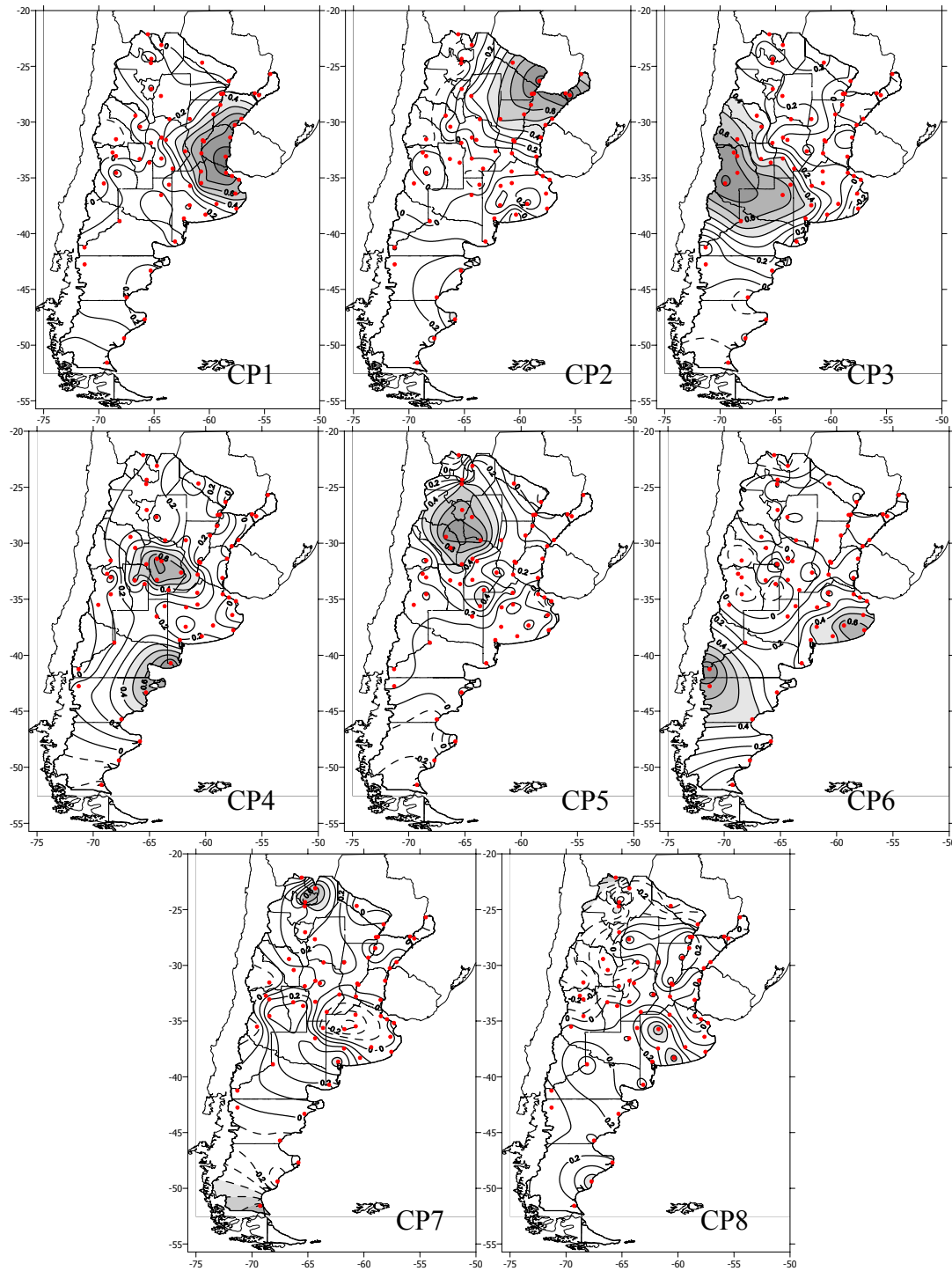


Figura 4. Distribución espacial de los factores de peso para las primeras 8 componentes principales rotadas mediante el método Varimax, obtenidas a partir de las cantidades de días secos de verano. Los valores negativos se muestran en líneas de trazos. Los valores superiores a  $r = |0.4|$  se encuentran sombreados.

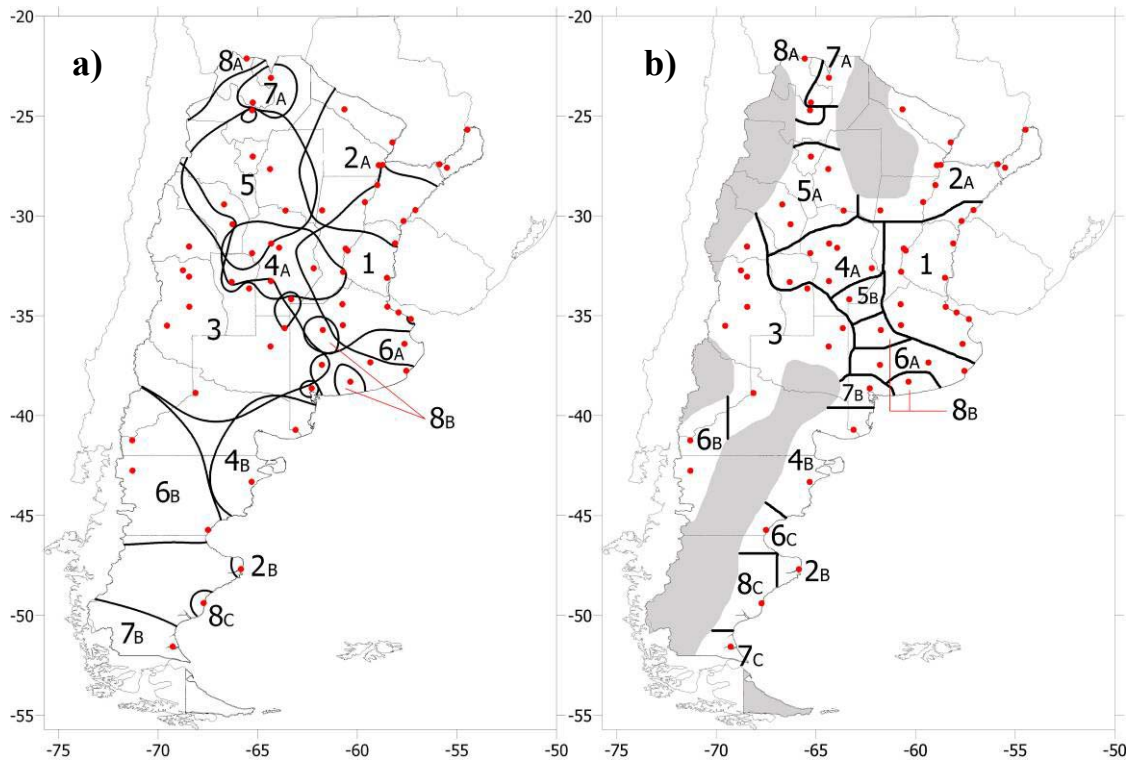


Figura 5. Ídem Figura 3 para el caso del ACP rotadas.

#### ii. Método de k-means.

Una vez analizados los resultados de la regionalización basada en ACP, se describen los resultados asociados a la aplicación del método de k-means, a fin de realizar una comparación posterior entre ambos métodos. Es importante destacar que la aplicación de este método en principio se realizó directamente sobre las cantidades de días secos, obteniendo en ese caso una regionalización poco satisfactoria, la cual generó grupos cuya distribución geográfica fue muy amplia (resultado no mostrado). Es por esto que se consideró la alternativa de aplicar este método utilizando como variables de entrada los factores de peso de las componentes retenidas rotadas anteriormente.

En la Figura 6 se muestra el gráfico del estadístico pseudo-F en función del número de grupos k. Los máximos locales en este estadístico indican un número de grupos óptimo, por lo tanto se buscó identificar estos máximos y se evaluaron las regiones obtenidas para los distintos picos, a fin de determinar la distribución espacial más adecuada respecto a la climatología. Un primer máximo aparece claramente para 5 grupos, mientras que se

observan máximos secundarios en 7, 10 y 12 grupos. En base a la evaluación de las regiones resultantes luego de la regionalización, se determinó que un total de 5 grupos presentaron una distribución razonable, la cual se muestra en la Figura 7. Cabe considerar que el número asignado a cada región no representa ningún tipo de ordenamiento jerárquico. En función de la cantidad media de días secos para el verano (Rivera 2009), la región 2 se separó en las sub-regiones 2A y 2B. A pesar de que el contraste en los valores medios de la cantidad de días secos entre las estaciones de la sub-región 4C y la sub-región 4D no es tan marcado, se optó por dividir ambas regiones de todas formas dadas las diferencias climáticas existentes en la precipitación (Paruelo y otros 1998). Lo mismo sucedió para la sub-región 4B, aunque en este caso la diferencia en las cantidades medias de días secos resulta ser más evidente respecto a la región 4D. A través de esta metodología se obtuvieron como regiones principales las regiones Centro-Este (5); Noreste (3A); del Centro y Sur Bonaerense (4B); Noroeste (2A y 2B); Centro-Oeste (1B) y Patagónica (4C y 4D), las cuales son climáticamente coherentes.

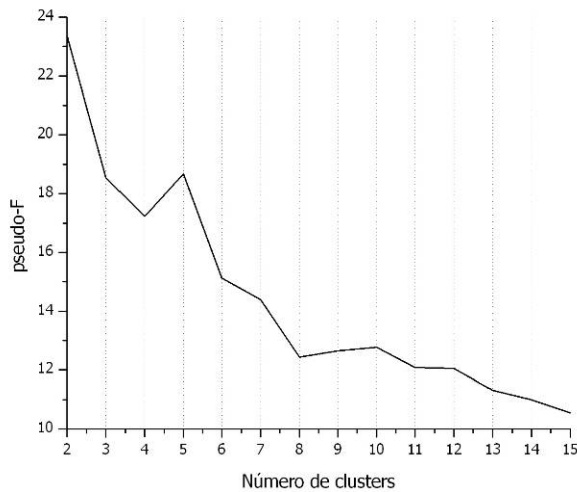


Figura 6. Resultados del estadístico pseudo-F para los distintos números de grupos.

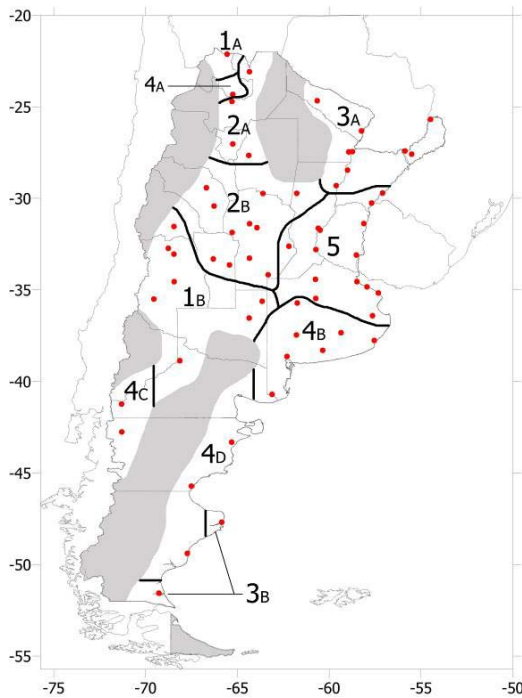


Figura 7. Regiones obtenidas mediante la aplicación del método de k-means para el trimestre de verano. Las áreas grisadas corresponden a regiones sin datos.

## 5. COMPARACIÓN DE LOS MÉTODOS Y DETERMINACIÓN DEL MÉTODO MÁS ADECUADO

La comparación de los resultados obtenidos a través de ACP rotadas y k-means combinado con ACP fue posible, dado que ACP utilizó la varianza contenida en la base de CP retenidas y el método de k-means fue utilizado considerando como matriz de entrada esta base de componentes, con lo cual se logró una consistencia en la información introducida en ambos métodos. Si se hubiese utilizado el método de k-means directamente sobre las cantidades de días secos, el método estaría utilizando el total de la varianza, con lo cual una comparación entre estas técnicas de regionalización podría resultar confusa.

Para el trimestre de verano aparecen algunos rasgos espaciales comunes a ambos métodos, como son las regiones del Noreste, Centro-Este, Centro-Oeste y Noroeste (Figuras 5 y 7). Se consideró que para este trimestre el método de k-means brinda los mejores resultados, dado que no presentó una distribución espacial tan heterogénea como se observa para el caso de ACP. Además, la mayoría de las sub-regiones obtenidas mediante ACP abarca una porción pequeña del país, comprendiendo en muchos casos solo una estación meteorológica, lo cual indica que el método no es adecuado en la representación de los patrones típicos de verano. A pesar de la aleatoriedad que presenta la precipitación en la mayor parte del país, se buscó definir y caracterizar las regiones desde el punto de vista de la climatología sinóptica. Esto permitirá un posterior análisis de las regiones a través del estudio de los procesos responsables de la generación de la precipitación, lo cual excede los objetivos de este trabajo.

## 6. DISCUSIÓN Y CONCLUSIONES

En este trabajo se determinaron regiones climáticas homogéneas desde el punto de vista de la variabilidad temporal de las cantidades de días secos de verano sobre la República Argentina. La regionalización fue llevada a cabo mediante dos métodos ampliamente utilizados en el ámbito meteorológico, el análisis de componentes principales (ACP) y el método de k-means.

El método de ACP fue utilizado en modo-S. Luego de evaluar tres criterios de corte para determinar la cantidad de componentes a retener para el análisis, se procedió a rotar las mismas mediante el método Varimax. Se evaluaron dos técnicas para la definición de las regiones, obteniendo que el criterio del máximo factor de peso permitiera una regionalización que posibilita distinguir adecuadamente las regiones obtenidas.

El método de k-means al ser aplicado directamente sobre las cantidades de días secos brindó resultados no satisfactorios. En vista de este inconveniente y con la intención de reducir las dimensiones de la matriz de entrada, el método fue aplicado sobre la base de factores de peso surgida del ACP, lo cual brindó una metodología de regionalización alternativa.

Luego de la evaluación de las regionalizaciones obtenidas a través de ambos métodos, se determina que el agrupamiento generado a partir de la aplicación del método de k-means es el más adecuado, dado que forma regiones que posibilitan un análisis desde el punto de vista de la climatología sinóptica. Para el caso de la regionalización obtenida mediante ACP, se obtienen muchas sub-divisiones, las cuales pueden responder a procesos asociados a escalas más pequeñas, como la mesoescala, y resultan en un incremento en la heterogeneidad de las divisiones climáticas. Se logró sub-dividir el territorio en regiones congruentes respecto de distintas propiedades de la variable como su variabilidad temporal y sus valores medios, lo cual dio como resultado seis regiones principales: las regiones Centro-Este; Noreste; Centro y Sur Bonaerense; Noroeste; Centro-Oeste y Patagónica, las cuales son climáticamente coherentes.

Ciertos aspectos de ambas metodologías deben ser tenidos en cuenta al momento de su aplicación, dado que las dos herramientas poseen diversas instancias en las cuales el usuario debe tomar decisiones respecto a su forma de aplicación. Tanto el número de componentes significativas retenidas para su rotación, como el número óptimo de regiones a ser generadas mediante del método de k-means, resultan de un proceso de selección subjetivo, a pesar de los criterios de corte utilizados en este trabajo. Luego, la inspección gráfica de las regionalizaciones puede ser útil al momento de determinar el número de regiones adecuado, pero

en ningún caso se obtendrá una regionalización totalmente objetiva dadas las instancias previas de cálculo. Otro aspecto a tener en cuenta es que la comparación entre las dos regionalizaciones obtenidas fue posible dado que ambos métodos retienen el mismo porcentaje de varianza respecto a la varianza contenida en los datos originales. De haber aplicado el método de k-means directamente sobre las cantidades de días secos, la comparación directa de los resultados no hubiese sido del todo adecuada.

**Agradecimientos:** Al Servicio Meteorológico Nacional por la provisión de los datos para la realización de este trabajo. Este trabajo fue realizado con apoyo del proyecto de investigación European Community's Seventh Framework Programme (FP7/2007-2013) en el marco del subsidio N°212492 (CLARIS LPB. A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin); y los siguientes proyectos: UBA X170, UBA X605 y BID 1728/OC-AR-PICT 38273.

## REFERENCIAS

- Anderberg, M. R., 1973. Cluster Analysis for Applications. Academic Press, 359 pp.
- Baeriswyl, P.-A. y Rebetz, M., 1997. Regionalization of Precipitation in Switzerland by Means of Principal Component Analysis. *Theor. Appl. Climatol.*, **58**, 31-41.
- Balling, R. C. y Goodrich, G. B., 2007. Analysis of drought determinants for the Colorado River Basin. *Clim. Change*, **82**, 179-194.
- Barreira, S. y Compagnucci, R. H., 2007. Sea ice concentration temporal variability over the Weddell Sea and its relationship with tropical sea surface temperature, in Antarctica: A Keystone in a Changing World – Online Proceedings of the 10th ISAES, edited by A. K. Cooper and C. R. Raymond et al., USGS Open-File Report 2007-1047, Short Research Paper 029, 5 p.; doi:10.3133/of2007-1047.srp029.
- Calinski, R.B. y Harabasz, J., 1974. A dendrite method for cluster análisis. *Commun. Stat.*, **3**, 1-27.

- Cattell, R. B., 1966. The scree test for the number of factors. *J. Multiv. Behav. Res.*, **1**, 245-276.
- Chambers, L., 2001. Classifying rainfall districts: a south Western Australian study. *Aus. Met. Mag.*, **30**, 91-103.
- Compagnucci, R. H. y Vargas, W. M., 1985. Regionalización del campo de presión del mes de julio para la parte austral de Sudamérica. *Geoacta*, **13** (1), 71-79
- Compagnucci, R. H. y Araneo, D. C., 2005. Identificación de áreas de homogeneidad estadística para los caudales de ríos andinos argentinos y su relación con la circulación atmosférica y la temperatura superficial del mar. *Meteorológica*, **30** (1y2), 41-53.
- Comrie, A. C. y Glenn, E. C., 1998. Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Clim. Res.*, **10**, 201-215.
- Coronato, F. y Bisigato, A., 1998. A temperature pattern classification in Patagonia. *Int. J. Climatol.*, **18**, 765-773.
- Cuadrat Prats, J. M. y Vicente-Serrano, S. M., 2004. Comportamiento de las sequías en la península Ibérica: Análisis mediante el Standardized Precipitation Index, en: García Codron y otros (Eds.) (2004). *El Clima entre el Mar y la Montaña*. Asociación Española de Climatología y Universidad de Cantabria, Serie A, nº 4, Santander.
- Demirel, M., Mariano, A. J., Kahya, E., 2007. Performing k-means analysis to drought principal components of Turkish rivers. 27th AGU Hydrology Days, Fort Collins, Colorado, March 19-21.
- Drosowsky, W., 1993. An analysis of Australian seasonal rainfall anomalies: 1950-1987. I: Spatial patterns. *Int. J. Climatol.*, **13**, 1-30.
- Fovell, R. G., Fovell, M.-Y. C., 1993. Climate zones of the conterminous United States defined using cluster analysis. *J. Climate*, **6**, 2103-2135.
- Gong, X. y Richman, M. B., 1995. On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate*, **8**, 897-931.
- Green, P. E., 1978. *Analysing Multivariate Data*. The Dryden Press: Illinois, USA, 519 pp.
- Jayawardene, H. K. W. I., Sonnadara, D. U. J., Jayewardene, D. R., 2005. Spatial interpolation of weekly rainfall depth in the dry zone of Sri Lanka. *Clim. Res.*, **29**, 223-231.
- Jolliffe, I. T., 1986. *Principal Component Analysis*. Springer-Verlag. 271 pp.
- Kahya, E., Demirel, M. y Piechota, T., 2007. Spatial grouping of annual streamflow patterns in Turkey. 27th AGU Hydrology Days, Fort Collins, Colorado, March 19-21.
- Kalayci, S. y Kahya, E., 2006. Assessment of streamflow variability modes in Turkey: 1964-1994. *J. Hydrol.*, **324**, 163-177.
- Kaiser, H. F., 1958. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200.
- Karl, T R., Koscielny, A. J. y Diaz, H. F., 1982. Potential errors in the application of principal component (eigenvector) analysis to geophysical data. *J. Appl. Meteorol.* **21**, 1183-1186.
- Lana, X., Serra, C. y Bargaño, A., 2001. Patterns of monthly rainfall shortage and excess in terms of the Standardized Precipitation Index for Catalonia (NE Spain). *Int J Climatol*, **21**, 1669-1691.
- Milligan, G. W. y Cooper, M. C., 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, **50**, 159-179.

- Paruelo, J. M., Beltrán, A., Jobbágy, E., Sala, O. E. y Golluscio, R. A., 1998. The Climate of Patagonia: general patterns and controls on biotic processes. *Ecología Austral*, **8**, 85-101.
- Phillips, I. D. y Denning, H., 2007. Winter daily precipitation variability over the South West Peninsula of England. *Theor. Appl. Climatol.* **87**, 103-122.
- Raziei, T., Bordi, I. y Pereira, L. S., 2008. A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrol. Earth Syst. Sci.*, **12**, 1309-1321.
- Richman, M. B. y Lamb, P. J., 1985. Climatic pattern analysis of 3- and 7-day summer rainfall in the central United States: Some methodological considerations and a regionalization. *J. Climate Appl. Meteor.*, **24**, 1325-1343.
- Richman, M., 1986. Rotation of Principal Components. *J. Climatol.* **6**, 293-335.
- Rivera, J. A., 2009. Variabilidad espacio-temporal de los días secos en Argentina. Tesis de Licenciatura en Ciencias de la Atmósfera. Universidad de Buenos Aires.
- Romero, R., Ramis, C., Guijarro, J. A. y Sumner, G., 1999. Daily rainfall affinity areas in mediterranean Spain. *Int. J. Climatol.* **19**, 557-578.
- Soulé, P. T., 1990. Spatial patterns of multiple drought types in the contiguous United States: a seasonal comparison. *Clim. Res.*, **1**, 13-21.
- Sumner, G., 1996. Daily precipitation patterns over Wales: towards a detailed precipitation climatology. *Transactions of the Institute of British Geographers, New Series*, **21** (1), 157-176.
- Vicente-Serrano, S. M., González-Hidalgo, J. C., de Luis, M. y Raventós, J., 2004. Drought patterns in the Mediterranean area: the Valencia region (East-Spain). *Clim. Res.*, **26**, 5-15.
- White, D., Richman, M., Yarnal, B., 1991. Climate regionalization and rotation of principal components, *Int. J. Climatol.*, **11**, 1-25.
- Wilks D.F., 2006. *Statistical Methods in the Atmospheric Sciences (Second Edition)*. Academic Press, 627 pp.