

RECONSTRUCCIÓN DE TEMPERATURAS A ESCALA DIARIA EN EL PARQUE NATURAL DEL MONCAYO (ESPAÑA)

Roberto SERRANO-NOTIVOLI, José María CUADRAT, Luis Alberto LONGARES ALADRÉN, Edurne MARTÍNEZ DEL CASTILLO, Klemen NOVAK, Miguel Ángel SAZ SÁNCHEZ, Ernesto TEJEDOR, Martín DE LUIS ARRILLAGA

¹*Depto. de Geografía y Ordenación del Territorio. Universidad de Zaragoza*
rs@unizar.es

RESUMEN

El presente trabajo muestra una metodología para reconstruir series completas de temperatura máxima y mínima diarias a partir de modelos regresivos diarios y locales utilizando los metadatos inherentes a los datos climáticos: latitud, longitud y altitud. Se han utilizado 66 observatorios del conjunto del Macizo del Moncayo procedentes de las redes de la Agencia Estatal de Meteorología, la Confederación Hidrográfica del Ebro y el Parque Natural del Moncayo, cubriendo un rango altitudinal de 2.000 metros y un periodo temporal desde 1950 hasta 2012.

El trabajo de reconstrucción se divide en tres fases: i) Control de calidad exhaustivo a escala diaria comparando cada dato observado con un conjunto de vecinos cercanos mediante ajuste de modelos construidos en función de los parámetros geográficos (latitud, longitud, y altitud) y eliminando aquellos que superan los umbrales establecidos; ii) Segmentación de las series originales para evitar inhomogeneidades; iii) Relleno de lagunas en toda la serie mediante el mismo método extrayendo un valor predicho en función de los datos vecinos depurados.

Palabras clave: Temperatura, reconstrucción, escala diaria, Moncayo, España.

ABSTRACT

Present work shows a new methodology to reconstruct complete maximum and minimum temperature series from regressive daily and local models using inherent metadata of climatic data: latitude, longitude and altitude. 66 observatories in whole Moncayo massif have been used from different networks: Spanish Meteorological Agency, Ebro River Basin, and Moncayo Natural Park, covering an altitudinal range of 2,000 meters and a temporal period from 1950 to 2012.

Reconstruction work has been divided in three phases: i) Exhaustive quality control at daily scale comparing each observed data with a near set of neighbors through the fitting of models constructed with geographic parameters (lat., long, and alt.) and removing those which are upper of thresholds; ii) Segmentation of original series to avoid inhomogeneities; iii) Gap filling in all series through the same method extracting a predicted value depending on debugged neighboring data.

Key words: Temperature, reconstruction, daily scale, Moncayo, Spain.

1. INTRODUCCIÓN

La problemática de la disponibilidad de dato diario en climatología es una constante en cualquier proyecto de investigación que requiera de información climática. Actualmente, las fuentes de datos oficiales ofrecen series incompletas, y en muchos casos con errores e inhomogeneidades que no se ajustan a los requerimientos de calidad que exige la investigación científica.

Los trabajos desarrollados en control de calidad y relleno de información climática en datos mensuales son abundantes (Brunetti *et al.*, 2006; González-Hidalgo *et al.*, 2011; Mitchell y Jones, 2005; Aguilar *et al.*, 1999), y todos ellos utilizan métodos de relleno de lagunas considerando la estructura completa de las series del entorno en su escala temporal, que en el caso de aplicarlos a datos en escala diaria, son susceptibles de contaminar la variabilidad de la serie final. Los trabajos sobre dato diario también han sido tratados (Cano y Gutierrez, 2004; Herrera, 2011; Herrera *et al.*, 2012; El-Kenawy, 2012), pero suelen aplicar los métodos o alguna parte de ellos de las escalas de trabajo mensuales. Metodológicamente, ambas escalas no pueden ser tratadas de la misma manera, ya que climáticamente el comportamiento no es el mismo. El dato diario introduce mucha más incertidumbre y variabilidad, con un porcentaje mucho mayor de aparición de datos extremos, que a escala mensual podrían considerarse datos anómalos. La importancia de obtener un dato climático a escala diaria puede ser escasa si lo que se necesita es obtener una serie de tendencias a largo plazo en series muy largas de información, para esto sería suficiente con el dato mensual. Sin embargo, existen otras aplicaciones que sí lo requieren: estudios muy localizados espacialmente; estudios de base para predicciones de cambio climático; estudios de ecología forestal, etc. Además, a partir de una serie de datos diarios de calidad y homogeneizados, se pueden extraer valores mensuales que sí son adecuados para observar tendencias.

1.1 Área de estudio

El Parque Natural del Moncayo se ubica en la vertiente Este del macizo con una altura máxima de 2.314 metros. Se eleva desde el Valle del Ebro salvando un desnivel de 1.800 metros. Su ubicación geográfica le convierte en uno de los más claros exponentes de los ecosistemas de montaña de transición entre el ambiente mediterráneo y atlántico, condicionando la acusada variación altitudinal unos notables contrastes en las condiciones climáticas que quedan reflejadas en la transición existente de la vegetación (Martínez del Castillo *et al.*, 2012). Es una barrera que impide la entrada fluida de los frentes del Noroeste en el tramo medio del Valle del Ebro, enfría el viento durante los meses de invierno, y recibe la mayor parte de las precipitaciones convectivas en este sector de la Cordillera Ibérica.

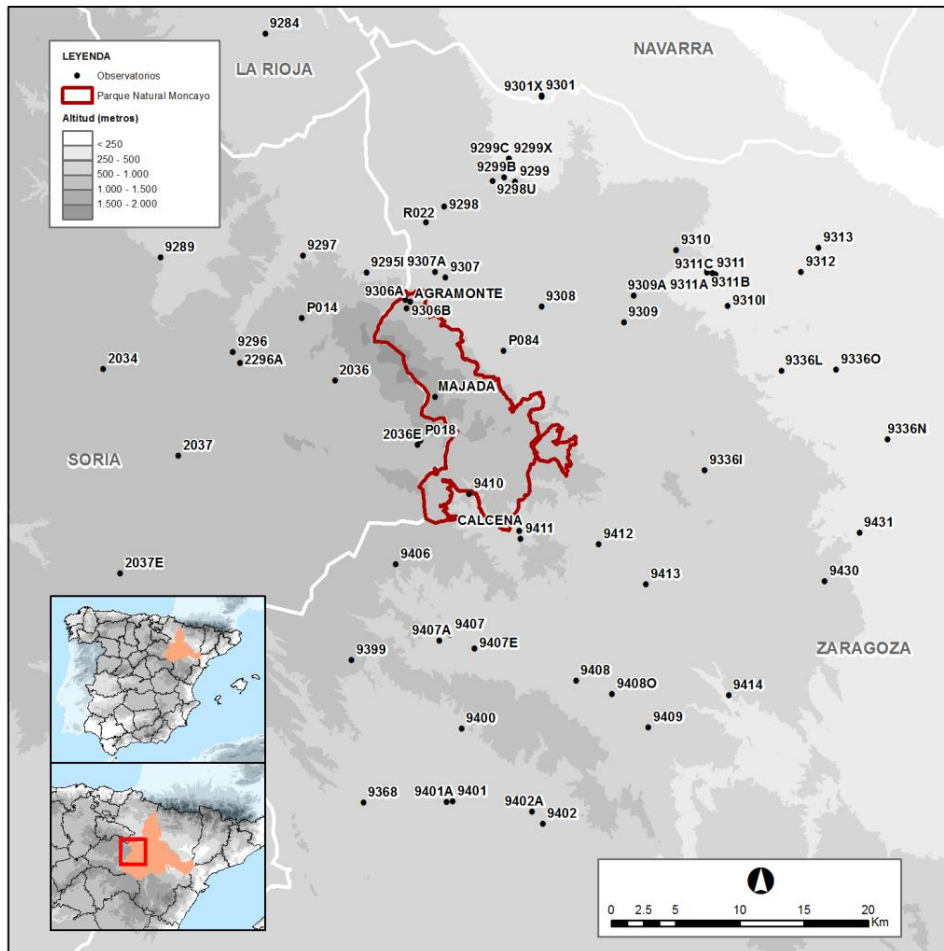


Fig. 1: Localización del área de estudio y ubicación de los observatorios utilizados.

2. DATOS Y METODOLOGÍA

Para el presente estudio, se han utilizado un total de 66 observatorios con registro de temperatura. De todos ellos, 59 forman parte de la red de la Agencia Estatal de Meteorología (AEMET), 4 corresponden a la Confederación Hidrográfica del Ebro (CHE), y los 3 restantes pertenecen al Parque Natural del Moncayo y son el objetivo de la reconstrucción climática en este trabajo (Agramonte, Calcena, y Majada Baja).

2.1 Control de calidad

El control de calidad es un paso fundamental para eliminar los datos incorrectos y comprobar la consistencia de los mismos (Feng *et al.*, 2004). Tanto las series extraídas del Banco de Datos Climatológico Nacional de AEMET, como las provenientes de la CHE, o los datos obtenidos de las estaciones del PN del Moncayo, producen bases de datos en estado bruto, carentes de cualquier tipo de filtrado, y susceptibles de contener todo tipo de errores. Al trabajar con datos a escala temporal diaria, la variabilidad es mucho más manifiesta que con variables mensuales o anuales. La aparición de valores extremos que pueden ser confundidos con datos anómalos (outliers) es mucho más frecuente, por tanto es imprescindible adoptar una metodología que permita identificar

este tipo de anomalías correctamente.

Las metodologías utilizadas para el control de calidad en datos mensuales consideran la comparación de la serie candidata con las de referencia en toda su dimensión temporal, es decir, desde el inicio hasta el fin de cada una de las series como en Esteban *et al.* (2009) o en González-Hidalgo *et al.* (2011). El método propuesto en este trabajo compara, para cada registro diario, las circunstancias que se dieron en todos los observatorios del entorno que condicionaron el valor de la variable en cada uno de ellos, esto es, se hace un control, día por día, de cómo la altitud, la latitud y la longitud, influyeron en el dato de temperatura de cada observatorio y en qué medida. Así, se estima un valor promediado entre todas las estaciones vecinas que servirá de referencia para observar las anomalías que puede haber entre éste (valor de referencia construido con las observaciones vecinas para un día concreto) y la estación objetivo a corregir (valor observado en la estación que se está trabajando).

Al estudiar las series climáticas en comparación a las de sus vecinos, se pueden determinar los días concretos en los que el valor de la variable climática de un observatorio tiene una varianza muy superior a la supuesta por la propia variabilidad local del clima. El objetivo es conseguir una serie de valores Z para cada observatorio, que representan la varianza extrema del valor original respecto a la modelización de los valores predichos de sus vecinos. Para obtener los valores Z se trabaja con los datos disponibles en cada día de todos los vecinos (para áreas más amplias es necesario filtrar el número de vecinos por distancia). Primero se estandarizan (1) a través de la media y la desviación estándar para que sean comparables con la serie candidata

$$Est_i = \left(\frac{CNVECI_i - MMVECI_i}{\sigma_i} \right) \sigma_{cand} + \bar{x}_{cand} \quad (1)$$

Donde Est_i es el valor estandarizado para el día concreto; $CNVECI_i$ es el dato observado del vecino en el día i ; $MMVECI_i$ es la media móvil de un mes de los valores observados de cada vecino; σ_i es la desviación estándar del conjunto de valores de un mes de cada vecino; σ_{cand} es la desviación estándar del conjunto de valores de un mes de la serie candidata; y \bar{x}_{cand} es la media móvil de un mes de los valores observados de la serie candidata.

Los valores estandarizados previamente se ajustan a los parámetros de altitud, latitud y longitud mediante el modelo de regresión que mejor se ajuste. Este ajuste se hace mediante polinomios desde orden 1 (ajuste lineal) hasta 4. El que mejor r^2 produce es el método elegido para ese parámetro en concreto y ese día.

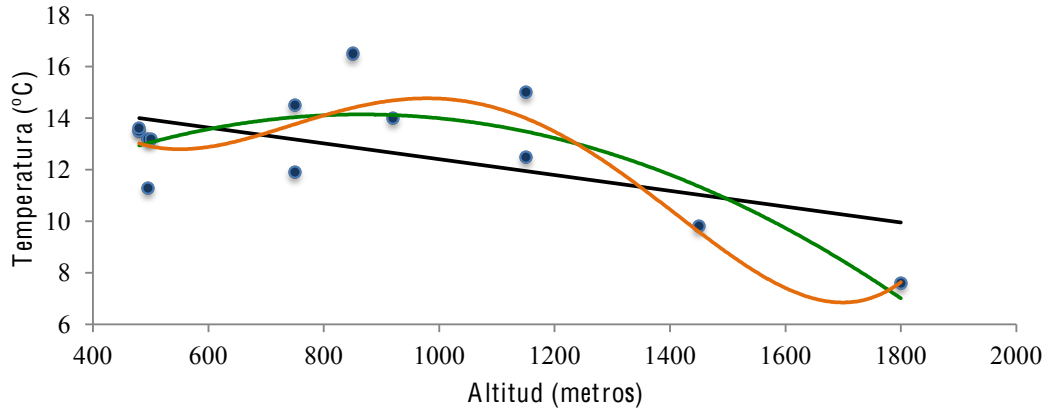


Fig. 2. Diferentes métodos de ajuste de los datos de temperatura máxima para un día concreto a los valores de altitud. Desde regresión lineal (negro) ($r^2 = 0.31252$), hasta modelos de regresión con polinomios de orden 4 (naranja) ($r^2 = 0.75481$), pasando por una regresión con orden de polinomio 2 (verde) ($r^2 = 0.6823$).

Para adecuar los datos predichos a la serie candidata se estandarizan y se promedian para obtener unos predichos finales (2) en función de la r^2 obtenida para cada modelo.

$$PFinal_i = \frac{(PParcial_{i\ alt}r^2_{i\ alt})+(PParcial_{i\ lon}r^2_{i\ lon})+(PParcial_{i\ lat}r^2_{i\ lat})}{r^2_{i\ alt}+r^2_{i\ lon}+r^2_{i\ lat}} \quad (2)$$

Donde $PFinal_i$ es el valor Predicho Final para un día; $PParcial_{i\ alt}$ es el Predicho Parcial de la Altitud para ese día; $r^2_{i\ alt}$ es el valor de r^2 para el modelo de predicción de la altitud de ese día; ; $PParcial_{i\ lon}$ es el Predicho Parcial de la Longitud; $r^2_{i\ lon}$ es el valor de r^2 para el modelo de predicción de la longitud; $PParcial_{i\ lat}$ es el Predicho Parcial de la Latitud; y $r^2_{i\ lat}$ es el valor de r^2 para el modelo de predicción de la latitud.

El cálculo final de los valores Z (3) se obtiene del cociente entre la serie de diferencias calculada entre el valor predicho final y su correspondiente valor observado (Dif_i), y la desviación estándar de esa misma serie de diferencias

$$Z_i = \frac{Dif_i}{\sigma_{Dif}} \quad (3)$$

Esta serie de valores Z representa una serie de anomalías que se interpreta como el número de desviaciones estándar que la serie original se aleja de la serie de referencia estandarizada a partir de los datos de todos los observatorios vecinos, e individualizada para cada uno de los días de la serie candidata.

Cuando se ha realizado el primer cálculo de valores Z , se eliminan, de los valores originales, aquellas Z cuyo valor es superior a 3,5 (en valor absoluto). Estos valores indican los días en los que el valor original está 3,5 desviaciones estándar por encima o por debajo respecto al predicho a través de la modelización de sus vecinos, por eso se eliminan de la base de datos original, sobre la que se vuelve a repetir la iteración hasta que ya no existan. En este trabajo se ha eliminado un 5% de datos sobre el total de observaciones que se han considerado anómalos en función de estos criterios.

2.3 Relleno de lagunas

Las series originales de datos no son continuas, de hecho tienen huecos que responden a días en los que el observatorio no registró nada, o que han sido eliminados por considerarse anomalías ajenas a la respuesta digital de los instrumentos de medida.

Si estos huecos han sido superiores a 30 días, la serie se fragmenta en ese punto y pasa a considerarse como otra nueva. La razón de esta fragmentación es que, al tratar los datos de una misma serie como periodos continuos independientes, se elimina una posible contaminación de las medias y desviaciones estándar en el caso de que ambos periodos fuesen inhomogéneos (es decir, que tuviesen una señal climática diferente que respondiese a efectos ajenos a la propia evolución del clima).

El resultado de la fragmentación provoca que las 66 series originales de temperatura se conviertan en 360 en el caso de las máximas, y 363 en el caso de las mínimas. Tras el relleno de lagunas de cada una de ellas la serie final es un promedio ponderado por el número de datos originales de cada segmento, manteniendo los originales que no han sido eliminados en el proceso de detección de anomalías.

El proceso de relleno se compone de 4 etapas bien diferenciadas y todas ellas necesarias para obtener una serie final de valores robusta estadísticamente:

i) Estandarización de vecinos: Se realiza a partir de los datos de los vecinos depurados mediante la eliminación de las anomalías calculadas previamente.

ii) Relleno a periodo común: Se seleccionan como vecinos las 10 estaciones más cercanas a la candidata que, para el periodo con dato en ésta, compartan al menos 30 días observados. Esto asegura un relleno con homogeneidad temporal de al menos un mes para cada dato diario rellenado.

iii) Siguiendo iteración, utilizar los datos generados en la anterior para continuar con el relleno: después del primer relleno se genera una nueva matriz de datos que incorporarán los datos originales depurados, y los predichos por un primer relleno. En sucesivas iteraciones, esta matriz se utilizará como series de vecinos, y volverán a rellenar todos los huecos de la serie candidata original. Este proceso se repite tantas veces como sea necesario hasta que estén rellenas todas las series en su periodo completo.

iv) Relleno final: El último relleno se realiza sobre la matriz de datos en la que ya todas las estaciones están rellenas completamente, desde el primer día hasta el último. A partir de aquí solamente es necesario hacer una modelización de todos los datos de los vecinos para cada candidata en los días en los que no había observación originalmente, esto evitará variabilidades que no respondan a la tendencia original de la serie, pues el valor rellenado para cada día será un promedio de las series estandarizadas construidas con los vecinos más cercanos a ellas. El resultado final es un conjunto de series completas desde el inicio hasta el final sin un solo hueco, que mantiene la variabilidad de las series originales, matizada por las situaciones locales tanto espacial como temporalmente.

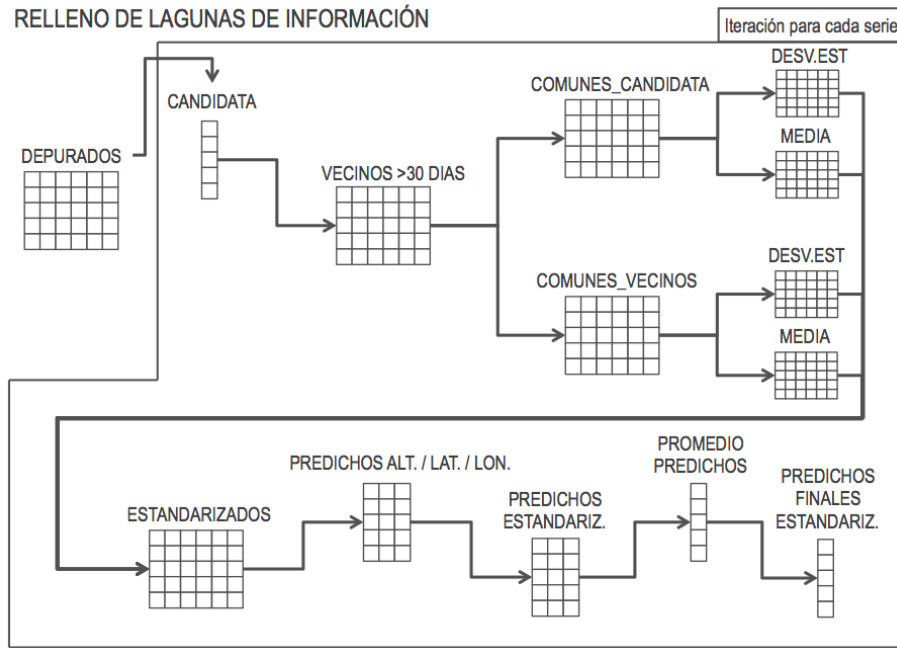


Fig. 4. Proceso de cálculo de los predichos para relleno.

3. RESULTADOS Y DISCUSIÓN

El ajuste de los valores predichos a los observados cuando se trabaja sobre el dato diario (del que se han modelizado todos los predichos) es algo inferior a la situación que representan las medias mensuales. Esta diferencia viene dada por el aumento de la variabilidad del clima con el aumento de la escala temporal. De esta manera, comparando por pares todos los observados iniciales con sus respectivos predichos, se obtiene un valor de r^2 medio de 0.9529 para las máximas, y de 0.9137 para las mínimas, un ajuste bastante bueno teniendo en cuenta que se trata de dato diario.

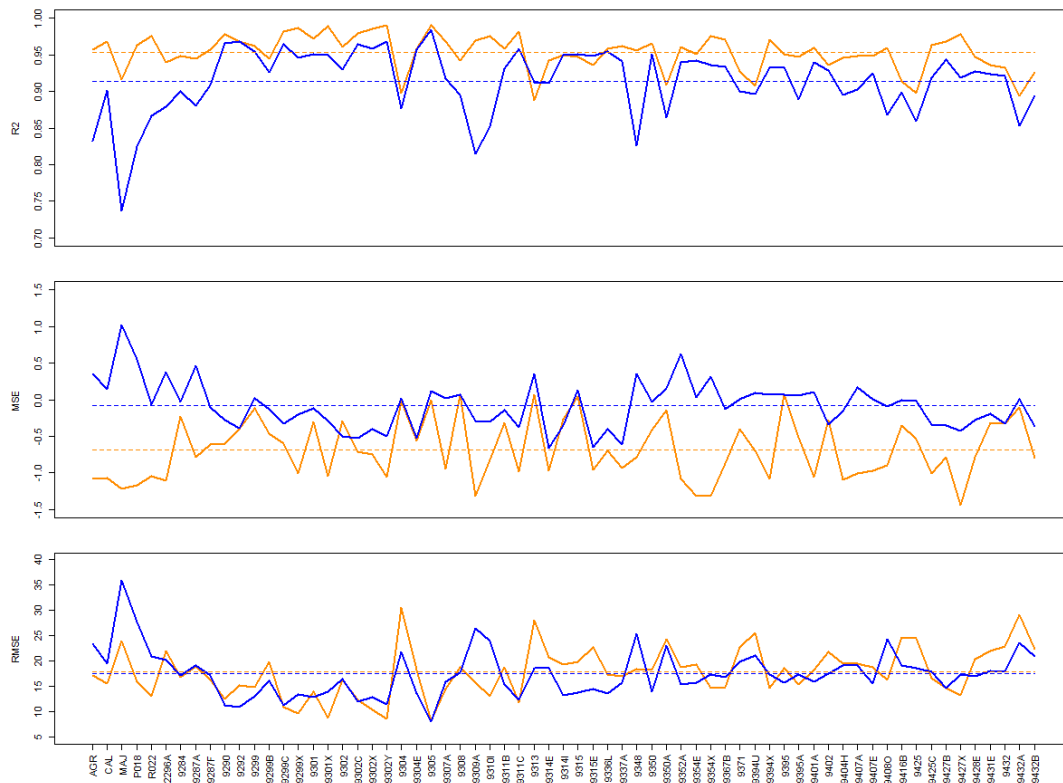


Fig. 6. Comparativa Observados-Predichos en dato diario para todas las estaciones de temperatura máxima (naranja) y mínima (azul). a) Valores de r^2 en correlación de Pearson; b) Valores de Error Medio Absoluto; c) Valores de Error Cuadrático Medio

Los resultados de la modelización muestran unos resultados realmente buenos para las temperaturas máximas, y buenos para las mínimas, especialmente en Majada Baja (1.798 m.s.n.m.) (Figs. 7c.1 y 7c.2), donde el ajuste entre observados y predichos tiene un coeficiente de 0.74, esta zona es proclive a sufrir fenómenos de inversión térmica que no son bien detectados por el modelo. En la estación de Agramonte (1.090 m.s.n.m.) (Figs. 7a.1 y 7a.2), la situación es diferente pero el modelo funciona mejor con valores de r^2 de 0.83 para las mínimas y 0.95 para las máximas.

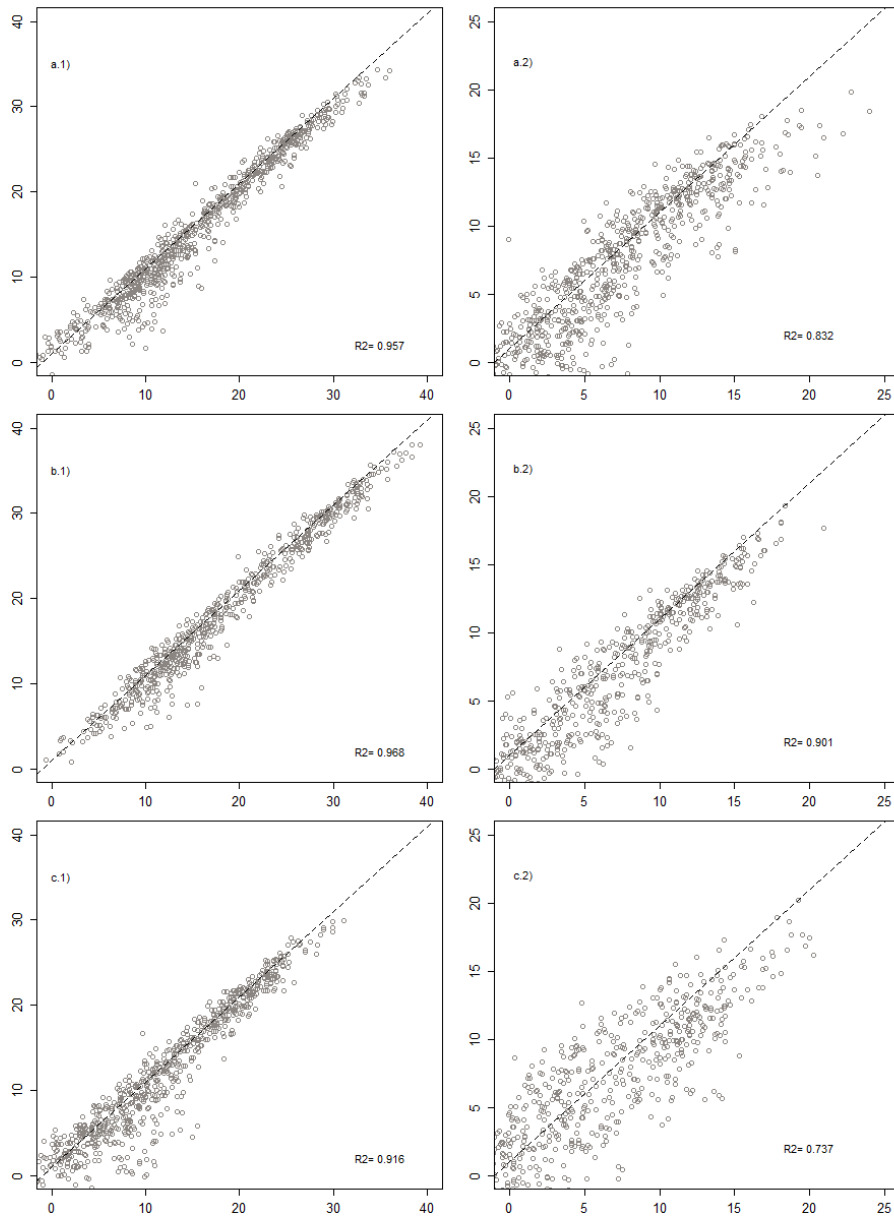


Fig. 7. Comparativa Observados-Predichos en dato diario para las estaciones de temperatura máxima y mínima. a) Agramonte; b) Calcena; c) Majada Baja. Col. izquierda: Temp. máxima; Col. Derecha: Temp. mínima

En general las temperaturas mínimas tienen más problemas para ser modelizadas correctamente que las máximas. En la Fig. 8 se muestra la relación entre el ajuste por correlación entre predichos y observados y la altitud. La causa es que los modelos de orden polinomial muy alto extrapolan los resultados al estar la observación fuera de rango.

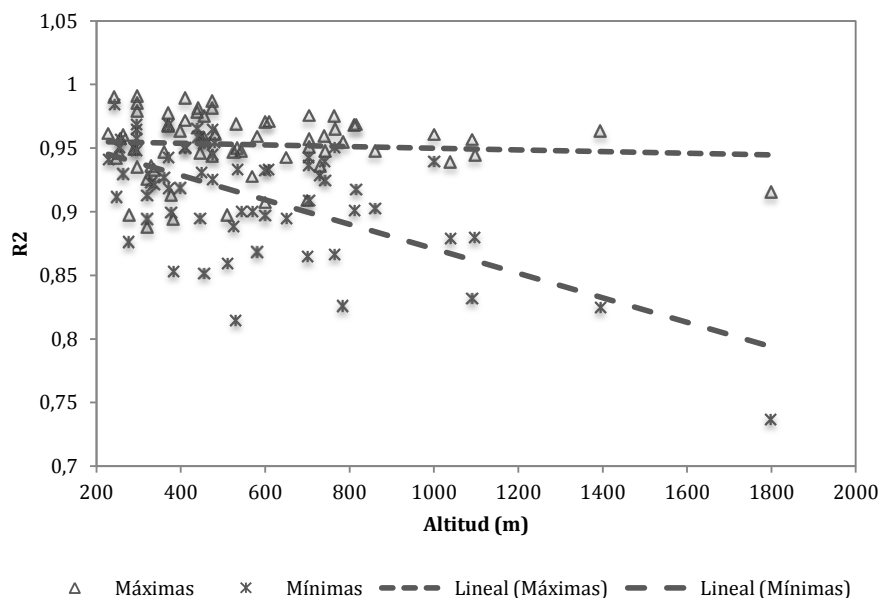


Fig. 8. Comparativa entre la correlación Observados-Predichos y la altitud de las estaciones

Se ha testado la homogeneidad de las tres estaciones del interior del Parque (Agramonte, Calcena, y Majada Baja) a través del agregado mensual del promedio de todas sus segmentaciones rellenadas. Mediante el software HOMER (Mestre *et al.* 2013) se han buscado puntos de ruptura (*breakpoints*) en los que haya habido un cambio en la variabilidad de la serie que no responda a factores climáticos. Para determinar un *breakpoint* como válido, se consideraron como ciertos aquellos concentrados alrededor de un mismo año y en un mínimo de tres series vecinas pertenecientes a la misma subred (en este caso el conjunto de estaciones en un radio de 20 kilómetros alrededor del parque) (Kuglitsch *et al.*, 2009), y todo ello contrastado con el metadato disponible, si lo hubiera. No se han detectado puntos de ruptura claros, por lo que no ha procedido la corrección de ninguna de las tres series.

4. CONCLUSIONES

En comparación a otros trabajos de escala diaria, Herrera (2011) acepta como validación un valor de R de 0.67 ($R^2=0.45$) para dato diario en su base de datos de temperatura Spain02, Cano (2004) no ofrece ningún valor de correlación o de error medio para todo el periodo estudiado, y El-Kenawy (2013) solamente da valor de R para dos estaciones, 0.92 y 0.93 ($R^2=0.85$ y 0.86 respectivamente).

En el presente estudio se han encontrado coeficientes de determinación más bajos en estaciones a mayor altura (Majada Baja, 1.800 m.s.n.m.). El ajuste de los datos en estas estaciones es menor, ya que el polinomio utilizado para interpolar el valor final está construido en muchas ocasiones con pocos datos (cada día cambia el número de vecinos) y todos ellos a menor altitud de estas estaciones candidatas a gran altitud, con lo que el valor predicho que se obtiene es una extrapolación que no encaja perfectamente en la serie original.

Para evitar valores predichos en altura que no se ajusten correctamente a modelos de orden de polinomio altos, se ha asignado en todos los días el modelo de orden 1

(regresión lineal) a partir de la altitud de la estación más elevada, de esta manera predice de manera menos irreal el valor final en altura para aquellas ubicaciones susceptibles de sufrir *overfitting* en su predicción (aquellas que están fuera del rango de los input).

Por otra parte, la estandarización de fragmentos que no contienen un año completo de observaciones, corre el riesgo de estar sesgada por los propios datos, ya que al estandarizar, por ejemplo, solamente con datos de verano o de invierno, modifica la variabilidad de la serie de referencia que rellena. Una posible corrección, a falta de realizar pruebas concluyentes, puede ser la de calcular una serie inicial de referencia a través de la modelización de los datos brutos, y extraerles su ciclo anual (media de cada día para todos los años), que después se añadiría a la serie de anomalías calculadas.

Las series resultantes del relleno tienen todos los días del periodo con un dato modelizado que mantiene la variabilidad y la media de la serie original, aunque se da en algún caso, especialmente en las temperaturas mínimas y en estaciones a mayor altura, que el ajuste no es tan perfecto como en las temperaturas máximas, hecho que puede deberse a una estandarización sobre periodos cortos, inferiores a un año, que no reflejan la variabilidad climática total que representa ese punto en el espacio. Este procedimiento será corregido en versiones posteriores del método, sin embargo, las tres estaciones objeto del trabajo, que disponían originalmente de tan solo tres años de observaciones (2010-2012) han sido rellenadas hasta 1950 con ajustes de r^2 superiores en muchos casos a 0.90, lo que indica que este método puede mejorar en algunos aspectos pero que es válido para la reconstrucción de valores diarios.

5. REFERENCIAS

- AGUILAR, E., LÓPEZ, J.M., BRUNET, M., SALADIÉ, O., SIGRÓ, X, LÓPEZ, D. (1999) Control de calidad y proceso de homogeneización de series térmicas catalanas. En: RASO, J.M. Y MARTÍN-VIDE, J. (Eds): *La climatología española en los albores del siglo XXI*. Publicaciones de la Asociación Española de Climatología (AEC), Serie A, nº 1 pp. 15-23.
- BRUNETTI, M., MAUGERI, M., MONTI, F., NANNI, T. (2006). “Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series”. *International Journal of Climatology*, 26, pp. 345–381.
- CANO, R., GUTIERREZ, J.M. (2004). Relleno de lagunas y homogeneización de series de precipitación en redes densas a escala diaria. En: DIEGO, C., GARCÍA CODRÓN, J.C., RASILLA, D., FERNÁNDEZ DE ARRÓYABE, P., GARMENDIA, C. (Eds.): *El Clima entre el mar y la montaña*. Publicaciones de la Asociación Española de Climatología (AEC), Serie A, nº 4. Santander. Pp. 431–440.
- EL-KENAWY, A. (2012): *Spatio-temporal variability of surface air temperature in Northeastern Spain*. Tesis doctoral.
- EL-KENAWY, A., LÓPEZ-MORENO, J.I., STEPANEK, P., VICENTE-SERRANO, S. (2013): “An assessment of the role of homogenization protocol in the performance of daily temperature series and trends: application to northeastern Spain”. *International Journal of Climatology*. 33, pp. 87–108
- ESTEBAN, P., NINYEROLA, M., PROHOM, M. (2009): “Spatial modelling of air temperature and precipitation for Andorra (Pyrenees) from daily circulation patterns”. *Theoretical Applied Climatology* 96, pp. 43–56

- FENG, S., HU, Q., QIAN, W. (2004): "Quality control of daily meteorological data in China, 1951-2000: A new dataset". *International Journal of Climatology*, 24(7), pp. 853-870
- GONZÁLEZ-HIDALGO, J.C., BRUNETTI, M., DE LUIS, M. (2011), "A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945-November 2005)", *International Journal of Climatology*, vol. 31, no. 5, pp. 715-731.
- HERRERA, S. (2011): *Desarrollo, validación y aplicaciones de Spain02: Una rejilla de alta resolución de observaciones interpoladas para precipitación y temperatura en España*. Tesis doctoral. <http://www.meteo.unican.es/en/tesis/herrera>
- HERRERA, S., GUTIÉRREZ, J.M., ANCELL, R., PONS, M.R., FRÍAS, M.D., FERNÁNDEZ, J. (2012): "Development and Analysis of a 50 year high-resolution daily gridded precipitation dataset over Spain (Spain02)". *International Journal of Climatology* 32, pp. 74–85.
- KUGLITSCH, F.G., AUCHMANN, R., BLEISCH, R., BRÖNNIMANN, S., MARTIUS, O., STEWART, M. (2012): "Break detection of annual Swiss temperature series", *Journal of Geophysical Research*, 117, D13105.
- MARTÍNEZ DEL-CASTILLO, E., SERRANO-NOTIVOLI, R., NOVAK, K., LONGARES-ALADRÉN, L.A., ARRECHEA, E., DE-LUIS, M., SAZ-SÁNCHEZ, M.A. (2012): Cuantificación de los gradientes climáticos altitudinales en la vertiente Sur del Macizo del Moncayo a partir de una nueva red de estaciones automáticas en altura. En: RODRÍGUEZ PUEBLA, C., CEBALLOS, A., GONZÁLEZ, N., MORÁN TEJEDA, E., HERNÁNDEZ, A. (Eds): *Cambio Climático. Extremos e Impactos*. Publicaciones de la Asociación Española de Climatología (AEC), Serie A, nº 8. Salamanca.
- MESTRE, O., DOMONKOS, P., PICARD, F., AUER, I., ROBIN, S., LEBARBIER, E., BÖHM, R., AGUILAR, E., GUIJARRO, J.A., VERTACHNIK, G., KLANCAR, M., DUBUISSON, B., STEPANEK, P. (2013). HOMER : "HOMogenisation softwarE in R- methods and applications". *Időjárás*, 117, pp. 47–67.
- MITCHELL, T., JONES, P. (2005): "An improved method of constructing a database of monthly climate observations and associated high-resolution grids". *International Journal of Climatology* 25, pp. 693–712.
- WILLMOTT, C.J., MATSUURA, K. (2006), "On the use of dimensioned measures of error to evaluate the performance of spatial interpolators", *International Journal of Geographical Information Science*, 20, pp. 89–102.