X Congreso Internacional AEC: Clima, sociedad, riesgos y ordenación del territorio
DOI: http://dx.doi.org/10.14198/XCongresoAECAlicante2016-37

RECONSTRUCCIÓN E INCERTIDUMBRE EN SERIES DE PRECIPITACIÓN DIARIA INSTRUMENTAL

Roberto SERRANO NOTIVOLI[1,2], Santiago BEGUERÍA[2], Miguel Ángel SAZ[1], Martín DE LUIS[1]

[1] *Departamento de Geografía y Ordenación del Territorio. Instituto Universitario de Ciencias Ambientales de Aragón (IUCA). Universidad de Zaragoza.*

[2] *Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Zaragoza.*

rs@unizar.es

RESUMEN

Se presenta un nuevo método de reconstrucción de series de precipitación diaria que permite filtrar y completar series originales de precipitación y crear series continuas en cualquier punto del territorio. De forma independiente, para cada día y localización se calculan dos valores de referencia: una predicción binomial (PB) que expresa la probabilidad de que un día sea húmedo o seco y una predicción de magnitud (PM) que estima la cantidad de precipitación. Para calcular estos dos valores de referencia se utilizan regresiones logísticas multivariantes con los datos de los diez observatorios más cercanos, usando como covariables la latitud, longitud y altitud de estas diez observaciones y de la serie candidata. El cálculo de estos valores de referencia nos permite: 1) aplicar un control de calidad independiente para cada dato de precipitación observado; 2) estimar valores de precipitación en los días sin observación; 3) crear nuevas series en lugares donde no existía observación a partir de las estaciones reconstruidas vecinas; y 4) crear mallas regulares de valores diarios de precipitación. Las estimaciones de precipitación diaria incluyen una estimación del error estándar para cada dato, que puede ser utilizada para evaluar la incertidumbre de las predicciones asociadas a cada momento y lugar. Para facilitar el uso de los procesos descritos se ha creado un paquete de funciones en lenguaje R denominado reddPrec, disponible para cualquier usuario en el repositorio oficial. El método se ha testado sobre 2.150 observatorios del sector noreste de España con valores de precipitación diaria desde 1950 hasta 2012.

Palabras clave: precipitación, reconstrucción, incertidumbre, error, rejilla.

ABSTRACT

A new daily precipitation series reconstruction method, which lets filter and complete original precipitation series and create continuous ones in any location of territory, is presented. Two reference values are computed for each day and location separately: a binomial prediction (PB) which is referred to the probability of occurrence of a wet or dry day and a magnitude prediction (PM) that estimates the amount of precipitation. To compute these two reference values are used multivariate logistic regressions using latitude, longitude and altitude as covariates. The calculation of these reference values let us: 1) apply a quality control to each values and day separa-

tely; 2) estimate precipitation values in days with no observation; 3) create new series from neighbor reconstructed stations where there were no observation; and 4) create regular grids of daily precipitation data. The daily precipitation estimations include an error estimation for each value that can be used to assess the related uncertainty of predicted precipitation for each moment and location. To ease the use of these processes, a functions package was created in R language named reddPrec, that is available to any user at official repository.

Key words: precipitation, reconstruction, uncertainty, error, grid.

1. INTRODUCCIÓN

La creciente demanda de datos climáticos de cada vez mayor resolución espacial y temporal requiere de metodologías automatizadas de control de calidad y reconstrucción que produzcan bases de datos fiables y continuas en el tiempo y el espacio. Actualmente existen varias bases de datos de precipitaciones a escala mensual y anual a nivel mundial (Mitchell and Jones, 2005; Lawrimore *et al*., 2011; Becker *et al*., 2013), regional (Klein-Tank *et al*., 2002; Li *et al*., 2009) y también para la península Ibérica (Ninyerola *et al*., 2007, Gonzalez-Hidalgo *et al*., 2011). Sin embargo, la precipitación diaria ha sido relegada en la investigación del clima a un segundo plano por la dificultad existente para trabajar con observatorios de diferentes características y especialmente de diferentes fuentes. A pesar de todo, para la región española se han desarrollado algunas bases de datos a escala diaria con diferentes aproximaciones metodológicas (Vicente-Serrano *et al*., 2010, Belo-Pereira *et al*., 2011; Herrera *et al*., 2012).

El presente trabajo describe una metodología de reconstrucción de series de precipitación diaria basada en la distribución espacial de las observaciones individualizadas para cada día por separado, aportándose para cada dato reconstruido un valor de error estimado que identifica su fiabilidad. Con objeto de testar la fiabilidad del protocolo metodológico empleado, éste se ha aplicado en la provincia de Teruel, un ámbito geográfico complejo desde el punto de vista orográfico y climático, donde aparece información pluviométrica oficial de hasta 7 fuentes distintas y con valores de precipitación que oscilan entre los menos de 400 mm del norte de la provincia y los más de 1.200 de las áreas mejor expuestas del suroeste. Primero se aplica un control de calidad en el que se descartan datos anómalos, para rellenar a continuación los datos faltantes. Con la base de datos completa es posible crear nuevas series de datos donde antes no existía observación e incluso crear mallas regulares (grids). La metodología explicada en este trabajo se ha implementado en un paquete de R denominado reddPrec, disponible en descarga libre y gratuita desde (https://cran.r-project.org/web/packages/reddPrec/index.html).

2. DATOS Y MÉTODOS

2.1. Datos

Se utilizaron un total de 2.150 estaciones de precipitación con datos diarios desde 1950 hasta 2012 del interior la provincia de Teruel y aquellas en un radio de 100 km alrededor de los límites provinciales (Fig. 1). De ellas, 1.676 procedían del Banco

Nacional de Datos de AEMET, 138 del SAIH Júcar, 124 del SAIH Ebro, 120 de la red del SIAR del Ministerio de Agricultura Alim. y Medio Ambiente, 65 del Servei Meteorològic de Catalunya, 26 del SAIH Tajo y 1 del SAIH Duero. En el periodo de estudio, el conjunto de las estaciones acumulaba un 69,7% de datos faltantes.

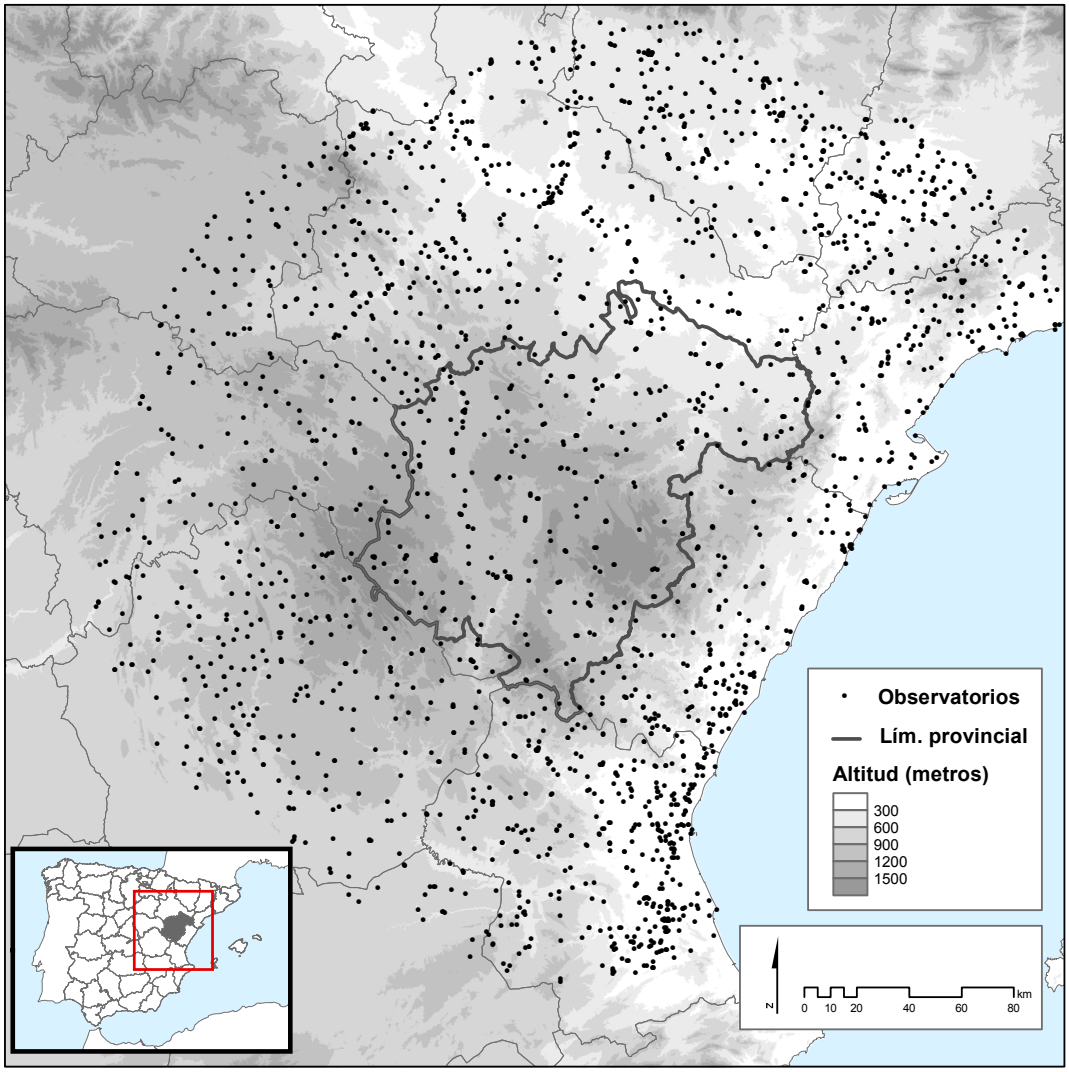

Fig. 1: Localización de los observatorios utilizados.

2.2. Cómputo de valores de referencia

Todo el proceso de reconstrucción se basa en el cálculo de valores de referencia (VR) tal y como se detalla en Serrano-Notivoli *et al*. (2016a). Estos valores se calculan para cada día y localización, y están basados en los datos observados en el entorno de cada una de ellas. La base del cálculo es una regresión logística multivariante:

$$P_{i,l} = \beta_{0_{i,l}} + \beta_{1_{i,l}} alt_l + \beta_{2_{i,l}} lat_l + \beta_{3_{i,l}} lon_l + \varepsilon_{1_{i,l}} \quad \text{(Ec. 1)}$$

donde $P_{i,l}$ es el valor predicho para un día y localización determinados, $\beta_{n_0^3}{}_{i,l}$ son los coeficientes de la regresión, alt_l, lat_l y lon_l son la altitud, la latitud y la longitud de la localización candidata y $\varepsilon_{1_{i,l}}$ es el error asociado al modelo de regresión.

Esta regresión se utiliza para calcular dos valores predichos que condicionarán el VR final: Una predicción binomial (PB) que expresa la probabilidad de que un día sea húmedo o seco y una predicción de la magnitud (PM) que estima la cantidad de precipitación recibida. En función del cálculo de uno u otro, el modelo expresado en Ec.1 se verá matizado. En el caso del cálculo de PB se utilizarán los datos de los 10 observatorios más cercanos codificados como una variable binomial, y en el caso del cálculo de PM se utilizarán los valores originales de esos observatorios pero reescalados entre 0 y 1, siendo 0 la mitad de la precipitación mínima de todos ellos y 1 el valor máximo más el rango (diferencia entre el máximo y el mínimo). Este reescalamiento de los valores originales permite estimaciones por encima del valor máximo observado en el entorno, pero establece un límite superior e inferior adaptado a cada localización y momento específico.

Finalmente, el valor de referencia (VR) se obtiene mediante una combinación de los dos valores predichos descritos. Si el valor de PB es inferior a 0,5 se considera que el día ha sido seco en esa localización y por lo tanto el RV asignado es de cero. Si por el contrario, el PB es superior o igual a 0,5 se considera que el día ha sido húmedo y RV tomará el valor de PM predicho.

2.3. Reconstrucción

El proceso básico de reconstrucción de cualquier variable climática se desarrolla en dos fases: Control de calidad y relleno de lagunas.

2.3.a. Control de calidad

Se aplicó un control de calidad a todas las observaciones utilizadas. Como los datos observados fueron examinados uno a uno de manera independiente, se utilizaron los cinco criterios definidos en Serrano-Notivoli *et al*. (2016b). Estos criterios pueden ser modificados para adaptarlos a la realidad climática de cada zona, pero en este caso se consideraron adecuados para determinar umbrales a partir de los cuales extraer de la base de datos original aquellos valores que no correspondiesen a la variabilidad climática propia del área de estudio.

1. A1: El valor observado es superior a cero y sus diez vecinos más cercanos son cero.
2. A2: El valor observado es cero y sus diez vecinos más cercanos son superiores a cero.
3. A3: La magnitud del valor observado es 10 veces superior o inferior al VR correspondiente.
4. A4: El valor observado es cero, PB es superior a 0,99 y PM es superior a 5 mm.
5. A5: El valor observado es superior a 5 mm, PB es menor de 0,01 y PM es inferior a 0.1 mm.

2.3.b. Relleno de lagunas

Utilizando la base de datos filtrada por los criterios de control de calidad descritos, se calcularon nuevos VR pero esta vez para todos los días y localizaciones sin dato

original. Como cada VR se calcula cada día por separado, es habitual que en cada uno se utilicen vecinos diferentes. Con el fin de evitar la aparición de inhomogeneidades por esta razón, se dividió la precipitación media observada en cada serie y la precipitación media predicha para esa misma serie en los días con observaciones. Esto produce una serie de valores predichos (VR) con la misma media que los observados. Para evitar que los coeficientes introdujeran variaciones que no correspondieran a la variabilidad climática propia de cada serie, los coeficientes se calcularon mensualmente de manera independiente. La series finales están compuestas por los valores originales filtrados y los valores faltantes rellenados con los VR corregidos por la serie de observados correspondiente.

Finalmente, y para asegurar la consistencia climática de la base de datos resultante, se rellenaron aquellas, aquellas que tenían más de 10 años de datos originales.

2.4. Creación de nuevas series de datos

Utilizando la misma metodología de aplicación de modelos individualizados para cada día y localización, es posible obtener series de nueva creación en cualquier punto del territorio. El único requisito previo es el de tomar como punto de partida la base de datos reconstruida, es decir, con todas las series filtradas, completas y sin huecos, ya que la nueva localización candidata deberá tener los mismos 10 vecinos todos los días para no introducir inhomogeneidades en la nueva serie.

En este caso, se optó por crear una malla regular de pares de coordenadas separadas entre sí 5 kilómetros sobre la provincia de Teruel. Se crearon 689 puntos en total. Hay que tener en cuenta que la malla resultante no representa una superficie continua de precipitación, sino que contiene estimaciones individualizadas para cada uno de los puntos que representan los pares de coordenadas. No se trata entonces de un proceso de interpolación espacial al uso, sino que es un proceso de inferencia estadística para una localización concreta x, y, z. Para la estimación de precipitación diaria en esta malla se utilizaron las 1.258 estaciones reconstruidas y completadas en el paso anterior.

3. RESULTADOS

3.1. Reconstrucción

De las 2.150 series iniciales, las 1.258 reconstruidas tenían un 54,1% de datos faltantes. Tras la detección y eliminación de datos en el proceso de control de calidad, los huecos aumentaron hasta el 55,1%.

Los criterios seleccionados para el control de calidad eliminaron prácticamente el doble de datos en la primera mitad del periodo respecto a la segunda (Fig. 2). Esto se debe principalmente a la densidad espacial de datos por día. Hacia los primeros años, la cantidad de datos disponible era mucho menor que al final, por lo tanto, para seleccionar los 10 vecinos más cercanos era necesario cubrir distancias más grandes y, por tanto, comparar el dato observado con datos de un entorno que en muchas ocasiones no guardaba patrones de distribución espacial de la precipitación similares en ese día. Este problema afecta especialmente a este tipo de metodologías de selección espacial, pero es recurrente en todo tipo de trabajos de reconstrucción climática con cualquier variable.

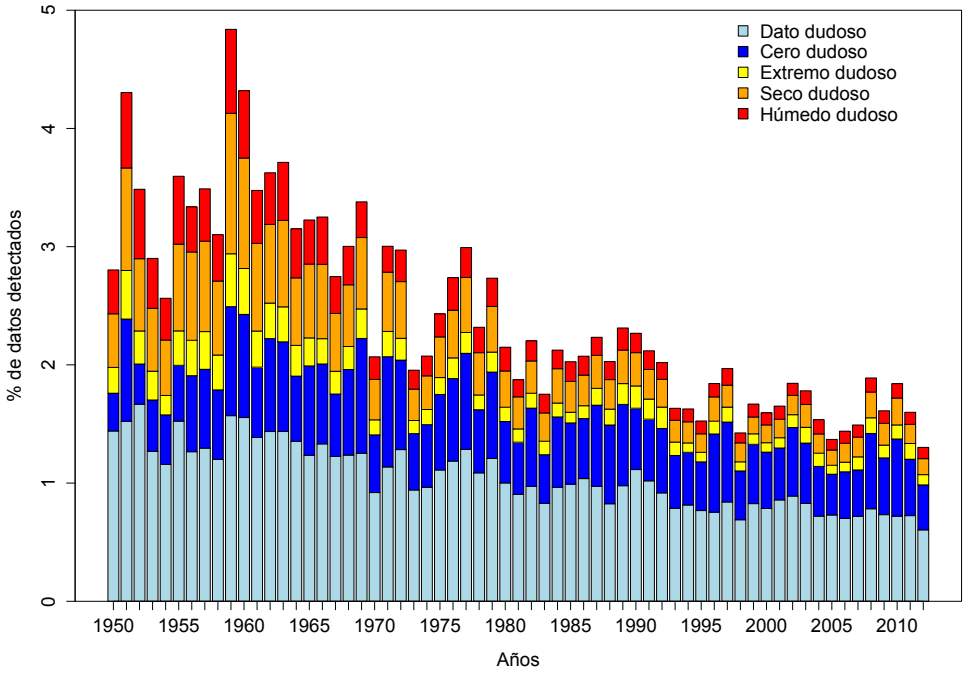

Fig. 2: Porcentaje anual de datos eliminados según criterio.

Tras el control de calidad se rellenaron todos los huecos utilizando los VR calculados a partir de la base de datos filtrada. Para cada estación, se obtuvo una serie completa con los valores originales que no fueron eliminados en el proceso de control de calidad, y los valores estimados en los huecos (Fig. 3).

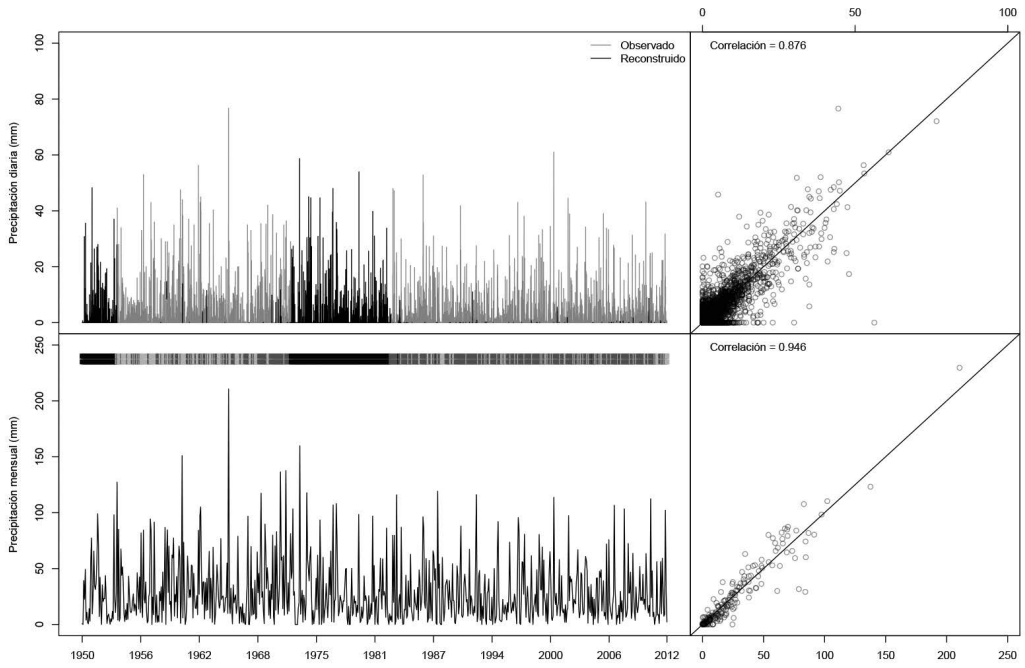

Fig. 3: Reconstrucción del observatorio de precipitación de 9547 (La Puebla de Híjar). Arriba: valores observados y reconstruidos diarios y correlación entre ellos (arriba derecha). Abajo: valores observados y reconstruidos mensuales y correlación entre ellos (abajo derecha)

3.2. Creación de nuevas series de datos

A partir de las estaciones reconstruidas se estimaron valores de precipitación diaria para cada uno de los puntos en la malla regular de 5x5km. De cada punto de malla se computaron las estimaciones de precipitación y sus errores, así como los agregados mensuales y anuales (Fig. 4)

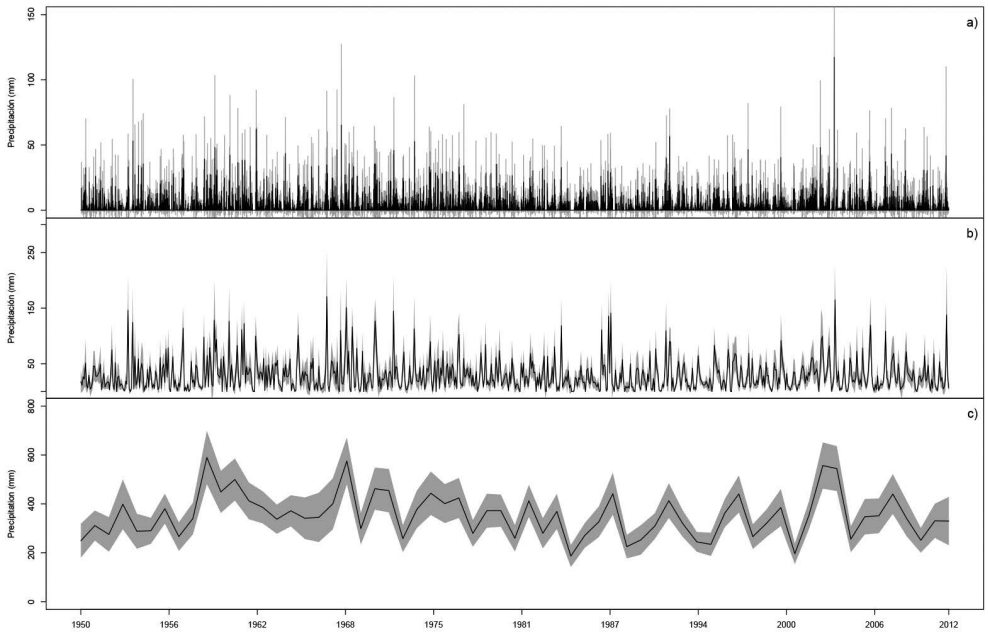

Fig. 4: Serie temporal correspondiente a un punto de malla creado a partir de las estaciones reconstruidas. a) Valores diarios; b) Mensuales; c) Anuales. Las áreas sombreadas en gris representan las bandas de confianza al 95% de cada dato.

El agregado de precipitación anual muestra patrones de distribución espacial lógicos para el territorio (Fig. 5 izda.). La zona de mayores valores (> 1.200mm) se ubica en la Sierra de Albarracín (sector Suroeste) con un segundo máximo en las elevaciones del Maestrazgo (sector Este). Los errores estándar asociados modelo de estimación de la precipitación diarias también fueron agregados anualmente y muestran un patrón espacial coherente pero diferente de la climatología de la zona (Fig. 5 centro).

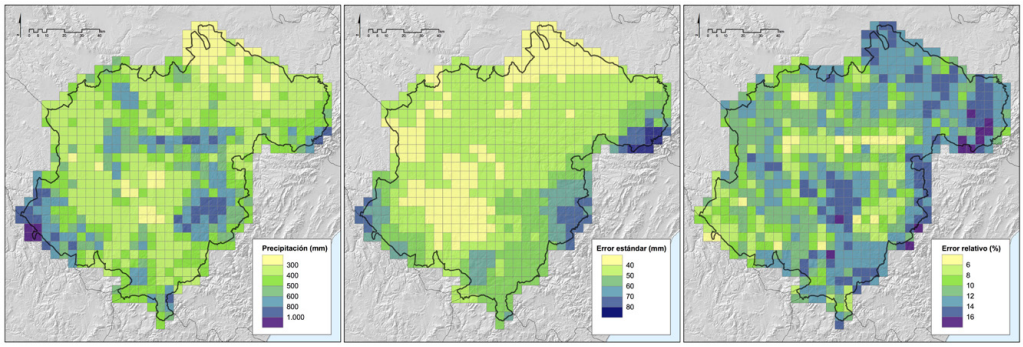

Fig. 5: Precipitación anual (izquierda), error estándar de la estimación (centro) y error relativo (derecha).

En este caso las zonas de mayor error se ubican en la zona oriental de la provincia por encima de los 70 mm en muchos casos. Los errores relativos, calculados como la ratio entre los errores estándar y el valor de precipitación, muestran la incertidumbre real de cada uno de los datos estimados en cada punto de malla. En este caso la variabilidad espacial es mucho mayor. La distribución de los errores relativos no tiene porqué estar relacionada con la climatología espacial del territorio, simplemente muestra aquellas zonas donde es más difícil hacer una estimación correcta de la precipitación. Normalmente se trata de zonas con alta irregularidad en la distribución de la variable.

4. DISCUSIÓN

En este trabajo se presenta una metodología de reconstrucción de series diarias de precipitación que incluye: 1) un control de calidad aplicado a cada día y observatorio; 2) el relleno de lagunas de información de las series filtradas mediante el cálculo de valores de referencia a partir de dos valores predichos que determinan la ocurrencia de precipitación o no (PB) y la magnitud de la misma (PM); y 3) la posibilidad de crear nuevas series de datos a partir de los observatorios reconstruidos, incluso en formato rejilla.

La principal diferencia de este método respecto a los habitualmente utilizados es que la mayoría requieren de series de partida que tengan un mínimo de datos disponibles para que sean comparables con las de su entorno. En este caso, cualquier serie con al menos un dato será útil, ya que la reconstrucción se hace día a día por separado sin asumir ninguna similitud previa entre series. Utilizando los diez observatorios más cercanos a la localización candidata se consigue que la estimación que devuelve el modelo refleje el carácter local de la precipitación para ese punto. Sin embargo, el inconveniente principal es que en las situaciones en las que no hay una densidad de observaciones suficiente, se tiene que ampliar el radio de búsqueda para encontrar esos diez vecinos, lo que puede provocar, en algunos casos que, por ejemplo, se detecten como valores sospechosos algunos que realmente no lo sean por ser comparados con otros que quizá no corresponden a los patrones de distribución habituales de esa localización.

El uso de modelos de regresión logística multivariante para el cálculo de los valores de referencia tiene la ventaja de la flexibilidad de las predicciones en función de los datos de origen. Uno de los efectos no deseados de este tipo de modelos es que pueden producir extrapolaciones cuando el observatorio candidato está fuera de rango, por ejemplo cuando se trata de localizaciones extremas (en la costa, en zonas muy altas sin vecinos por encima, etc.). Para amortiguar este efecto, se introdujeron una asíntota superior e inferior mediante el reescalado de los datos originales. A pesar de que seguirá produciéndose extrapolación, ésta será más suave y continua en el espacio.

La capacidad de crear de nuevas series es clave para desarrollar climatologías en localizaciones donde no existen observaciones, un problema habitual en cualquier estudio climatológico. Este método es capaz de crear nuevas series en formato rejilla, siempre considerando que no se trata de una superficie continua de precipitación, sino de una estimación puntual para localizaciones específicas. La diferencia con otras rejillas ya disponibles es que para cada uno de los datos estimados, se aporta un valor

del error asociado al mismo que resulta muy útil ya que la mayoría de los métodos de creación de grids o rejillas reducen, en diferentes proporciones, la varianza respecto a los datos observados (Beguería *et al*., 2015).

AGRADECIMIENTOS

Este estudio ha sido posible gracias a los proyectos de investigación CGL2012-31668, CGL2015-69985-R, CGL2011-24185 y CGL2014-52135-C3-1-R, financiados por el Ministerio de Economía y Competitividad (MINECO) y fondos FEDER. Los investigadores agradecen también la ayuda al Gobierno de Aragón a través del "Programa de grupos de investigación" (grupos "H38, Clima, Cambio Global y Sistemas Naturales" y "E68, Geomorfología y Cambio Global").

REFERENCIAS

Beguería, S., Vicente-Serrano, S.M., Tomás-Burguera, M., Maneta, M. (2015). Bias in the variance of gridded data sets leads to misleading conclusions about changes in climate variability. International Journal of Climatology, 36(9), 3413-3422. doi: http://dx.doi.org/10.1002/joc.4561.

Belo-Pereira, M., Dutra, E., Viterbo, P. (2011): Evaluation of global precipitation data sets over the Iberian Peninsula. J. Geophys. Res., 116, D20. doi: http://dx.doi.org/10.1029/2010JD015481.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. Earth Syst. Sci. Data, 5, 71–99. doi: http://dx.doi.org/10.5194/essd-5-71-2013

Herrera, S., Gutiérrez, J., Ancell, R., Pons, M., Frías, M., Fernández, J. (2012). Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). International Journal of Climatology, 32(1), 74-85. doi: http://dx.doi.org/10.1002/joc.2256.

González-Hidalgo, J.C., Brunetti, M., de Luis, M. (201)1. A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945-November 2005). International Journal of Climatology, 31(5), 715-731. doi: http://dx.doi.org/10.1002/joc.2115.

Klein-Tank, A.M.G., Wijngaard, J-B-, Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J.A., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V., Petrovic, P. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology. 22, 1441–1453. doi: http://dx.doi.org/10.1002/joc.773

Li, Q., Zhang, H., Liu, X., Chen, J., Li, W. (2009). A mainland China homogenized historical temperature dataset of 1951-2004. Bulletin of the American Meteorological Society. 90, 1062–1065. doi: http://dx.doi.org/10.1175/2009BAMS2736.1

Mitchell, T.D., Jones, P.D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology, 25, 693–712.

Ninyerola, M., Pons, X., Roure, J.M. (2007). Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. Theoretical and Applied Climatology, 89(3), 195-209. doi: http://dx.doi.org/10.1007/s00704-006-0264-2.

Serrano-Notivoli, R., Beguería, S., Saz, M.A., de Luis, M. 2016a. Spatially-based reconstruction of daily precipitation instrumental data series. International Journal of Climatology. En revisión.

Serrano-Notivoli, R., de Luis, M., Beguería, S., Saz, M.A., 2016b. Spatially-based quality control for daily precipitation datasets. Geophysical Research Abstracts, 18, EGU2016-16456. Viena, Austria.

Vicente-Serrano, S., Beguería, S., López-Moreno, J., García-Vera, M., Stepanek, P. (2010). A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. International Journal of Climatology, 30(8), 1146-1163. doi: http://dx.doi.org/10.1002/joc.1850.