

Regionalización de caudales. Propuesta metodológica para la identificación de regiones homogéneas

Marta del Carmen Paris
Graciela Viviana Zucarelli

Universidad Nacional del Litoral, Argentina

La variabilidad espacial de las condiciones del escurrimiento, comúnmente encontrada en las cuencas hidrográficas, requiere de una complementación de las metodologías de estimación de información, a partir de los datos registrados puntualmente, que permita considerar el alcance de su representatividad. Esto lleva a definir zonas en donde se mantenga cierta constancia de las características hidrológicas; es decir, la delimitación o identificación de regiones hidrológicamente homogéneas. No obstante, en muchas ocasiones no es trivial determinar un criterio para establecer dichas regiones. En este trabajo se presentan los resultados alcanzados por la aplicación de métodos estadístico-matemáticos multivariados (análisis de agrupamiento, componentes principales y gráfica de Andrews) en la identificación de las regiones homogéneas. Tal propuesta metodológica es una alternativa en donde se contempla no sólo el análisis de los datos multidimensionales, sino también su evaluación integrada con la caracterización conceptual del sistema. El área de aplicación abarca 11 cuencas del noroeste de Argentina, que cuentan con registros históricos de 605 crecidas. Sin embargo, en esta zona del país, de marcada aridez, en donde se torna indispensable realizar un manejo eficiente de los recursos hídricos, sólo tres de tales cuencas continúan siendo aforadas en la actualidad. La identificación de regiones homogéneas presentada en este trabajo ha permitido establecer pautas para el diseño de una red de mediciones que considera la representatividad de las observaciones y su potencialidad para la transferencia de información.

Palabras clave: regiones homogéneas, regionalización, análisis de agrupamiento, componentes principales, curvas de Andrews, Argentina.

Introducción

El agua constituye un recurso limitado en áreas áridas y semiáridas. Esto requiere que se establezcan estrategias de gestión adecuadas para garantizar su utilización eficiente. Para concretar dicha meta es indispensable conocer las particularidades de los sistemas hídricos que comprende el área. Sin embargo, la escasez de información de series históricas es una situación frecuente. Además, esto se agrava fundamentalmente por la poca cantidad de cuencas aforadas, la discontinuidad de las series y la confiabilidad de los datos disponibles (incertidumbre y representatividad).

Precisamente, el propósito principal de la regionalización hidrológica es la transferencia de información de cuencas aforadas a las no aforadas, que pertenecen a una misma región hidrológica (Tucci, 1987; Franchini y Suppo, 1996). Esta discretización espacial se realiza frecuentemente teniendo en cuenta las características geomorfológicas y climáticas —que definen la respuesta hidrológica de la región— y puede resultar en zonas sin continuidad geográfica, aunque su comportamiento hidrológico sea semejante (Andrade y Hawkins, 2000). Dicho de otro modo, sus respuestas hidrológicas pueden ser comparadas, facilitando de esta manera la transferencia de información.

Dado que la cuenca hidrográfica es usualmente considerada como unidad de gestión de los recursos hídricos comprometidos en la misma, la identificación de regiones hidrológicamente homogéneas es un aspecto que merece ser tenido en cuenta a la hora de realizar estimaciones de aquellas variables hidrológicas involucradas en el planteamiento de estrategias de gestión (como caudales de diseño, eventuales escenarios de regulación, proyecto de obras y capacidad de los cursos de agua, entre otros) y contemplar la heterogeneidad de ambientes presentes dentro de esta unidad de gestión.

En este trabajo se presentan los resultados alcanzados por la aplicación de métodos estadístico-matemáticos multivariados (análisis de agrupamiento, componentes principales y curva de Andrews) en la identificación de las regiones homogéneas en 11 cuencas del noroeste de Argentina, con miras a la definición estratégica sobre la continuidad de los puestos de aforo que permitan obtener estimaciones representativas del caudal mediano ($Q_{50\%}$ =caudal asociado con el 50% de probabilidad de ocurrencia), que cuantifica la disponibilidad hídrica del recurso agua.

Metodología

La evaluación de los sistemas hídricos involucra la cuantificación sistemática de aquellas variables consideradas relevantes para su estudio y caracterización. En este sentido, los métodos estadísticos adquieren gran importancia en el manejo de dichos datos. Ello se debe fundamentalmente a su capacidad de optimización y síntesis en la generación de información.

Además, al considerar numerosas variables, el problema en estudio presenta una múltiple referenciación y/o se conforman bases de datos de gran tamaño. Esta situación sugiere la necesidad de adoptar un enfoque multivariado para el procesamiento y análisis, ya que de lo contrario sólo se realizaría la interpretación circunscripta e independiente de cada una de estas variables.

Métodos como el análisis de agrupamiento, análisis factorial, componentes principales y análisis de correspondencias, están dirigidos a la identificación de asociaciones o grupos de entidades similares, contemplando las particularidades de los datos observados (diferentes escalas de medida, unidades, órdenes de magnitud, etcétera). Además, las diferentes alternativas de representación de los resultados obtenidos (dendrogramas, diagramas de cargas, distribución espacial de las marcas, curva de Andrews, etcétera), facilitan la interpretación de los mismos.

Si bien existen diferentes estrategias para el tratamiento de datos, la elección de alguna de ellas

depende de: a) las interrelaciones entre las variables, b) las preferencias del investigador y c) la aplicación del principio de la parsimonia, donde la simplicidad de la interpretación es un aspecto primario (Brown, 1998).

En todos los casos, la caracterización conceptual preliminar del problema a analizar, el examen y la preparación de los datos básicos que se utilizarán en el tratamiento multivariado es una etapa importante, que incluye:

- La selección de las variables que intervienen en este análisis multidimensional sobre la base de su representatividad, en relación con los fenómenos en estudio, caracterización univariada previa, cantidad, estacionariedad y/o homogeneidad de las series de datos, etcétera.
- El pretratamiento o transformación de los datos en función de su normalidad, escalas, unidades y orden de magnitud de cada variable. La solución para esto es trabajar con el logaritmo natural de los datos para corregir su asimetría por no ajuste a la distribución estadística normal y realizar una estandarización de los mismos mediante la tipificación:

$$Y_{ij} = \frac{X_{ij} - \bar{X}_i}{S_i} \quad (1)$$

donde X_{ij} representa el valor de la j -ésima observación de la i -ésima variable; $\bar{X}_i$ es la media de la variable X_{ij} ; S_i representa la desviación estándar de la variable X_{ij} , e Y_{ij} es la representación de la j -ésima observación de la i -ésima variable transformada, para solucionar los problemas de escala, unidades y órdenes de magnitud.

- El análisis primario de las relaciones entre variables u observaciones, que permite no sólo fundamentar la elección del tipo de método a utilizar, sino también colabora en la posterior interpretación de los resultados.

Es importante resaltar que, en general los métodos multivariados sustentan o inician su desarrollo en la denominada matriz de similitud, esto es: una matriz calculada con el conjunto de datos originales (matriz de datos básicos), que expresa las similitudes entre las series de observaciones de cada variable (modo R) o de variables en cada punto de observación (modo Q). Para ello, una adecuada selección del coeficiente utilizado en la valoración de las similitudes presentes en los datos —coeficiente de correlación de Pearson (r), distancia

euclídeana (d), coseno de theta ($\cos \theta$), etcétera—permitirá el reconocimiento de las particularidades, convergencias o relaciones contenidas en la matriz de datos básicos.

Particularmente, el análisis de agrupamiento (Brown, 1998), también conocido como análisis cluster, o simplemente cluster, en modo Q , es una técnica destinada a estudiar las posibles similitudes entre unidades observacionales o puestos de medida. En cambio, su alternativa de aplicación en modo R indaga sobre las vinculaciones entre las variables consideradas. En cualquiera de los casos, modo Q o modo R , esta metodología permite clasificar un conjunto heterogéneo de datos en grupos o asociaciones homogéneas con propiedades características diferenciales, en función de maximizar la distancia o diferencia entre ellos. Es usual la generación de dendrogramas o fenograma, como el de la ilustración 1, para la representación gráfica de las asociaciones sostenidas por este método. A partir de la identificación de los dos grupos principales se continúa con la discriminación de subgrupos u otras asociaciones, cuya vinculación es creciente en términos de similitud. Así, por ejemplo, en la ilustración 1 se presentan dos grupos principales: $A-B-C$ y $D-E$, en donde $A-B$ y $D-E$ constituyen núcleos. En este caso es necesario establecer cuál es el nivel de discriminación que, en términos de similitud, aporta la mejor definición de asociaciones homogéneas; dicho de otro modo, hasta qué orden jerárquico se continúa con el agrupamiento.

Por otra parte, el análisis factorial por componentes principales o análisis de componentes principales (Jöreskog *et al.*, 1976), en modo R y modo Q , permite la reducción del número original de variables o unidades observacionales mediante la definición de nuevas entidades que resultan de una combinación lineal de las variables u observaciones originales. Estas nuevas variables se definen sobre la base del cálculo de autovalores y autovectores contenidos en la matriz de similitud que, a su vez, se obtiene del conjunto de datos

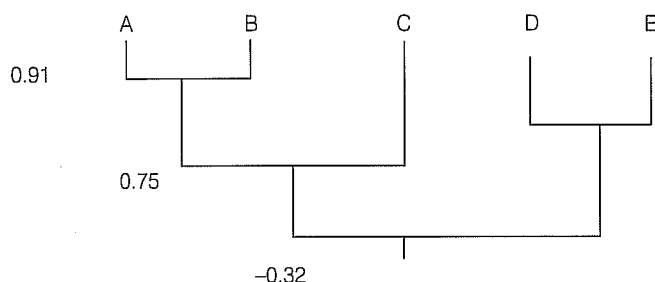
originales. Esta transformación de n variables originales (que presentan un cierto grado de correlación) en otras p nuevas variables incorrelacionadas es, desde el punto de vista matemático, linealmente independiente y define geométricamente un conjunto de componentes ortogonales. En este caso, es necesario establecer cuál es el número p adecuado para representar convenientemente los fenómenos en análisis de una manera parsimoniosa, pero sin perder demasiada información. Para ello suelen considerarse los siguientes criterios: a) evidentemente, como se pretende que $p < n$, la varianza representada por las p nuevas variables, componentes o factores deberá ser menor que la varianza total del conjunto original de n variables; generalmente un 80% es aceptable; b) además, y dado que con esta reducción cada una de las n variables originales tendrá una varianza representada menor que la original, que se denomina comunalidad (Jöreskog *et al.*, 1976), deberá tener un umbral del 80%; c) los autovalores retenidos deberían ser mayores a uno o bien tener una contribución mínima del 10% en la varianza.

De la aplicación de esta metodología se deriva la definición de factores, "scores" o marcas, "loadings" o cargas, que permiten identificar procesos o fenómenos naturales asociados con estas nuevas identidades y el área de afectación de las mismas en donde se mantendría el comportamiento homogéneo de las variables relacionadas con los factores y que precisamente los caracterizan (Paris *et al.*, 1995).

Teniendo en cuenta además, que los datos n -dimensionales pueden ser representados como puntos en un espacio n -dimensional (y del mismo modo con las m observaciones), Andrews (1972) propuso un método simple y útil para representar gráficamente datos multidimensionales en sólo dos dimensiones.

De acuerdo con ello, en cada sitio de registro o punto de observación es posible definir una función armónica en donde los coeficientes que afectan a las funciones seno y coseno, y el término constante son definidos por los valores cuantificados para las n variables. Esto deriva en una expresión como la siguiente:

Ilustración 1. Dendrograma.



$$f_x(t) = \frac{X_1}{\sqrt{2}} + X_2 \sin(t) + X_3 \cos(t) + X_4 \sin(2t) + X_5 \cos(2t) + \dots \quad (2)$$

donde $X_1, X_2, \dots, X_n$ son las variables cuantificadas en cada sitio; t varía en el intervalo $-\pi$ y π . En forma general, la expresión (2) puede escribirse de la siguiente manera:

$$f_x(t) = \frac{X_1}{\sqrt{2}} + \sum_{i=1}^{m1} X_{2i} \sin(it) + \sum_{i=1}^{m2} X_{2i+1} \cos(it) \quad (3)$$

donde:

$$m1 = \frac{n-1}{2} = m2 \quad \text{si } n \text{ es impar} \quad (4)$$

$$m1 = \frac{n}{2} \quad \text{y} \quad m2 = \frac{n}{2} - 1 \quad \text{si } n \text{ es par} \quad (5)$$

Así, considerando los m sitios de medición, es posible obtener un gráfico de las llamadas curvas de Andrews, como el que se presenta en la ilustración 2.

A partir de este gráfico será posible identificar las semejanzas en el comportamiento hidrológico de acuerdo con la similitud morfológica de estas funciones. La curva promedio de cada grupo representará sus características medias a través de sus coeficientes (que son las variables hidrológicas consideradas).

En este trabajo se presenta una propuesta metodológica que involucra la utilización de los métodos estadísticos-matemáticos multivariados anteriormente descritos en forma conjunta para, a partir de una evaluación integrada de los resultados aportados por

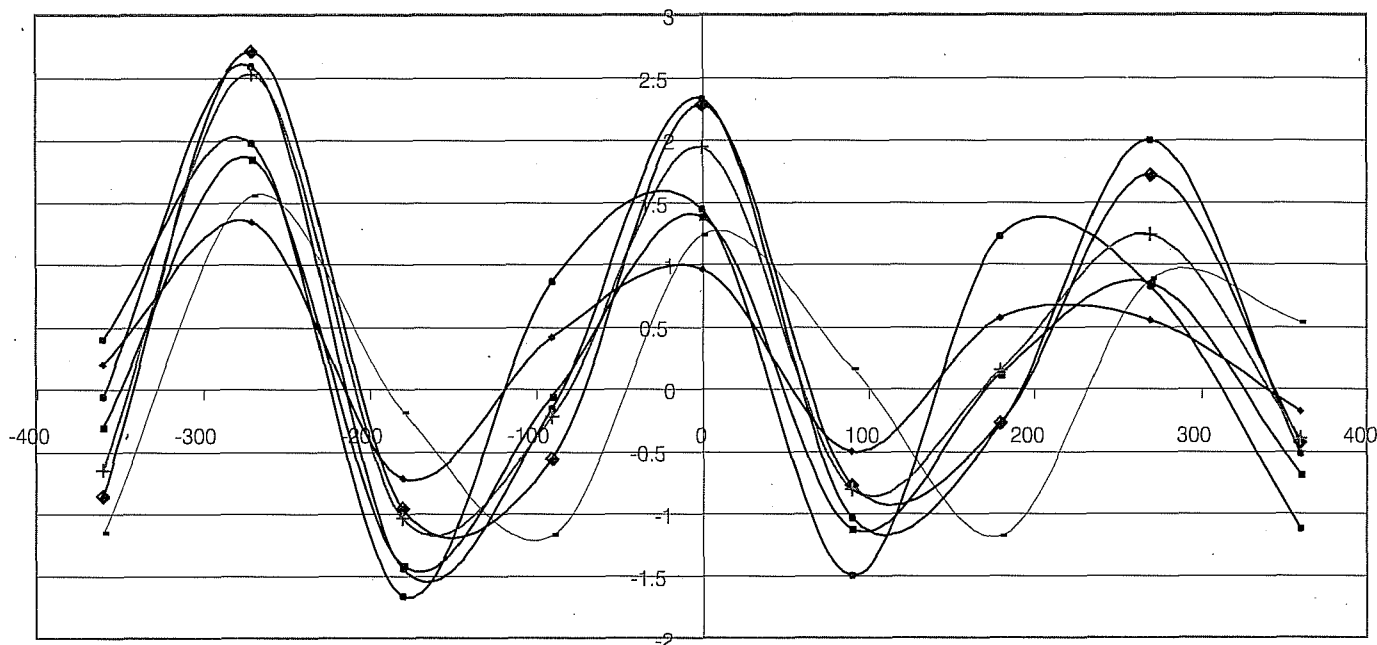
estas técnicas y la caracterización conceptual previa del sistema, definir zonas de comportamiento hidrológico homogéneo e identificar aquellas variables de mayor significación en la expresión funcional del caudal mediano.

Área de aplicación

La ilustración 3 muestra la ubicación del área de aplicación de la propuesta metodológica. Ésta comprende fundamentalmente el territorio de la provincia de Tucumán y porciones menores de las provincias de Catamarca y Santiago del Estero, Argentina. En esta región se analizó un total de 605 crecidas, correspondientes a 11 subcuencas, de las cuales ocho pertenecen a la cuenca del río Salí-Dulce, una a la cuenca del río Pasaje o Salado, una a la Alta Cuenca del Juramento y la restante a la cuenca del Salar de Pipanaco (EVARSA, 1997). El cuadro 1 presenta la ubicación de los puestos de aforo, el área de aporte a estas estaciones, el periodo de registro, etcétera.

Tucumán es una de las provincias argentinas con una mejor distribución espacial de los recursos hídricos. Su geografía presenta dos regiones claramente definidas: llanuras en el este y montañas en el oeste, en las que las sierras del Aconquija forman una barrera climática para los vientos provenientes del sudeste que descargan su humedad en el sector oriental (500 a 2,000 mm de precipitación anual) y dan lugar a una densa red hidrográfica.

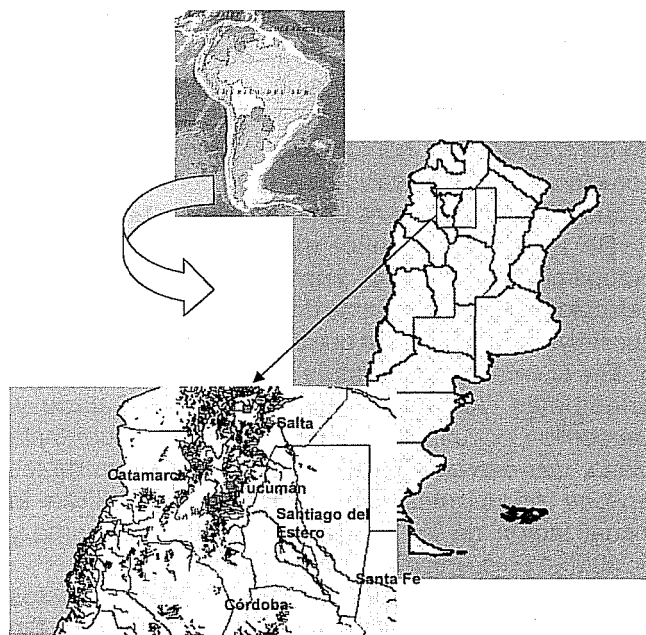
Ilustración 2. Curvas de Andrews.



Cuadro 1. Características de las cuencas analizadas.

Río	Estación	Latitud sur	Longitud oeste	Provincia	Área (km ²)	Periodo	Cantidad crecidas
Membrillo	Las Higueras	27°18	65°42	Tucumán	40	1977-1983	40
Salí	San Vicente	26°37	65°11	Tucumán	3,720	1974-1978	17
Salado	El Arenal	26°13	63°45	Santiago del Estero	40,000	1978-1983	22
Las Cañas	Potrero del Clavillo	27°24	65°59	Tucumán	1,000	1974-1983	23
Solco	Las Higueras	27°18	65°42	Tucumán	140	1974-1983	81
Lules	Potrero de las Tablas	26°21	65°21	Tucumán	600	1973-1983	83
Cochuna	Los Hornitos	27°20	65°55	Tucumán	155	1973-1982	44
Santa María	Pie de Médano	26°59	66°15	Catamarca	4,435	1970-1983	88
Angostura	Tafí del Valle						
	km 52	26°55	65°41	Tucumán	500	1973-1977	10
Los Sosa	Ruta 307						
	km 19	27°06	65°36	Tucumán	620	1974-1983	57
Belén	Playa Larga	27°35	67°00	Catamarca	4,300	1966-1982	140

Ilustración 3. Ubicación del área de aplicación y de los puestos de aforo (fuente: Subsecretaría de Recursos Hídricos de la República Argentina, 2002).



Esta abundancia de agua permitió el asentamiento de poblaciones en la zona del pie de monte con un importante desarrollo agrícola, especialmente caña de azúcar y citrus (Tineo, 1995). El río más importante de la provincia es el Salí, que recibe el aporte de distintos tributarios: Angostura, Lules, Los Sosa, Cochuna, etcétera, y desagua aguas abajo

en el embalse de río Hondo, dando lugar al río Dulce en la provincia de Santiago del Estero. El río Santa María, que nace en Catamarca, atraviesa el sector occidental de Tucumán y llega hasta Salta.

Por su parte, el río Belén es el principal tributario de una extensa red de ríos y arroyos que aportan a un receptor final: el salar de Pipanaco. El área de aporte a esta cuenca extensa cerrada se ubica en su totalidad en la provincia de Catamarca. El río Salado, en cambio, se origina al norte del área de aplicación, en la provincia de Salta, donde se denomina Pasaje o Juramento, y atraviesa diagonalmente el territorio santiagueño para ingresar a la provincia de Santa Fe, donde desemboca en el río Paraná (Cases I Asociatas, 1995).

El cuadro 2 reúne algunas características de las cuencas analizadas: nombre de la cuenca, el área de drenaje A (km²) y los valores máximo, mínimo y medio de caudal Q_i (m³/s), velocidad v (m/s), profundidad h (m) y ancho w (m) para cada uno de los ríos.

Tal como se observa en la ilustración 4, la respuesta hidrológica de las cuencas en estudio es muy diferente, al igual que las áreas de aporte presentadas anteriormente en el cuadro 2. Investigaciones anteriores (Zucarelli y Ceirano, 2002) han permitido establecer: a) un resumen de las relaciones hidráulicas encontradas en estos cursos, dado que la forma y el tamaño de las secciones transversales de los canales aluviales se relacionan con los valores de caudales y, b) conforme a los valores obtenidos para el coeficiente K_{go} (ecuación de Glushkov) (Farías, 2000) se han podido asociar los ríos por las características de sus lechos desde rocosos (K_{go} cercano a 1.4; Las Cañas, Lules, Cochuna, Belén) hasta arenosos

Cuadro 2. Características de las cuencas analizadas.

Río	Caudal (m ³ /s)	Velocidad (m/s)	Profundidad (m)	Ancho (m)
	Máximo Mínimo Medio	Máximo Mínimo Medio	Máximo Mínimo Medio	Máximo Mínimo Medio
Membrillo	88.97 2.49 31.73	6.04 0.44 2.93	1.21 0.09 0.80	28.00 5.00 18.33
Salí	338.16 9.77 132.17	3.29 0.84 2.55	1.86 0.42 1.05	58.00 25.00 43.55
Salado	165.57 23.44 73.15	2.46 0.64 1.49	1.43 0.70 0.98	60.00 32.00 49.90
Las Cañas	313.73 12.15 110.82	4.56 0.73 1.91	6.67 1.19 3.13	18.00 13.00 15.52
Solco	175.93 11.93 66.72	7.50 1.67 3.72	2.45 0.58 1.13	49.00 18.00 26.96
Lules	208.73 10.61 55.93	2.19 0.48 1.00	6.67 1.23 3.98	28.00 16.00 20.03
Cochuna	208.42 3.43 27.63	1.48 0.33 0.60	7.14 0.89 3.72	52.00 10.00 18.98
Santa María	144.94 0.11 20.97	1.36 0.09 0.33	4.00 0.74 1.55	51.00 2.50 26.21
Angostura	34.51 7.80 17.23	0.91 0.31 0.59	2.41 1.24 1.76	16.00 14.00 15.11
Los Sosa	168.53 11.83 47.74	5.95 1.12 2.19	1.81 0.53 0.88	36.08 14.00 24.66
Belén	763.31 0.92 77.04	2.43 0.12 0.52	10.00 1.16 3.49	52.00 2.50 29.56

altamente erosionables (K_{go} aproximadamente igual a 6.5; Membrillo, Salí, Salado, Los Sosa). Merece recordar que la fórmula de Glushkov es una ecuación morfológica que establece una relación entre el ancho w y la profundidad h en un cauce estable. Esta relación, cuyo carácter inicial es puramente empírico, fue establecida sobre la base de observaciones llevadas a cabo en numerosos ríos situados en territorio de la antigua Unión Soviética durante el primer cuarto del siglo pasado y su uso se ha generalizado a lo largo de los años en diversos países de la región oriental de Europa (Farías, 2000).

$$K_{go} = \frac{w^{1/2}}{h} \text{ ecuación de Glushkov} \quad (6)$$

Resultados

En función de la escasa información disponible, la interrupción o discontinuidad en las mediciones y de la necesidad de optimizar mediciones futuras, la evaluación de las posibles similitudes con vistas a la definición de regiones hidrológicamente homogéneas colabora con la definición estratégica de aquellos puestos de aforo que permitan obtener estimaciones representativas del caudal medio, a fin de cuantificar la disponibilidad en los cursos superficiales.

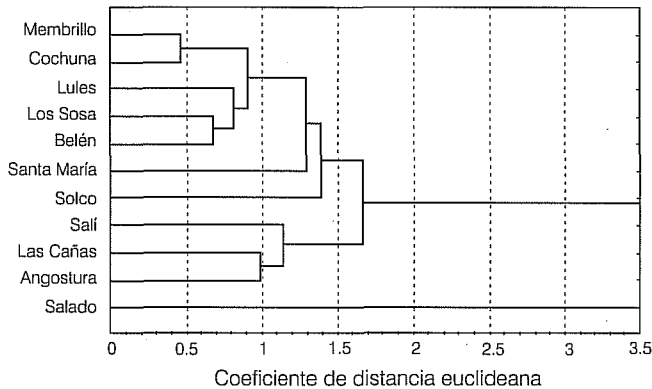
De este modo, la identificación de regiones hidrológicamente homogéneas se hizo sobre la base de los valores medios de las variables que caracterizan las secciones aforadas correspondientes a las cuencas de aporte superficial: área de drenaje A (km²), caudal Q (m³/s), velocidad v (m/s), profundidad h (m) y ancho de la sección transversal w (m). Los datos originales fueron transformados únicamente mediante una estandarización.

La ilustración 5 muestra las curvas de Andrews para cada uno de los puestos de observación disponibles en el área de estudio. Sin embargo, en ella no es posible realizar una identificación de unidades observacionales de comportamiento homogéneo diferencial tan directa como era de esperar, salvo en el caso del río Salado, por lo que su utilización no es tan simple como pareciera.

De acuerdo con ello, mediante el análisis de agrupamiento en modo Q , a partir de la matriz de similitud calculada con coeficiente de distancia euclídeana (d), se construye el dendrograma que se muestra en la ilustración 6. En el mismo es posible distinguir dos grupos principales: Angostura-Las Cañas-Salí-Salado y, Membrillo-Cochuna-Belén-Los Sosa-Lules-Solco-Santa María. Las curvas de Andrews para cada uno de los grupos principales (ilustración 7) posibilitan la visualización de los resultados del dendrograma y colaboran en su interpretación.

Como puede observarse en la ilustración 7, estos grupos tienen curvas con amplitudes diferentes. Sin embargo, ello constituye una primera aproximación en la identificación de homogeneidades, ya que no aporta información relativa a las características propias de cada grupo (caudal, velocidad, áreas, etcétera) ni tampoco es posible evidenciar con claridad la existencia de similitudes importantes dentro de cada uno de los grupos principales definidos por el método de agrupamiento.

Ilustración 6. Análisis cluster modo Q.



principales. La componente I (varianza=68.74%, relacionada con los mayores valores de caudal, velocidad y tirante) está integrada por los puestos correspondientes a los ríos Membrillo, Salí, Las Cañas, Solco, Lules, Cochuna, Santa María, Angostura, Los Sosa y Belén, con marcas positivas. La componente II (varianza=18.29%, relacionada con las variables área y ancho) está integrada por las estaciones de los ríos Membrillo, Solco, Lules, Cochuna, Santa María, Los Sosa y Belén, todos con marcas positivas, y las correspondientes a los ríos Salí, Salado, Las Cañas y Angostura, con marcas negativas. La presencia de marcas positivas en las dos componentes se interpreta en relación con las variables que definen cada factor; así, por ejemplo, la sección de control del río

Ilustración 7. Curvas de Andrews para los dos grupos principales.

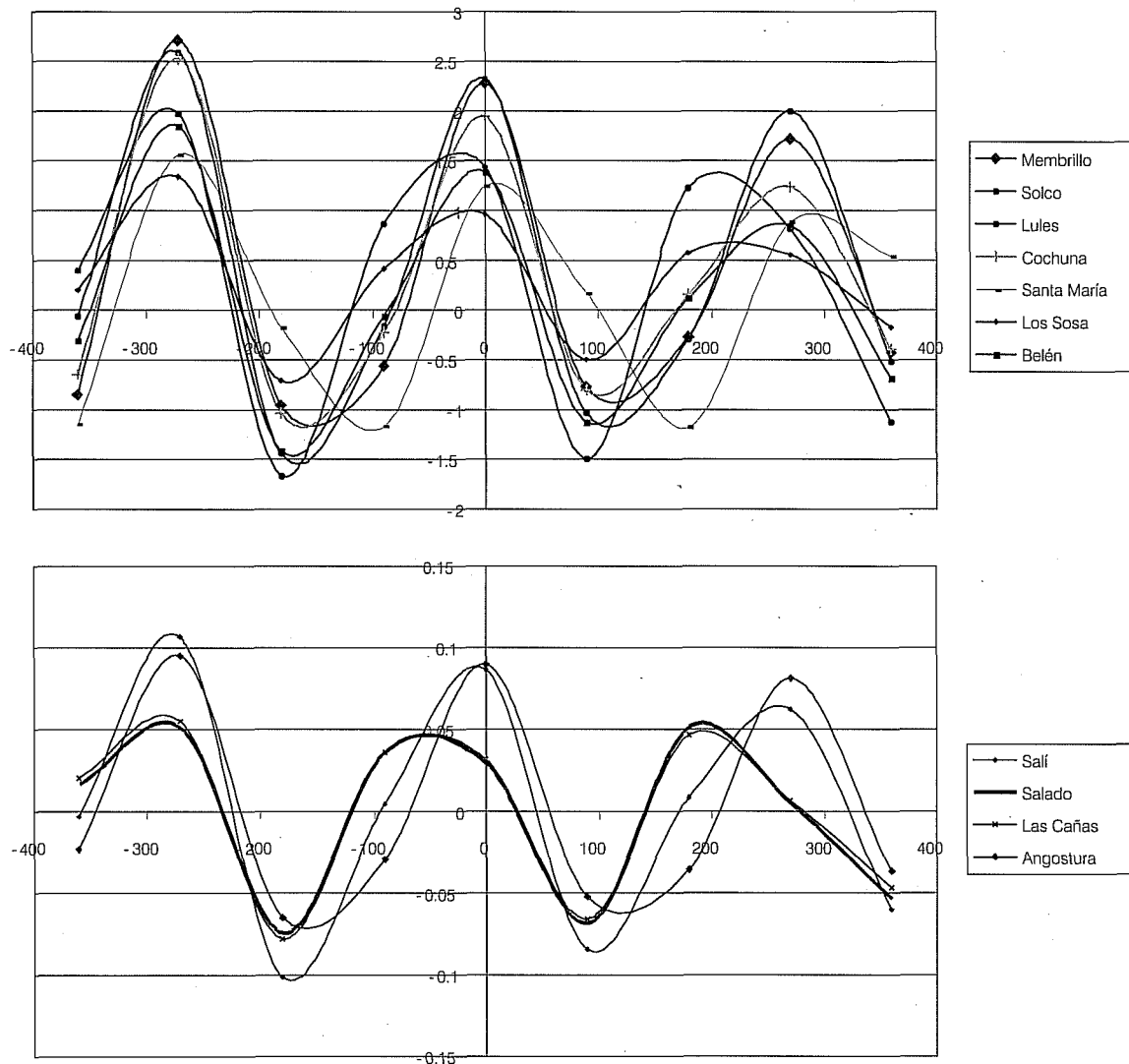


Ilustración 4. Caudales.

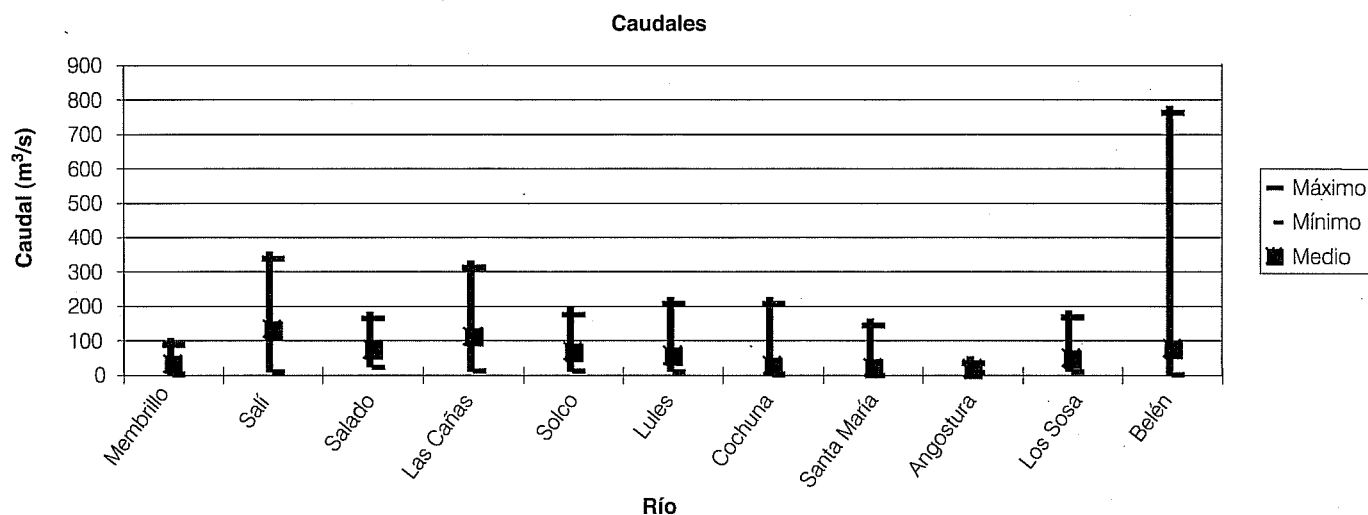
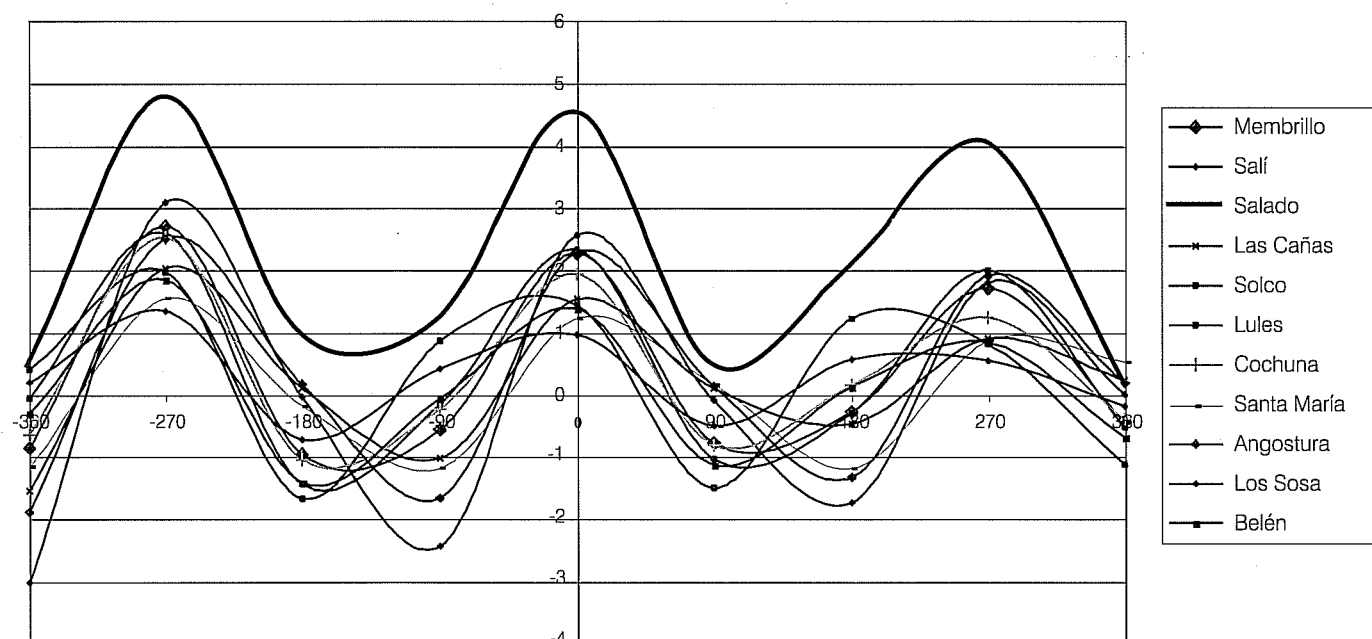


Ilustración 5. Curvas de Andrews para el área de aplicación.



Teniendo en cuenta, además, que estas curvas representan funciones armónicas, cuyos coeficientes están definidos por los valores cuantificados para las n variables, y que el orden de las variables en la función tiene notable influencia en los resultados obtenidos (Embrechts y Herzberg, 1991), se realiza el análisis de componentes principales en modo R , como metodología complementaria para contar con otra evidencia sobre la definición de los propios grupos homogéneos a través

de las marcas y establecer el orden más adecuado de las variables en la expresión de Andrews mediante el análisis de las cargas. En este caso, la matriz de similitud fue cuantificada con el coeficiente coseno de theta ($\cos \theta$). La mejor relación con el caudal la ofrece el ancho de la sección. El cuadro 3 presenta los resultados obtenidos.

Así, con una representatividad del 87% de la varianza total del conjunto de datos y una comunalidad mínima de 80% es posible diferenciar dos componentes o factores

Cuadro 3. Análisis de componentes principales modo R. Resultados.

Matriz de similitud – coeficiente cos θ

	Área	Caudal	Velocidad	Profundidad	Ancho
Área	1.00	0.32	0.30	0.53	0.80
Caudal	0.32	1.00	0.69	0.67	0.75
Velocidad	0.30	0.67	1.00	0.77	0.48
Profundidad	0.53	0.67	0.77	1.00	0.69
Ancho	0.80	0.75	0.48	0.69	1.00

Autovalor (orden)	Autovalor (valor)	Varianza	Varianza acumulada
1	3.4375	68.7493	68.7493
2	0.9147	18.2932	87.0425
3	0.4296	8.5914	95.6339
4	0.1892	3.7835	99.4174
5	0.0291	0.5825	100.000

Variable	Comunalidad (%)
Área	94
Caudal	80
Velocidad	86
Profundidad	82
Ancho	93

Cargas

Variable	Factor I	Factor II
Área	0.12	-0.96
Caudal	0.84	-0.29
Velocidad	0.92	-0.11
Profundidad	0.79	-0.43
Ancho	0.49	-0.82

Marcas

Río	Factor I	Factor II
Membrillo	1.18	1.27
Salí	1.07	-7.39
Salado	-0.41	-1.30
Las Cañas	0.61	-5.32
Solco	1.53	7.48
Lules	0.85	2.44
Cochuna	1.17	0.41
Santa María	1.18	0.72
Angostura	0.77	-2.83
Los Sosa	0.68	1.81
Belén	1.18	2.51

Solco registra importantes caudales y velocidades (componente I), pero también grandes anchos y área pequeña (componente II). Del mismo modo, marcas positivas en una componente y negativas en el otro factor merecen una consideración análoga. Tal es la sección sobre el río Salí, con gran caudal y área. La ilustración 8 muestra las curvas de Andrews correspondientes a estas dos componentes principales, considerando para la componente II las asociaciones definidas por las marcas positivas y negativas.

Debe notarse que estos grupos son idénticos a los obtenidos por el análisis de agrupamiento modo Q, pero como se introdujo el reordenamiento de las variables en la expresión de las curvas dado por los valores absolutos de las cargas obtenidas (según cuadro 3: área, velocidad, caudal, ancho, profundidad), esta nueva graficación refleja diferencias más marcadas entre las amplitudes y frecuencias en el trazado de las curvas de Andrews.

Posteriormente, con el objeto de obtener una mejor diferenciación entre los grupos, se realizó el análisis de componentes principales, pero en modo Q. El cuadro 4 muestra los resultados obtenidos. Considerando tres factores como componentes principales, la varianza total representada por los mismos es de 94.84%, conservando un umbral de comunalidad del 86%. La ilustración 9 muestra las asociaciones obtenidas de acuerdo con la magnitud de las cargas en valor absoluto para cada factor:

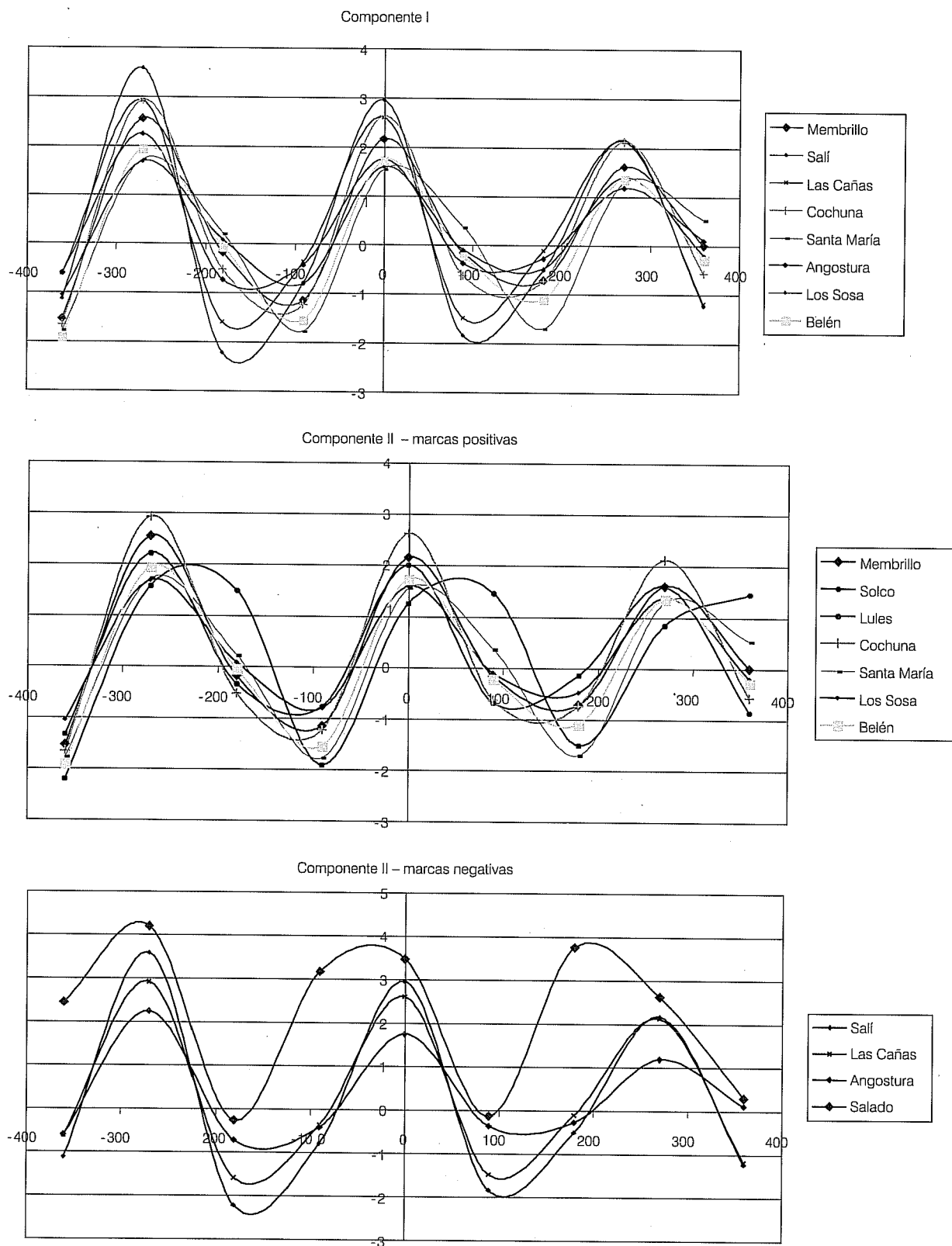
- Factor I – ríos Membrillo, Solco, Lules, Cochuna, Los Sosa, Belén.
- Factor II – ríos Salí, Las Cañas, Santa María, Angostura.
- Factor III – río Salado.

Debe notarse en este caso que la diferencia entre amplitudes (ordenadas) y formas en las curvas de los gráficos de la ilustración 9 son mayores incluso que en la ilustración 8.

De acuerdo con esta conformación de regiones homogéneas, las secciones de aforo correspondientes a los ríos Membrillo, Solco, Lules, Cochuna, Los Sosa y Belén tienen un comportamiento hidrológico semejante, de modo que sería posible la transferencia de información entre las mismas. Igualmente ocurre con las estaciones de aforo de los ríos Salí, Las Cañas, Santa María y Angostura.

A fin de validar tales resultados se analizan los hidrogramas medios para dichos puestos de aforo (ilustración 10). Como se observa en estos gráficos, las asociaciones obtenidas no serían válidas, pero es importante tener en cuenta la representatividad de tales

Ilustración 8. Curvas de Andrews. Análisis de componentes principales modo R.



Cuadro 4. Análisis de componentes principales modo Q. Resultados.

Autovalor (orden)	Autovalor (valor)	Varianza	Varianza acumulada
1	8.30	75.53	75.53
2	1.15	10.43	85.96
3	0.98	8.89	94.84
4	0.47	4.22	99.07
5	0.10	0.92	99.99
del 6 al 11	< 0.001	< 0.01	100.00

Río	Comunalidad (%)
Membrillo	99
Salí	97
Salado	86
Las Cañas	94
Solco	86
Lules	95
Cochuna	99
Santa María	96
Angostura	96
Los Sosa	96
Belén	98

Cargas

Río	Factor I	Factor II	Factor III
Membrillo	0.72	-0.60	-0.31
Salí	0.29	-0.89	-0.29
Salado	0.13	-0.27	-0.87
Las Cañas	0.31	-0.79	-0.46
Solco	0.88	-0.29	0.04
Lules	0.78	-0.16	-0.55
Cochuna	0.72	-0.55	-0.39
Santa María	0.63	-0.73	0.18
Angostura	0.33	-0.88	-0.25
Los Sosa	0.83	-0.31	-0.41
Belén	0.88	-0.40	-0.18

hidrogramas medios en relación con los valores extremos (caudales máximos y mínimos medios) observados (ilustración 11). Sin embargo, las curvas de permanencia registradas en dichas estaciones, de acuerdo con esta conformación de regiones, sí reflejan órdenes de magnitud de los caudales de cada grupo y formas homogéneas en las asociaciones anteriormente mencionadas (ilustración 12).

Conclusiones y recomendaciones

Los resultados alcanzados en estas subcuencas del noroeste argentino con varios métodos estadísticos-matemáticos y su evaluación integrada con la caracterización conceptual previa del sistema han permitido identificar regiones hidrológicas homogéneas en forma consistente y corroborar la influencia que tiene el ordenamiento de variables en la expresión de Andrews, obteniendo mejoras significativas en la representación gráfica cuando se consideraron las vinculaciones existentes entre el caudal y el resto de las variables consideradas (en donde el ancho de la sección contabilizó el mayor coeficiente de similitud).

Se destaca la importancia en la utilización conjunta de métodos conocidos y aplicados individualmente que, siguiendo un proceso convergente, dan lugar a la propuesta metodológica presentada en este trabajo: el análisis de agrupamiento en modo Q para lograr una primera aproximación en la clasificación del conjunto de datos y evaluación de sus similitudes —es decir, como análisis primario de las relaciones entre las observaciones disponibles—, análisis de componentes principales en modo R como metodología complementaria que colabore en la definición de los propios grupos homogéneos a través de las marcas, y establecer el orden más adecuado de las variables en la expresión de Andrews mediante el análisis de las cargas y posteriormente el análisis de componentes principales en modo Q para una mejor delineación de los grupos. Todo ello sobre la base de una adecuada selección de variables, el pretratamiento de los datos, la adecuada selección de los coeficientes de similitud utilizados, la caracterización previa del sistema en estudio y la elaboración paralela de las curvas de Andrews para la evaluación de los resultados obtenidos con cada una de estas técnicas.

En las regiones hidrológicas así definidas, las curvas de permanencia reflejan órdenes de magnitud de los caudales y formas homogéneas, evidenciándose además la no representatividad de los hidrogramas medios en la caracterización de las regiones, dado el amplio rango registrado entre los caudales máximos y mínimos medios.

En relación con la constancia u homogeneidad de las características hidrológicas en las regiones, y teniendo presente que sólo tres de las cuencas continúan siendo aforadas en la actualidad, se propone una recomendación para que los organismos encargados de operar la red de mediciones hidrológicas mantengan la continuidad de registro en al menos una estación por región hidrológica definida. Esto es de fundamental importancia, ya que no sólo permitirá contar con estimaciones regionalizadas de caudales en esta zona del país, en donde se torna

Ilustración 9. Curvas de Andrews. Análisis de componentes principales modo Q.

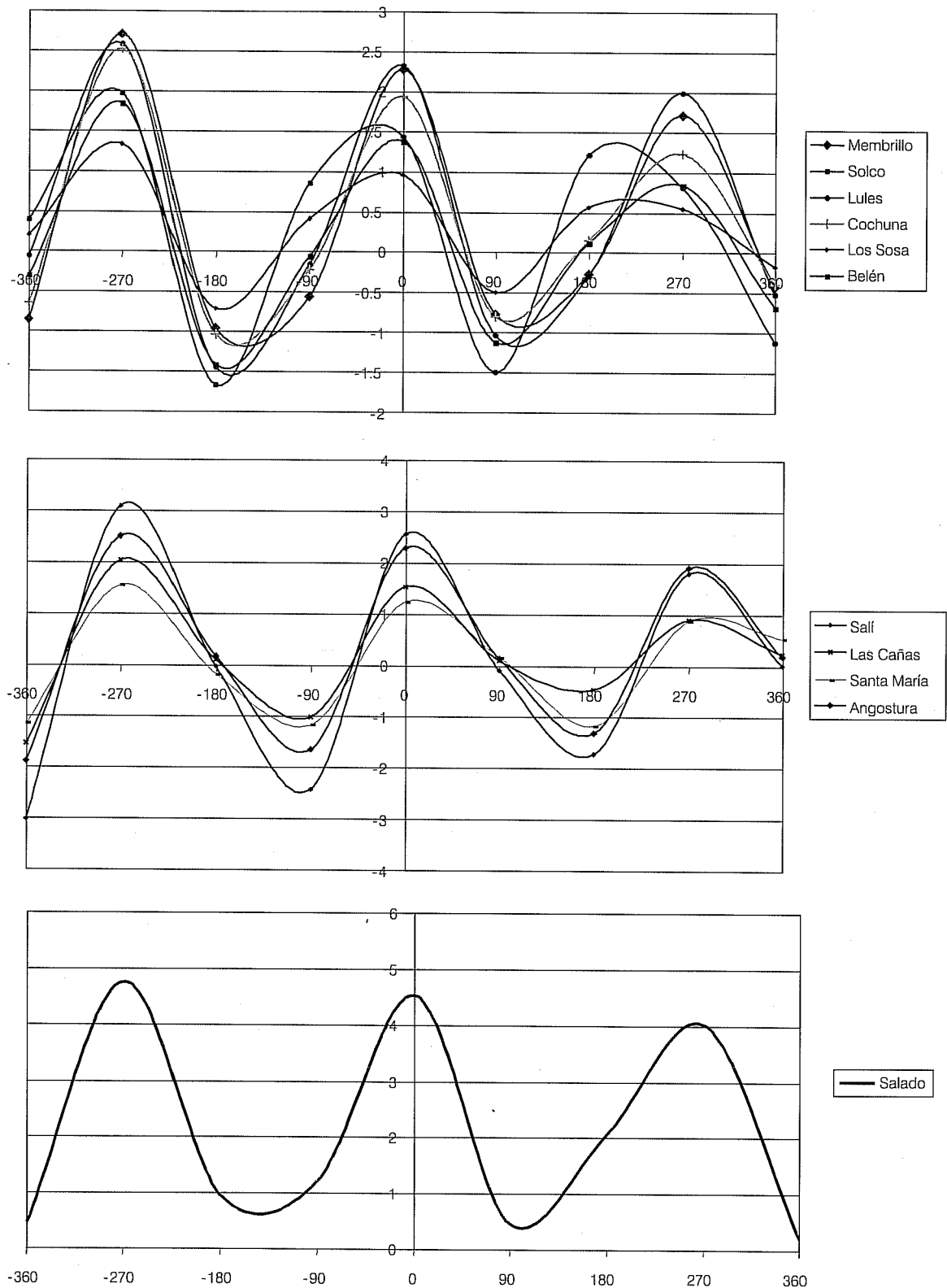


Ilustración 10. Hidrogramas medios para las estaciones de cada región.

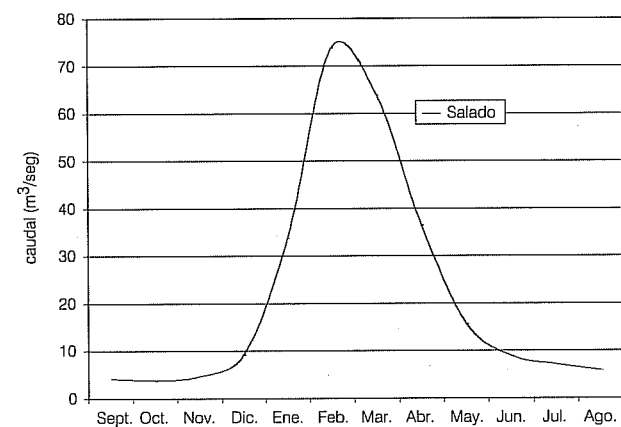
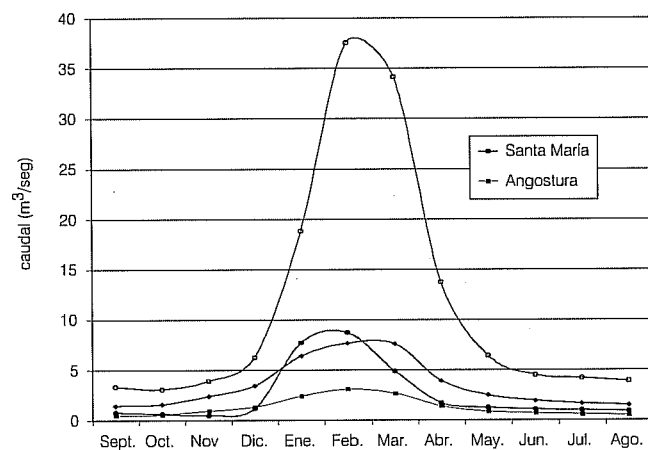
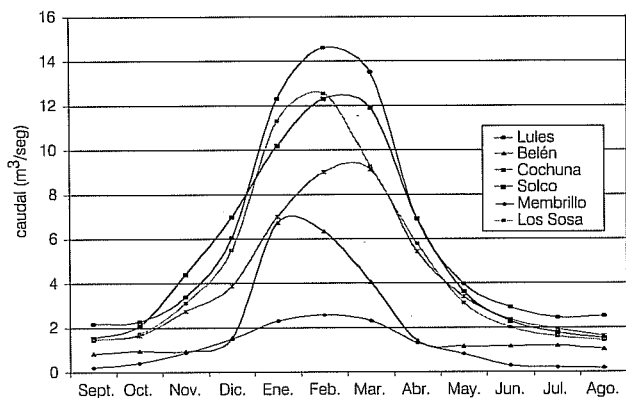


Ilustración 11. Hidrogramas medios para la estaciones Lules, Los Sosa y Las Cañas.

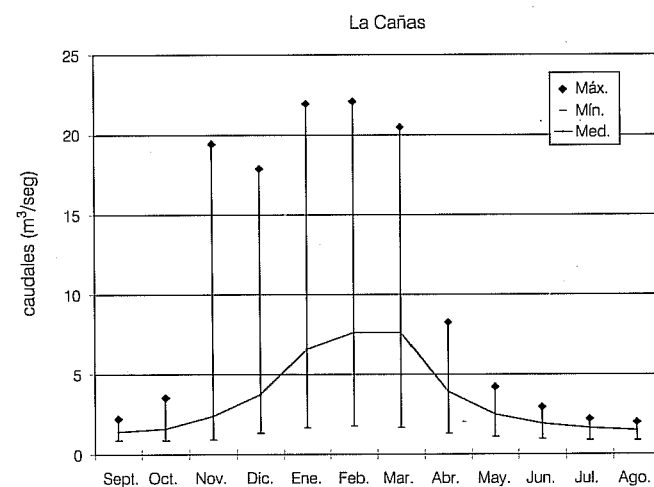
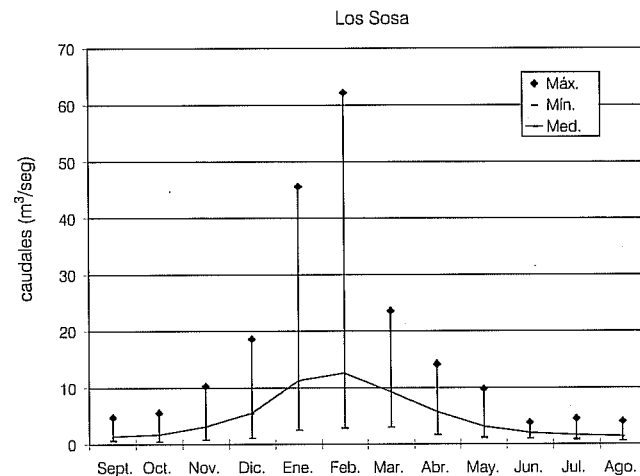
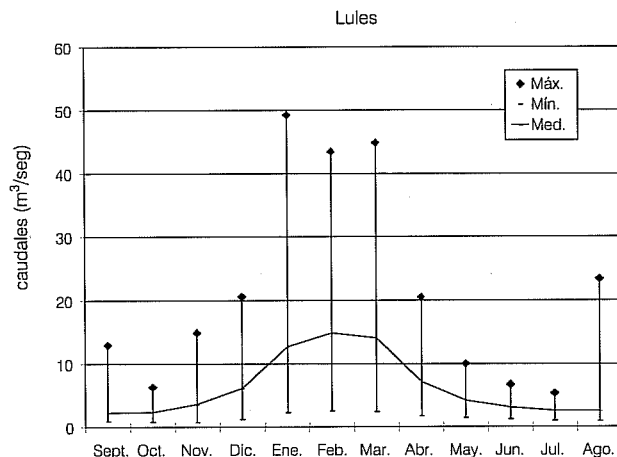
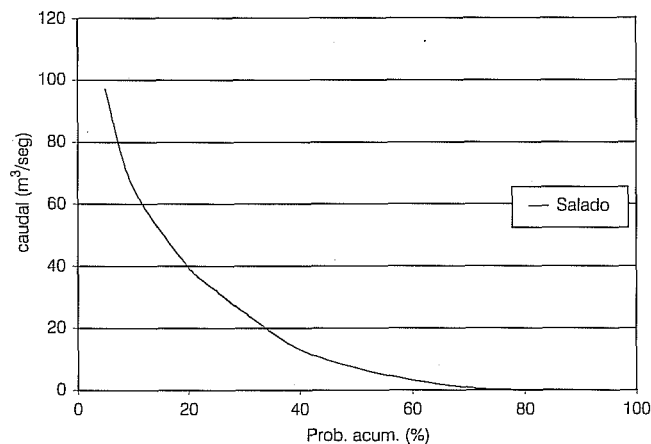
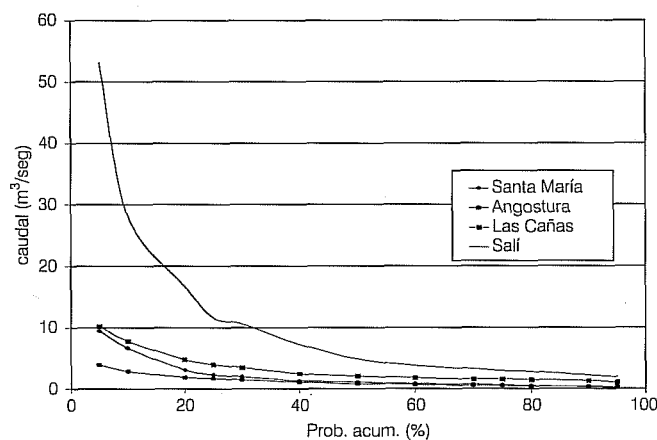
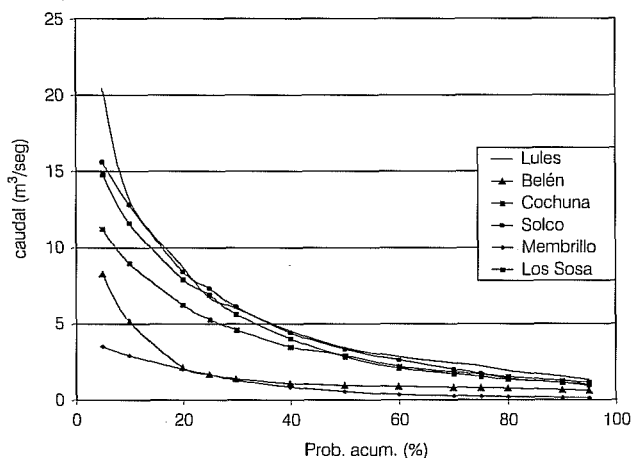


Ilustración 12. Curvas de permanencia para las estaciones de cada región.



indispensable realizar un manejo eficiente de los recursos hídricos, sino que a la vez suministrará datos básicos para la estimación y el seguimiento de crecidas en las zonas de descarga de estos ríos, donde son frecuentes situaciones de inundaciones y anegamientos.

A la luz de los resultados obtenidos, se propone dar continuidad las investigaciones mediante la aplicación de esta propuesta metodológica para la identificación de regiones hidrológicas homogéneas en otras zonas o por incorporación de otras variables (precipitación, densidad de drenaje, etcétera) que contribuyan a los fines de la delimitación, considerando, asimismo, la posibilidad de sólo usar un periodo común en función de la disponibilidad de registros.

Agradecimientos

Al personal de la región del noroeste argentino (NOA) de la Empresa EVARSA por facilitar los datos básicos para la realización del presente trabajo.

Recibido: 03/09/2003
Aprobado: 16/12/2003

Referencias

- ANDREWS, D. Plots of high-dimensional data. *Biometrics*. Vol. 28, 1972, pp. 125-136.
- ANDRADE, E. y HAWKINS, R. Aplicação da função de Andrews na avaliação da regionalização de bacias em regiões áridas e semi-áridas. *Revista Brasileira de Recursos Hídricos*. Vol. 5, núm. 4, octubre, 2000, pp. 17-24.
- BROWN, C. *Applied multivariate statistic in geohydrologic and related sciences*. Springer Verlag, 1998, 248 pp.
- CASES I ASOCIATAS, S.A. *Mi país, la Argentina*. Buenos Aires: Editorial Arte Gráfico, Editorial Argentina, S.A., ISBN 84-599-3442-X, 1995.
- EMBRECHTS, P. y HERZBERG, A. Variations of Andrews' plots. *International Statistical Review*. Vol. 59, núm. 2, 1991, pp. 175-194.
- EVARSA. *Estadística hidrológica*. Tomos I y II. Presidencia de la Nación. Secretaría de Recursos Naturales y Desarrollo Sustentables. República Argentina, 1997.
- FARÍAS, H. *Formulación racional de la relación ancho-profundidad para cauces aluviales estables*. Congreso Nacional del Agua 2000, Santiago del Estero, Argentina, 2000.
- FRANCHINI, M. y SUPPO, M. Regional analysis of flow duration curves for a limestone region. *Water resource management*. Vol. 10, 1996, pp. 199-218.
- JÖRESKOG, K., KLOVAN, J. y REYMENT, R. *Geological factor analysis*. Amsterdam: Elsevier Scientific Publishing Company, 1976.

PARIS, M., TUJCHNEIDER, O., FILI, M., PÉREZ, M. y D'ELIA, M.
Hidrogeoquímica de las aguas subterráneas en el área de Villa María Grande (provincia de Entre Ríos, República Argentina). VI Simposio de Hidrogeología. Sevilla, España, 23 al 27 de octubre de 1995. Libro de la colección "Hidrogeología y recursos hidráulicos". Tomo XX. Sección 4: *Geoquímica de las aguas subterráneas*. 1995, pp. 221-234.
SUBSECRETARÍA DE RECURSOS HÍDRICOS DE LA REPÚBLICA ARGENTINA. *Atlas digital de los recursos*

hídricos superficiales de la República Argentina. 2002.
TINEO, A. *90 años de investigación hidrogeológica en Tucumán*. *Correlación Geológica*. Núm. 11, ISSN: 217428, 1995, pp. 289-294.
TUCCI, C. *Hidrología*. Porto Alegre, Brasil: Universidad Federal do Rio Grande do Sul Editora, 1987, 942 pp.
ZUCARELLI, G. y CEIRANO, E. Relaciones hidráulicas en cauces de cuencas del noroeste Argentino. *Revista del CURIHAM*. Vol. 8, núm. 2, 2002 pp. 136-148.

Abstract

PARIS, M. DEL C. & ZUCARELLI, G.V. Flow regionalization. An approach to identify homogeneous regions. Hydraulic engineering in Mexico (in Spanish). Vol. XIX, no. 4, October-December, 2004, pp. 5-19.

The spatial variability of the flow conditions, which often exists in the watershed, needs a complementation of the methodologies to estimate the obtained information from punctual data to take into account its representativity. This leads to define zones where certain constancy of the hydrological characteristics is maintained. Nevertheless, in many cases it is not trivial to determine a criterion (or criteria) to establish these regions. The obtained results by application of multivariate statistical-mathematical methods (cluster and principal components analysis and Andrews' plots) to identify homogeneous regions are presented here. This approach is an alternative that considers both the multidimensional data analysis and their integrated evaluation with the system conceptual characterization. The application area comprises 11 basins in the northwest of Argentina, which have historical records of 605 floods. However, in this country zone, with an important drought where it is very important to do an efficient water resources management, only 3 of these basin are nowadays gauging. The homogeneous regions identification resulting has allowed to define strategies to the network measurement design which considers the representativity of the observations and their potentially to transfer information.

Keywords: homogeneous regions, regionalization, cluster analysis, principal components, Andrews' Plot, Argentina.

Dirección institucional de los autores:

Ing. Marta del Carmen Paris
M. en I. Graciela Viviana Zucarelli

Universidad Nacional del Litoral,
Facultad de Ingeniería y Ciencias Hídricas,
Ciudad Universitaria,
Casilla de correo 217-(3000) Santa Fe, Argentina,
teléfono: 0054 342 4575 244 (internos) 150 y 167,
fax: 0054 342 4575 224,
martaparis@gigared.com, mparis@fich.unl.edu.ar,
zuca@fich1.unl.edu.ar