

PROPUESTA METODOLÓGICA PARA LA DETERMINACIÓN DE INHOMOGENEIDADES RELATIVAS EN LAS SERIES DE OBSERVACIONES

Leoncio GARCÍA-BARRÓN* y M^a Fernanda PITA**

* *Departamento de Física Aplicada II*

** *Departamento de Geografía Física y A.G.R. Universidad de Sevilla*

RESUMEN

El análisis de la estructura y evolución de las series temporales requiere un estudio previo de la calidad de los registros de observaciones, por lo que es preciso realizar pruebas sobre su homogeneidad interna. El método propuesto para detectar inhomogeneidades relativas se basa directamente en medidas de correlación entre periodos simultáneos del conjunto de observatorios objeto de estudio. El propio método proporciona mecanismos para efectuar la homogeneización.

Palabras claves: Calidad de los datos, Test de homogeneidad.

ABSTRACT

To analyse the structure and evolution of time series, a previous examination of the data quality is required. We propose a new method to identify relative inhomogeneities in data series, based upon correlation coefficients existing between them through successive simultaneous periods. The method provides mechanisms to homogeneize the series as well.

Key words: Data quality, Homogeneity test.

1. INTRODUCCIÓN

La fuente de investigación de las series climáticas son los registros de las variables meteorológicas realizados a lo largo de sucesivos años. Si las circunstancias durante la época de observación han obligado a efectuar cambios en la localización de los aparatos de medida, del instrumental, y/o del procedimiento, o el entorno del observatorio ha sufrido modificaciones, la calidad de la serie de observaciones puede verse afectada. A ello se une la existencia de “lagunas” en lapsos en que no se produjeron observaciones, o, al menos, no se conservan.

Como consecuencia de todo ello, previamente al análisis climático de las series de observaciones conviene comprobar la calidad de las mismas. En general, cada de una de ellas es sometida a prue-

bas de homogeneidad interna, por aplicación de algunos de los test específicos: Thom, Von Neumann, etc. (RODRÍGUEZ BARRERA *et al.*, 1996). De forma complementaria es posible someter a las series a un análisis comparado de homogeneidad relativa, en relación con los valores obtenidos en otros observatorios que se aceptan como de referencia. Los procedimientos más sofisticados para detectar inhomogeneidades relativas están frecuentemente basados en los métodos de Alexandersson (ALEXANDERSSON, 1986, 1997), o modificaciones sobre los mismos, (ALMARZA y LÓPEZ, 1994; AGUILAR *et al.*, 1999) y por tanto, precisan de la comprobación de los requisitos previos en la estructura de las series (normalidad, ...), exigiendo además cálculos relativamente prolijos. Existen, por otro lado, métodos más sencillos de detección de inhomogeneidades, tales como el test del cúmulo de dobles masas, el test de las ratios etc., pero que no siempre producen resultados satisfactorios y suficientemente concluyentes ni permiten derivar de ellos un proceso de homogeneización consecuente.

Para el análisis de las series térmicas y pluviométricas de larga duración en el suroeste de España hemos utilizado un método intermedio entre ambos (GARCÍA-BARRÓN, 2000), directamente basado en medidas de correlación entre periodos simultáneos del conjunto de observatorios objeto de estudio y que, en nuestra opinión, subsana estas carencias.

2. PROPUESTA METODOLÓGICA

El método se fundamenta en la premisa de que es factible identificar y, en su caso, corregir inhomogeneidades de series meteorológicas por contraste entre periodos simultáneos de distintos observatorios, con la condición de estar globalmente bien correlacionadas. Admitamos *a priori* que todas las series son aceptablemente homogéneas, aunque ninguna con carácter absoluto y definitivo, y que las posibles inhomogeneidades provocadas artificialmente han sido fortuitas, por lo que no tienen que coincidir en el mismo instante en distintos observatorios (esto resulta especialmente evidente cuando los observatorios tienen distinta titularidad, como es nuestro caso). Bajo esta hipótesis, valores excesivamente bajos en el coeficiente de correlación de periodos simultáneos de distintos observatorios indicarían alguna discordancia entre algunos de sus elementos correspondientes.

Consecuentemente con estas premisas, el sistema que empleamos para identificar las discordancias de las series es tomar tramos móviles simultáneos de n elementos para calcular el coeficiente de correlación, y avanzar sucesivamente a lo largo de tales series. Al desplazar el tramo progresivamente en el tiempo, se genera, por tanto, para cada pareja de series originales de observaciones, una nueva serie -reducida en n elementos- de coeficientes de correlación. Se trataría, pues, de calcular para cada par de observatorios la serie de coeficientes de correlación móviles para periodos de n años.

Una disminución de los valores del coeficiente indicaría que los últimos elementos incorporados a tal tramo presentaban discrepancia en su evolución temporal, reflejando consecuentemente la existencia de una inhomogeneidad en alguna de las series. La discordancia se calificaría como inhomogeneidad en los casos en que la caída del coeficiente fuera superior a determinado valor y en valor absoluto no sobrepasara un mínimo prefijado, o bien cuando se registrara un decremento

gradual en el mismo. Lógicamente, una vez detectada la inhomogeneidad, y como paso previo a su corrección, se procedería a comprobar la naturaleza de dicha inhomogeneidad y su posible origen a partir del análisis de la información disponible acerca del observatorio, el entorno del mismo y el proceso seguido por la información hasta llegar a nuestras manos.

Para verificar la adecuación del método a los fines que persigue, hemos elaborado un ejemplo simple, consistente en tomar al azar una serie de cincuenta elementos (a_i) –en este caso, calificaciones de un ejercicio de alumnos- que denominaremos serie A. A partir de ella se elabora una nueva serie B, en la que cada uno de los términos se obtiene como combinación lineal $b_i = 5a_i + 2a_{i+1}$ (los coeficientes se han elegido para que la correlación entre las series A y B sea del mismo orden que la que presentan las series meteorológicas entre distintos observatorios). La serie B, además, sufre una doble manipulación: a cada uno de los elementos entre b_{21} y b_{30} se le suma una cantidad constante de 5 unidades (simularía un posible cambio de aparato de observación, o de su emplazamiento, que condujera a un aumento constante en el valor de las observaciones); a partir del elemento b_{41} se incrementa progresivamente una unidad hasta el b_{50} (pretende simular la aparición de una tendencia creciente en alguna serie de observación). A la serie resultante final le denominamos serie C.

Calculamos los coeficientes de correlación móviles para cinco pares de valores entre las series A y C, cuya homogeneidad deseamos comprobar ($\{a_{1,5}; c_{1,5}\}$; $\{a_{2,6}; c_{2,6}\}$, ..., $\{a_{46,50}; c_{46,50}\}$) y obtenemos una nueva serie de coeficientes de correlación que nos permite discernir la existencia de las inhomogeneidades introducidas (ver figura 1). Los valores del coeficiente de correlación son superiores a 0,8 hasta el elemento 21, en el cual se produce un descenso hasta valores inferiores a 0,4 como consecuencia de la discontinuidad introducida. Puesto que trabajamos con tramos de cinco términos, a partir del 25 la serie nuevamente recupera su estabilidad hasta el elemento 31, en que se detecta una discrepancia respecto de los precedentes al cesar el incremento introducido y, por tanto, manifestarse como nueva ruptura de la continuidad de la serie. A partir del elemento

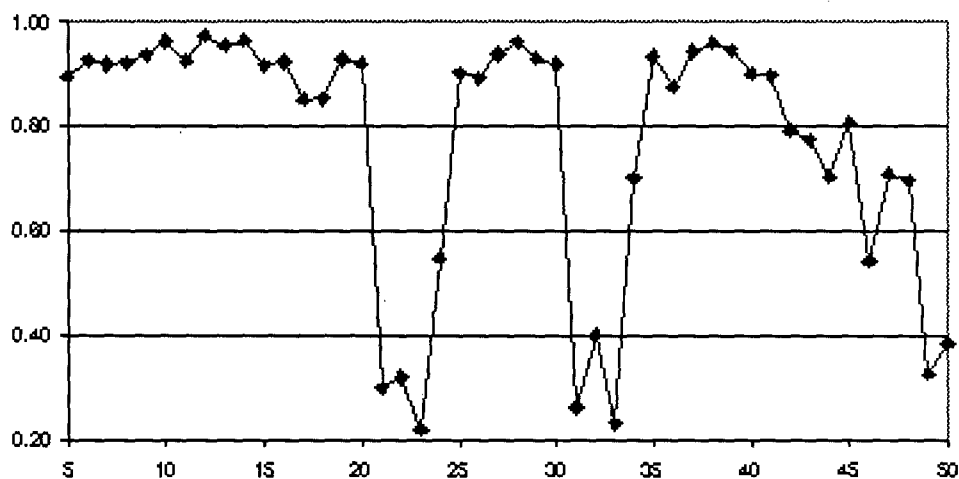


Figura 1: Ejemplo de aplicación del método de detección de inhomogeneidades relativas a partir del coeficiente de correlación.

41 se manifiesta una disminución progresiva del coeficiente de correlación (con dientes de sierra), consecuencia de la tendencia inducida en la serie C.

3. APLICACIÓN DEL MÉTODO A SERIES PLUVIOMÉTRICAS DEL SUROESTE DE ESPAÑA

Los observatorios del suroeste de España objeto de estudio (Badajoz, Córdoba, Sevilla-Tablada, Riotinto/Huelva, y San Fernando (Cádiz)) disponen de series centenarias así como de un historial bastante conocido. En este conjunto de observatorios se ha procedido inicialmente al relleno de las escasas lagunas existentes y a la unificación de los periodos de registros, con la correspondiente homogeneización interna, por medio de las rectas de regresión a partir de periodos de solapamiento con observatorios bien correlacionados (estaciones hidrográficas, observatorios con series de corta duración, etc). Una vez completadas las series iniciales se ha realizado un primer análisis comparativo con el auxilio de métodos gráficos, lo que ha permitido ratificar (o, en alguna ocasión, rectificar) las modificaciones introducidas. Dichas series constituyen para nosotros desde ese momento series de datos "observados", con independencia de que así sea en sentido estricto, o de que provengan de relleno de lagunas, o de la prolongación de la serie, reservando la acepción de "calculados" o "generados" a aquellos valores que resulten por elaboración en alguna fase del modelo de homogeneización propuesto.

Tras esta fase preparatoria se inicia el proceso de detección de inhomogeneidades relativas, el cual puede adoptar dos formas diferentes: la elaboración de las series de coeficientes de correlación móviles por pares de observatorios, testando así la homogeneidad de todos contra todos, (ello facilita la asignación de la inhomogeneidad al observatorio correspondiente), o bien la elaboración de las correlaciones móviles entre cada una de las series analizadas (X , Y , Z ...) y una nueva serie ficticia de referencia (R), que sería la media de las restantes series pluviométricas, excluida, desde luego, la propia serie objeto de análisis en cada caso. Ello implica que para cada serie X a analizar, se elaboraría una serie de referencia diferente (R_x). En el caso de las variables meteorológicas, y especialmente las pluviométricas, la gran variabilidad espacial y las abundantes peculiaridades ligadas a la naturaleza del territorio aconsejan mejor el empleo de la segunda forma, en la cual el peso de estas particularidades geográficas de detalle se atenúa, eliminándose así el efecto de meras perturbaciones de los rasgos generales de la distribución espacial de la precipitación. Por su parte, el empleo de la media en la obtención de la serie de referencia podría sustituirse por el de la media ponderada de las estaciones, siendo el factor ponderador el valor de la correlación existente entre cada una de las estaciones participantes en el proceso y la serie objeto de estudio. La realidad, sin embargo, es que no se mejora sustancialmente el proceso con la introducción de esta mejora, que, por otra parte, es ligeramente más laboriosa.

Hemos aplicado el modelo a nuestras series desde el año 1900 hasta 1998. Es destacable la fuerte asociación existente entre todas, que se pone de manifiesto en el alto valor de correlación registrado entre cada una de ellas y la serie denominada "suroeste", la cual resulta de la media aritmética de todas ellas (ver tabla 1). Se garantiza así el requisito previo para la utilización del procedimiento. Posteriormente, para cada uno de los observatorios (X) se ha elaborado la serie de referencia, (R_x), en este caso como promedio de los datos simultáneos del resto de los observatorios.

Tabla 1: COEFICIENTES DE CORRELACIÓN ENTRE CADA UNO DE LOS OBSERVATORIOS ESTUDIADOS Y EL PRÓMEDIO DE LA REGIÓN (1900-98)

Badajoz/Suroeste	Córdoba/Suroeste	Riotinto/Suroeste	S.Fernando/Suroeste	Sevilla/Suroeste
0.83	0.90	0.91	0.86	0.93

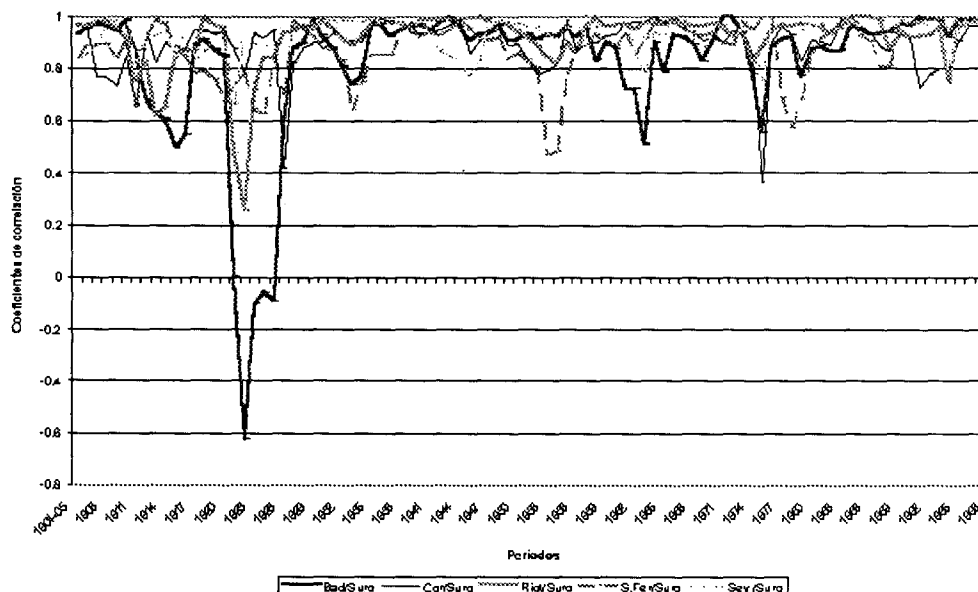


Figura 2: Aplicación del método de homogeneización a las series pluviométricas del suroeste de España (1901/05: 1993/98).

Con tramos sincrónicos de cinco elementos hemos generado la serie de los coeficientes de correlación (1900/05; 1901/06; ... ;1994/98) entre las series $\{X, R_x\}$ (ver figura 2). En las mismas observamos, como era previsible, que la casi totalidad de elementos alcanzan valores superiores a 0,8 lo que facilita la identificación de cada una de las discordancias relativas.

Destaca la anomalía de los valores de precipitación anual entre 1921 y 1925 en Badajoz, parcialmente coincidente con la de Ríotinto en 1921-1922, así como el bienio 1953-54 en San Fernando, el año 1963 en Badajoz y la anualidad de 1975 en Córdoba y Badajoz. Hay que resaltar que ninguna de estas discrepancias había sido prevista, al no asociarse con modificaciones de ubicación o del instrumental. Es también destacable el hecho de que no se han detectado tendencias discrepantes, siendo incluso los coeficientes de correlación ligeramente superiores durante el último tercio de siglo, aún cuando las condiciones de urbanización o industrialización de las localidades respectivas no sean uniformes.

Ante las discrepancias, en las fechas citadas, de la evolución temporal de la precipitación debemos

plantearnos si las mismas son debidas a deficiencias en el proceso de medida o registro, y proceder, en tal caso, a la homogeneización.

A modo ilustrativo hemos representado gráficamente los resultados del proceso para las estaciones de Badajoz y Tablada durante los últimos cincuenta años (ver figura 3).

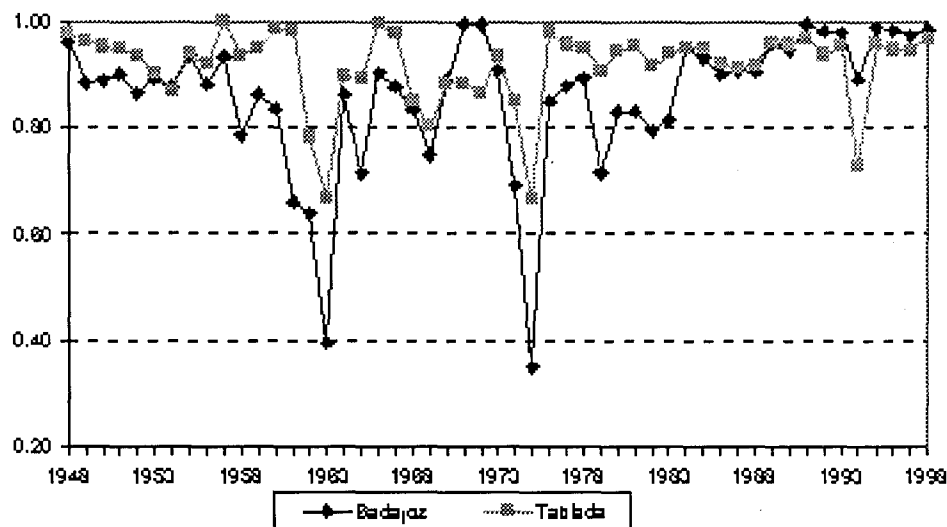


Figura 3: Aplicación del método a las estaciones de Badajoz y Tablada (1948-1998).

Observamos que los coeficientes de correlación por tramos de cinco años entre Badajoz y las respectivas series del Suroeste presentan disparidades, descendentes en 1963 y aislada en 1975. Esto a su vez provoca una disminución relativa de los coeficientes en los otros observatorios, inducida por el sistema de cálculo, como se percibe en Tablada. Una vez detectadas estas disparidades es preciso volver a las series de registros y efectuar un análisis comparado de su evolución.

4. HOMOGENEIZACIÓN

El procedimiento expuesto tiene la virtualidad de exponer las anualidades en las cuales el comportamiento de las magnitudes no ha sido análogo en distintos observatorios. La existencia de estas discrepancias no implica necesariamente defecto de calidad en la serie de observaciones, sino que puede ser reflejo de la diversidad real de comportamiento, con lo que se enriquecen las conclusiones. Por tanto, los primeros resultados que ofrece el método propuesto son la mera indicación cuantitativa de las discrepancias relativas existentes entre los observatorios. A partir de ese momento es importante proceder a un análisis de tales discrepancias al objeto de establecer hasta qué punto constituyen inhomogeneidades o no, y es finalmente el criterio del investigador el que establece cuándo el análisis comparado proporciona indicios suficientes para admitir error en el

proceso de medida-registro, si tiene base documental en que apoyarse, o métodos alternativos de confirmación/refutación (por ejemplo: amplitud térmica en el caso de las temperaturas, observación gráfica, etc.).

Aquellas discrepancias a las que el investigador les confiere carácter de inhomogeneidad pueden ser directamente subsanadas por el propio modelo. Para ello, en cada serie de observaciones se sustituirían los datos calificados como inhomogéneos por los valores resultantes de la recta de regresión establecida entre la serie estudiada (una vez eliminados los valores inhomogéneos) y su correspondiente serie de referencia.

En caso de que las inhomogeneidades fueran simultáneas en dos series de observaciones diferentes, habría que establecer un proceso iterativo de homogeneización. En este caso se recomienda empezar por rectificar la serie que presenta el menor coeficiente de correlación con la serie de referencia; con el nuevo valor se opera sobre la segunda serie, y así sucesivamente hasta estabilizar los coeficientes. En caso de que hubiera un número demasiado elevado de observatorios como para dificultar en exceso el proceso iterativo, es preferible anular la contribución de los observatorios sometidos a sospechas al elaborar la serie de referencia.

5. CONCLUSIONES

Conocemos muy escasos precedentes bibliográficos de la utilización de métodos de detección de inhomogeneidades similares a éste, es decir, fundamentados teóricamente en las propiedades de asociación entre elementos correspondientes de distintas series (CARREGA, 1994, LABORDE, 2000), y en ningún caso alcanzan el nivel de efectividad y precisión que atribuimos al que ahora presentamos.

Consideramos que, frente a métodos más complejos para evaluar la calidad de las series, posee la virtud de no imponer demasiados requisitos previos a la estructura interna de las mismas, siendo necesario, sin embargo, que las series estén bien correlacionadas entre sí.

La definición de forma objetiva de la serie temporal de referencia -aún cuando sea ficticia- permite la aplicación de sistemas de análisis simples y rigurosos. En consecuencia, la obtención de resultados cuantificados proporciona parámetros de interpretación directa, en cuanto a duración e intensidad, de las discrepancias evolutivas en las series iniciales de observaciones.

Determinadas las inhomogeneidades relativas, el mismo modelo propuesto aporta los mecanismos de homogeneización de cada una de las series meteorológicas de los observatorios analizados, lo cual constituye una ventaja adicional importante del método.

6. REFERENCIAS

AGUILAR, E. *et al.* (1999): "Control de calidad y proceso de homogeneización de series térmicas catalanas". En RASO, J.M. y MARTIN-VIDE, J.(Eds), "*La climatología española en los albores del siglo XXI*", Publicaciones de la Asociación Española de Climatología (AEC), Serie A, nº 1, pp15-23.

ALEXANDERSON, H. (1986): "A homogeneity test applied to precipitation data". *Journal of Climatology*, 6, 661-675.

ALEXANDERSON, H. (1997): "Homogeneization of Swedish data". *Int. Journal of Climatology*, 17, 25-34.

ALMARZA, C. y LÓPEZ, J.A. (1994): "Rachas húmedas y rachas secas durante el periodo instrumental de observaciones. El caso de San Fernando (Cádiz)". En PITA, M.F. y AGUILAR, M. (Eds), "*Cambios y variaciones climáticas en España*", Fundación El Monte y Universidad de Sevilla, 147-167.

CARRÉGA, P. (1994): "Les tendances actuelles du climat de la Côte d'Azur (France)". *Publications de l'AIC*, vol. 7, 173-182.

GARCÍA-BARRÓN, L. (2000): "*Análisis de series termopluviométricas para la elaboración de modelos climáticos en el suroeste de España*". Tesis Doctoral, Universidad de Sevilla, 156 pp + anexos.

LABORDE, J.P. (2000): "*Hydrolab (software para el tratamiento estadístico de las series meteorológicas e hidrológicas)*". Universidad de Niza.

RODRÍGUEZ BARRERA *et al.* (1996): "Revisión de los criterios de homogeneidad aplicados a variables meteorológicas". En "*Modelos y SIGS, Actas del VII Coloquio de Geografía Cuantitativa*", 96-111.