

# Serie de precipitación diaria en rejilla con fines climáticos

## Nota técnica 24 de AEMET

Candelas Peral García  
Beatriz Navascués Fernández-Victorio  
Petra Ramos Calzado



MINISTERIO  
DE AGRICULTURA Y PESCA,  
ALIMENTACIÓN Y MEDIO AMBIENTE





Aviso Legal: los contenidos de esta publicación podrán ser reutilizados, citando la fuente y la fecha, en su caso, de la última actualización

**Edita:**

© Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente  
Agencia Estatal de Meteorología  
Madrid, 2017

Catálogo de Publicaciones de la Administración General del Estado:

<https://cpage.mpr.gob.es>

NIPO: 014-17-009-5

<https://doi.org/10.31978/014-17-009-5>

Agencia Estatal de Meteorología (AEMET)

C/ Leonardo Prieto Castro, 8

28040 Madrid

<http://www.aemet.es/>



@Aemet\_Esp



<https://www.facebook.com/AgenciaEstataldeMeteorologia>

## Resumen

Se presenta una serie de precipitación acumulada en 24 horas en rejilla a escala diaria sobre España peninsular y Baleares, que comienza el 1 de enero de 1951 hasta la actualidad, y que se ha generado con el fin de que pueda servir de herramienta para la vigilancia y predicción del clima. Para ello, se han utilizado las series de observaciones del Banco Nacional de Datos Climatológicos de AEMET que fueron seleccionadas para la producción de escenarios regionalizados de cambio climático. El espaciado de la rejilla es de 5 km aproximadamente, si bien las estructuras de precipitación tienen una resolución inferior. El método utilizado para crear los campos de precipitación acumulada a escala diaria es un análisis por Interpolación Estadística, un método de estimación lineal que maneja bien distribuciones irregulares de observaciones, y produce el mínimo error de interpolación con una correcta formulación de las estadísticas de los errores de las estimaciones de las que parte para crear el campo analizado.

Como punto de partida, se ha utilizado el sistema de análisis de superficie del modelo de predicción numérica HIRLAM operativo en AEMET, el cual ha sido modificado y adaptado en ciertos aspectos para el desarrollo de esta serie en rejilla. Entre los más importantes están los parámetros que regulan las propiedades del análisis por Interpolación Estadística, que han sido sintonizados con las series de observaciones utilizadas. El análisis generado considera la influencia de la orografía, así como la variación estacional de la resolución de las estructuras de la precipitación, y de la distribución espacial de la propia precipitación y su variabilidad diaria.

Los diagnósticos que se han realizado para comprobar el funcionamiento del análisis, así como la validación de los campos de precipitación frente a observaciones independientes, muestran un correcto comportamiento del sistema desarrollado. Como parte del trabajo de evaluación, se han calculado también algunas estadísticas climatológicas a partir de la rejilla desarrollada. Los resultados que se obtienen son consistentes con los calculados en cada una de las estaciones que se han utilizado para generarla. También muestran acuerdo con las estadísticas climatológicas que proceden de la rejilla Spain02, de menor resolución y generada con anterioridad sobre un periodo más corto, si bien existen pequeñas diferencias entre ambas en las zonas de menor densidad de observaciones.

## Contenido

1	Introducción .....	5
2	Metodología .....	6
2.1	Observaciones utilizadas .....	6
2.2	Método de análisis .....	8
2.3	Cajas de análisis .....	9
2.4	Controles de calidad de las observaciones .....	9
2.5	<i>First guess</i> utilizado .....	10
2.6	Rejilla del análisis .....	11
3	Modelado de las estadísticas de los errores del <i>first guess</i> y de las observaciones .....	11
3.1	Modelos utilizados para representar las estadísticas de los errores del <i>first guess</i> y de las observaciones .....	11
3.2	Método para sintonizar los modelos de los errores del <i>first guess</i> y de las observaciones ..	12
3.3	Componente de la función de estructura dependiente de la distancia horizontal .....	13
3.4	Modelado de la anisotropía debida a la orografía .....	16
3.5	Varianzas del error del <i>first guess</i> y de las observaciones .....	17
4	Serie de precipitación diaria en rejilla desde 1951: Diagnósticos y validación .....	20
4.1	Diagnósticos del análisis .....	20
4.2	Validación frente a estaciones independientes .....	21
4.3	Estadísticas climatológicas obtenidas con la precipitación en rejilla .....	23
5	Conclusiones .....	27
6	Reconocimientos .....	28
7	Referencias .....	29

## 1 Introducción

En el ámbito científico, técnico, y también del público en general, existe una demanda creciente de una descripción detallada espacial y temporalmente de las variables climáticas esenciales, y entre ellas de forma prioritaria la temperatura y precipitación, que va más allá de disponer de series temporales de observaciones meteorológicas *in situ*. El conocimiento de la evolución de estas variables permite una vigilancia de la variabilidad del clima y del cambio climático más completa y detallada; pero también estudiar su impacto en los ecosistemas y en los numerosos sectores socioeconómicos sensibles al tiempo y al clima, validar las predicciones climáticas en las diferentes escalas temporales, postprocesar o combinarlas con las salidas de los modelos climáticos globales para generar predicciones adaptadas a las necesidades particulares de los distintos usuarios, producir predicciones climáticas con métodos empíricos etc. Estos requerimientos justifican los esfuerzos dedicados a la obtención de datos de las variables climáticas en rejillas espaciales a intervalos regulares de tiempo que abarquen el periodo más largo que sea posible.

Actualmente se encuentran disponibles series temporales de datos en rejilla procedentes de proyectos de reanálisis globales producidos por sistemas de Predicción Numérica del Tiempo (PNT), como por ejemplo ERA-interim (Dee et al., 2011) del Centro Europeo de Predicción a Plazo Medio, o los reanálisis de NCEP (Kalnay et al., 1996). Estos reanálisis aprovechan los avances que se han producido en los modelos numéricos de predicción meteorológica, y en sus sistemas de asimilación de datos de todo tipo de observaciones, para crear descripciones muy precisas del estado de la atmósfera a intervalos de tiempo regulares. Los sistemas utilizados se basan en versiones fijas de las componentes de asimilación de datos y del modelo de predicción a lo largo de todo el periodo, con el fin de evitar en lo posible la introducción de inhomogeneidades en la serie. En los últimos años, se han generado también regionalizaciones de estos reanálisis globales con modelos numéricos de área limitada, entre ellas la realizada con el modelo HIRLAM sobre Europa (Dahlgren et al., 2016). A pesar de que se ha ido incrementando la resolución espacial desde que comenzaron a generarse los reanálisis, sus campos aún no tienen la suficiente resolución espacial requerida para ciertas aplicaciones. Además, el campo de precipitación que proporcionan los reanálisis es normalmente una predicción del modelo, y solo en pocas ocasiones se utilizan observaciones pluviométricas para generar un análisis de precipitación en su postproceso. Este es el caso del proyecto de reanálisis HIRLAM sobre Europa (Landelius et al., 2016), que tras el reanálisis, ha realizado un paso posterior de análisis objetivo de precipitación con una resolución de 5 km. Los sistemas de análisis objetivo de las cadenas de PNT utilizan dos tipos de estimaciones: una predicción a corto plazo del modelo que actúa como primera estimación o *first guess*, y las observaciones. Tanto en el caso de que la precipitación sea una predicción del modelo (salida del reanálisis), como en el que exista un paso de análisis posterior que la utilice como *first guess*, la climatología del modelo numérico, que puede estar sesgada con respecto al clima real, está presente en los campos previstos del reanálisis y puede estarlo también en el del análisis si no ha sido posible corregirla con las observaciones disponibles. Por estos motivos, entre otros, se han desarrollado desde hace más de una década series temporales de precipitación y temperatura en rejilla complementarias a los proyectos de reanálisis, e independientes de los modelos PNT, mediante la aplicación de diferentes métodos de interpolación espacial de las observaciones climatológicas.

En Europa, una de las rejillas construida únicamente a partir de las observaciones con métodos de interpolación geoespacial, y que es ampliamente utilizada para aplicaciones climáticas es E-Obs (Haylock et al., 2008). En España, la colaboración entre AEMET y la Universidad de Cantabria hizo posible el desarrollo conjunto hace unos años de la rejilla Spain02 (Herrera et al., 2012), a escala diaria sobre España peninsular y Baleares, con una mayor resolución (20 km aproximadamente). Spain02 utiliza un conjunto de observaciones seleccionadas del Banco Nacional de Datos Climatológicos de AEMET, cuya densidad espacial supera a la de las empleadas por E-Obs en España. Posteriormente, se crearon nuevas versiones de Spain02, y más recientemente se ha ampliado su resolución espacial hasta alrededor de 11 km (más información en <http://www.meteo.unican.es/es/datasets/spain02> ).

Los Servicios Meteorológicos de diferentes países ejecutan rutinariamente desde hace años sistemas de análisis objetivo para generar campos de precipitación en rejilla, aunque no siempre con fines climáticos. Por ejemplo, en Francia, Météo-France desarrolló en los años 90 el sistema SAFRAN (Durand et al., 1993, 2009). En Canadá, el proyecto CaPA (Canadian Precipitation Analysis) del Servicio Meteorológico Canadiense puso en funcionamiento un sistema de análisis para producir campos de

precipitación acumulada en 6 horas a una resolución de 15 km sobre Norte América en tiempo real (Mahfouf et al., 2007).

Este documento describe la metodología y primera evaluación de una nueva serie de datos pluviométricos en rejilla de 5 km a escala diaria sobre España peninsular y Baleares, orientada a su aplicación en el ámbito del clima, que se ha generado desde el 1 de enero de 1951 hasta la actualidad. La metodología empleada difiere de la utilizada en Spain02, ya que está basada en el paquete de análisis objetivo de las variables superficiales de la cadena de PNT HIRLAM, que ha sido ampliado y adaptado a la variable de precipitación. Los campos de precipitación diaria producidos por una versión preliminar de este análisis, que se ejecuta en tiempo real con un propósito de vigilancia meteorológica con otro conjunto de observaciones, han sido evaluados y comparados a los generados con el sistema de análisis SAFRAN durante un periodo de un año en la cuenca del Ebro (Quintana et al., 2016).

El método de análisis utilizado en esta rejilla es Interpolación Estadística, al igual que en Landelius et al. (2016) para Europa, y que en Mahfouf et al. (2007) para Norte América. Sin embargo, a diferencia de Landelius et al. (2016) y Mahfouf et al. (2007), no se ha utilizado como *first guess* del análisis una predicción a corto plazo de un modelo numérico. Los parámetros que regulan el comportamiento del análisis también son diferentes a los descritos por estos autores y a la versión preliminar descrita en Quintana et al. (2016), ya que se han sintonizado utilizando las estadísticas de las observaciones pluviométricas asimiladas, y modelando aspectos tales como la variación espacial de la variabilidad diaria de la precipitación, la influencia de la orografía, y la dependencia estacional de todos estos parámetros y de la propia precipitación. El conjunto de observaciones con el que se construye el análisis no coincide exactamente tampoco con el de Spain02, ya que se han utilizado otros criterios para seleccionar las estaciones. Tampoco es el mismo que el utilizado por Landelius et al. (2016), puesto que en este proyecto sobre Europa se han asimilado observaciones de precipitación con una densidad espacial muy inferior a las utilizadas en la rejilla que se está presentando aquí.

La estructura de este documento es la siguiente. En la sección 2 se describe la metodología de análisis objetivo que se ha utilizado para generar la rejilla de precipitación diaria. La sección 3 expone cómo se han calibrado los parámetros que rigen el comportamiento del análisis y sus características. En la sección 4 se presentan los distintos aspectos que se han estudiado para evaluar la rejilla generada desde el año 1951 hasta la actualidad. Las principales conclusiones de este trabajo se apuntan en la sección 5.

## 2 Metodología

### 2.1 Observaciones utilizadas

Como la serie de precipitación en rejilla a escala diaria se ha desarrollado con fines climáticos, para generarla se ha utilizado un conjunto seleccionado de observaciones pluviométricas del Banco Nacional de Datos Climatológicos de AEMET. En concreto, las series de datos de precipitación acumulada en 24 horas de las estaciones que fueron escogidas para la producción de escenarios regionalizados de cambio climático con métodos empíricos. Estas series cumplen una serie de requisitos respecto a su completitud (existen más de 19 años con dato anual), y respecto a su homogeneidad (los datos de precipitación acumulada anual superan al menos el test de Alexandersson o el de Wald-Wolfowitz). La aplicación de estos criterios se describe en Brunet et al. (2008).

De esta forma, el número de series temporales utilizadas para la generación de la serie de precipitación en rejilla se reduce a 2300 aproximadamente, de las más de 9000 existentes en el Banco Nacional de Datos climatológicos. La Figura 1 muestra la distribución espacial de las estaciones pluviométricas disponibles en 1951 y a comienzos de la década de los 90. Como se puede observar, es muy irregular y ha variado a lo largo del tiempo. El número de estaciones disponibles para la generación de la rejilla ha evolucionado desde un mínimo de alrededor de 700 al comienzo de los años 50, hasta un máximo de más de 2000 entre comienzos de la década de los 80 y mediados de los 90. En la actualidad, el número de estaciones disponibles ha disminuido debido al reemplazamiento de algunas de ellas por estaciones automáticas, cuyos datos no se han utilizado para la generación de esta rejilla.

La distribución espacial de las estaciones utilizadas, y su evolución en el transcurso del tiempo, es diferente de unas Comunidades Autónomas a otras. En Galicia, el número de estaciones es pequeño y su densidad es bastante baja, salvo en la provincia de Ourense a partir de la década de los 60. En áreas montañosas, tales como sierra Morena, sierra Nevada y la sierra de Grazalema en el sur peninsular, la densidad espacial de estaciones disponibles es también muy baja. Sin embargo, en Baleares es muy superior ya incluso desde los años 50. El hecho de que haya tanta diferencia de unas zonas a otras, y también a lo largo del periodo para el que se quiere generar la rejilla, impone restricciones en la resolución del análisis, ya que la extensión espacial y la densidad de datos determinan respectivamente la mayor y la menor de las escalas que es posible analizar.

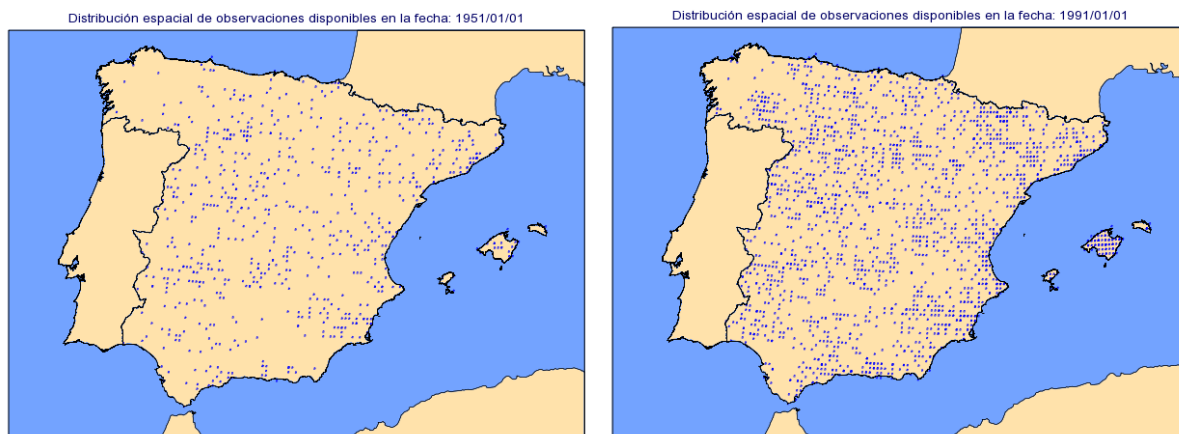


Figura 1: Distribución espacial de las observaciones disponibles en el año 1951 (izquierda), y en el año 1991 (derecha).

En la Figura 2 se representa el porcentaje de estaciones disponibles en los años 80-90 en función del límite máximo de distancia a la estación más próxima, en diferentes zonas geográficas. Se puede observar que, en España peninsular y Baleares, para el 99% de las estaciones existe otra estación a una distancia igual o inferior a 30 km. Sin embargo, si consideramos sólo las islas Baleares, para el 99% de las estaciones se encuentra otra a una distancia menor o igual que 15 km. En Galicia, para conseguir un 99% de observaciones, el límite máximo de distancia a la estación más próxima sube a 40 km.

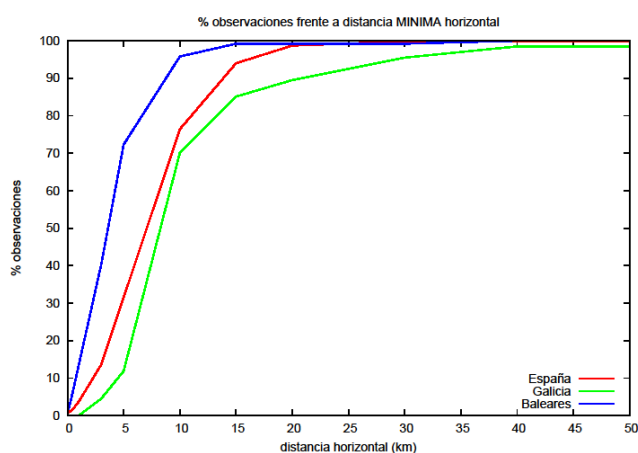


Figura 2: Porcentaje de observaciones cuya distancia a la más próxima es igual o inferior a diferentes umbrales representados en el eje x. España peninsular y Baleares (rojo), Galicia (verde), y Baleares (azul).

## 2.2 Método de análisis

Para cada día, desde el 1 de enero de 1951 hasta la actualidad, se ha ejecutado un sistema de análisis objetivo para producir un campo de precipitación diaria acumulada de 07 a 07 horas en una rejilla cuyas características se describen en la sección 2.6. El sistema empleado se basa en el análisis de superficie integrado en la cadena de predicción numérica HIRLAM, denominado SPAN (Rodríguez et al., 2003, Navascués et al., 2003), que fue ampliado a la variable precipitación (Quintana et al., 2016), y ha sido ajustado en este trabajo para representar las estructuras de precipitación acumulada en 24 horas a las escalas que permite la densidad de estaciones seleccionadas en la España peninsular y Baleares.

El método utilizado es el de Interpolación Óptima, también denominado Interpolación Estadística debido a las aproximaciones e hipótesis asumidas en su implementación. Este método de análisis maneja bien distribuciones irregulares de observaciones, y su algoritmo es el siguiente:

$$f_A(r_i) = f_B(r_i) + \sum_{k=1}^K W_{ik} [f_O(r_k) - f_B(r_k)] \quad (1)$$

donde  $f_A(r_i)$  es el valor analizado del campo  $f$  en el punto de rejilla  $r_i$  (en nuestro caso la variable  $f$  es la precipitación acumulada de 07 a 07 horas),  $f_B(r_i)$  es el valor de una primera estimación de ese campo, o *first guess*, de  $f$  en  $r_i$ ,  $f_O(r_k)$  y  $f_B(r_k)$  son los valores observados y del *first guess* respectivamente en el punto de observación  $r_k$ ,  $K$  es el número total de observaciones y  $W_{ik}$  el peso dado a cada incremento de observación  $[f_O(r_k) - f_B(r_k)]$  en el punto de rejilla  $r_i$  (Daley, 1991).

El método de Interpolación Estadística es, por tanto, un método de interpolación lineal que tiene por objeto construir la mejor estimación a partir de dos tipos de fuentes de información: un *first guess* y un conjunto de observaciones. Cada uno de estos tipos de información tiene unos errores que se suponen conocidos. El análisis resultante busca así filtrar el ruido y extraer la señal de las estimaciones a partir de las que se construye.

Para ello, se impone hacer mínima la varianza del error del análisis. Esta restricción determina  $W$ , la matriz de los pesos de todas las observaciones para todos los puntos de rejilla. Su fila  $i$ ,  $W_i$ , corresponde a los pesos de cada una de las observaciones en el punto de rejilla  $i$ , cuyos elementos son  $W_{ik}$ , y toma la forma:

$$W_i = B H_i^T (H B H^T + R)^{-1} \quad (2)$$

donde  $B$  es la matriz de covarianza de los errores del *first guess*, de dimensión  $n \times n$  (siendo  $n$  el nº total de puntos de rejilla para los que se resuelve el análisis),  $H$  es el operador que interpola los valores del campo  $f$  definido en los puntos de rejilla  $r_i$  a los puntos de observación  $r_k$ , de dimensión  $K \times n$ , y  $R$  es la matriz de covarianza de los errores de las observaciones. De esta forma,  $H B H^T$  es la matriz de covarianza de los errores del *first guess* en los puntos de observación, y por tanto tiene dimensión  $K \times K$ , al igual que  $R$ , la matriz de covarianza de los errores de observación.  $B H_i^T$  es el vector cuyos elementos son la covarianza de los errores del *first guess* entre el punto de rejilla para el que se obtiene el análisis,  $r_i$ , y cada uno de los puntos de observación,  $r_k$ .

El error (mínimo) de la interpolación en cada punto de rejilla  $i$  es el elemento  $ii$  de la diagonal de la matriz de covarianza de los errores del análisis,  $A$ :

$$A = (I - W H) B \quad (3)$$

donde  $I$  es la matriz identidad.

Los elementos más importantes del algoritmo de Interpolación Estadística son las matrices de covarianza de los errores de *first guess*,  $B$ , y de las observaciones,  $R$ , que son las que determinan el peso de cada observación en cada punto de rejilla, y por tanto las que controlan el resultado del análisis. En las estaciones de observación existe normalmente una discrepancia entre el valor observado y el valor estimado por el *first guess*, es lo que se denomina el incremento de observación. El propósito del análisis objetivo es filtrar estos incrementos de observación teniendo en cuenta las matrices de covarianza de los errores del *first guess* y de las observaciones,  $(H B H^T + R)^{-1}$ , para a continuación, extender estos incrementos filtrados a los puntos de rejilla del análisis. El algoritmo de Interpolación



Estadística utiliza para este último paso la estructura espacial de la covarianza de los errores del *first guess* ( $BH_i^T$ ). La ecuación (3) expresa cómo en el proceso de análisis por este método se reduce el error del *first guess* debido a la utilización de las observaciones.

### 2.3 Cajas de análisis

La obtención de los pesos  $W_i$  conlleva la inversión de la matriz  $(HBH^T + R)$  que es de dimensión  $K \times K$ . Si se utilizan todas las observaciones disponibles para producir el análisis en cada punto de rejilla, el tamaño de esta matriz es del orden de  $10^3 \times 10^3$ , lo que hace que el proceso de análisis sea muy costoso computacionalmente. En la práctica, como es habitual en otros análisis objetivos, se establece la hipótesis de que la Interpolación Estadística es local, lo que significa que para cualquier dominio sólo un número limitado de observaciones son importantes al calcular los incrementos de los análisis. Es decir, el dominio se divide en cajas de análisis rectangulares y dentro de cada una de ellas se realiza el análisis de forma independiente. Cada caja de análisis utiliza las observaciones de un área de búsqueda de datos circular, mayor que el área de la caja de análisis, en la que se encuentran todas las observaciones que intervienen en el análisis. El motivo por el cual el área de búsqueda de datos es mayor que la caja es para evitar discontinuidades en el análisis entre cajas contiguas.

En esta implementación del análisis de precipitación se han utilizado cajas de análisis de  $2^\circ \times 2^\circ$  con un radio de búsqueda de observaciones de hasta  $4^\circ$  desde el centro de la caja. Este tamaño de cajas de análisis de  $2^\circ$  es adecuado dada la longitud de escala en la horizontal de los errores del *first guess*, como veremos posteriormente.

### 2.4 Controles de calidad de las observaciones

El algoritmo de Interpolación Óptima supone errores de las observaciones no sesgados y distribuidos normalmente con varianza  $\sigma_o^2$ . Un paso preliminar del análisis consiste, por tanto, en identificar errores groseros no gaussianos mediante controles de calidad de los datos. Este sistema de análisis dispone de dos métodos consecutivos para decidir si las observaciones son o no aceptables.

El primer control es el chequeo frente al *first guess* (**F.G. check**) que limita las observaciones a un subconjunto cuya distancia al *first guess* queda dentro de unos límites definidos por un umbral. Elimina errores groseros en las observaciones. Si se cumple que:

$$\frac{(ob - f.g.)^2}{\sigma_o^2 + \sigma_B^2} > umb_{f.g.} \quad (4)$$

la observación se rechaza (donde  $\sigma_B^2$  es la varianza del error del *first guess*). Puede presentar problemas en casos extremos porque pueden estar rechazándose observaciones correctas que podrían llegar a corregir el *first guess* en el análisis.

Las observaciones que pasan este primer control son sometidas a un chequeo frente a vecinos (llamado **O.I. check**) que rechaza observaciones que no son coherentes con las observaciones de los alrededores. En este chequeo, se obtiene una estimación,  $an$ , del valor analizado en el punto donde está situada la observación a chequear, utilizando todas las observaciones del entorno. La diferencia entre esta estimación y la observación que se chequea se normaliza con la suma de los errores de la observación y de este análisis (que se puede calcular utilizando la ecuación 3). Si se supera un determinado umbral, la observación será tomada en cuenta en el análisis, aunque se marcará como sospechosa, pero si supera un segundo umbral la observación se rechazará. Este control será especialmente bueno en zonas con alta densidad de observaciones (en los que el error del análisis será menor), pero da problemas en zonas con baja densidad de datos, entre otros motivos porque en esas zonas el error del análisis será mucho mayor.

$$\frac{(ob - an)^2}{\sigma_o^2 + \sigma_A^2} > umb_{O.I.} \quad (5)$$

$$\frac{(ob-an)^2}{\sigma_o^2 + \sigma_A^2} > umb2_{O.I.} \quad (6)$$

En este trabajo los valores asignados a los umbrales utilizados en estos controles de calidad son muy elevados, haciendo que estos chequeos estén prácticamente desactivados. Los datos de precipitación utilizados proceden del Banco Nacional de Datos Climatológicos donde ya han pasado diferentes filtros, además de los criterios de completitud y homogeneidad de las series que han conducido a la selección de las estaciones que son utilizadas para generar esta rejilla. Por otro lado, no se desea que observaciones extremas y muy puntuales, por ejemplo debidas a situaciones convectivas relativamente frecuentes en determinadas zonas de nuestro dominio, sean rechazadas.

## 2.5 First guess utilizado

Normalmente, las variables que se analizan en una cadena de predicción numérica del tiempo, tales como por ejemplo, la temperatura o el geopotencial, utilizan como *first guess* una predicción a corto plazo. En el caso de la precipitación, la gran variabilidad que presenta en distintas escalas espaciales y la propia calidad de las predicciones de esta variable en las escalas más pequeñas y fenómenos más extremos, hacen que sea difícil utilizar como *first guess* una predicción de corto alcance del modelo, sobre todo si no se dispone de una red homogénea de observaciones de una densidad suficiente para corregir las escalas más cortas. Por otro lado, no contamos con una serie de predicciones de alta resolución a escala diaria desde 1951, que es el año en que se ha comenzado a generar los análisis diarios de precipitación acumulada en 24 horas a partir de los datos seleccionados del Banco Nacional de Datos Climatológicos de AEMET. Por estos motivos, en el análisis que se ha construido, se ha optado por utilizar como *first guess* un campo nulo, es decir, la primera estimación para el campo de precipitación consiste en que esta variable toma valor cero en todos los puntos de rejilla.

De esta forma, para la estimación de la matriz B de covarianza de los errores del *first guess*, se utiliza la matriz de covarianza de la propia precipitación observada. Evidentemente, el error de este *first guess* tiene un sesgo o bias. Las ecuaciones (1) y (2) producen un análisis no sesgado y minimizan la varianza del error de interpolación si los errores del *first guess* y de las observaciones no están sesgados. Sin embargo, el utilizarlas con un *first guess* con bias generaría un análisis sesgado y el error de interpolación no sería mínimo. Para poder seguir empleando este algoritmo, y producir el menor error de interpolación, se ha estimado el bias del *first guess* y se han amplificado los errores del *first guess* que se especifican en la matriz B, de covarianza de los errores del *first guess* (Dee and Da Silva, 1998).

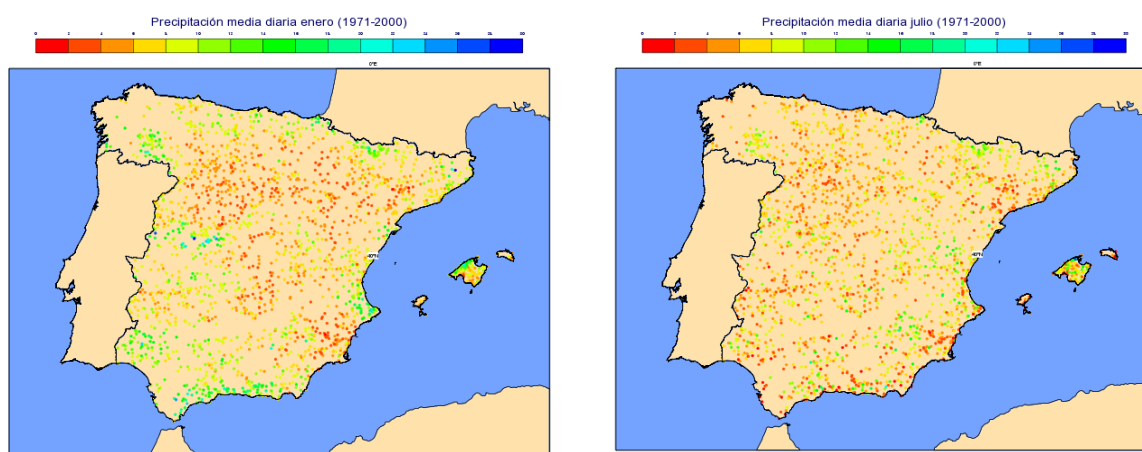


Figura 3: Distribución espacial del bias del *first guess* en las estaciones de observación (que coincide con la precipitación media diaria) en los meses de enero (izquierda), y julio (derecha).

Por la forma en la que se ha construido el *first guess*, su bias es la precipitación promedio en cada punto de observación (asumiendo que el error sistemático de las observaciones es nulo). En la Figura 3 se presenta la distribución espacial de esta precipitación promedio en un mes de invierno y otro de

verano, respectivamente. En ellas se aprecia que la precipitación promedio, y por tanto el sesgo del *first guess*, presentan una variación espacial y estacional.

Como se mostrará posteriormente en el apartado 3.5, la precipitación promedio se ha sumado a la varianza de la precipitación en el modelo de varianza del error del *first guess* en la matriz B asumida. De esta forma, el análisis se genera con las ecuaciones (1) y (2) para minimizar la varianza del error de interpolación, si bien no se puede descartar que el análisis presente algún sesgo al no haberlo tenido en cuenta explícitamente (Dee and Da Silva, 1998). En el apartado 4 se presentan los resultados de la evaluación de la rejilla generada, que incluye este posible sesgo del análisis.

## 2.6 Rejilla del análisis

La rejilla sobre la que se calcula el análisis es la rejilla lat-lon rotada del modelo de predicción numérica del tiempo HIRLAM (Undén et al, 2002), que se ejecuta operativamente por AEMET en alta resolución (modelo HNR) desde hace una década (Navascués et al., 2013). Con esta geometría, la distancia entre puntos de rejilla es aproximadamente constante en el dominio. La resolución del modelo HNR en la rejilla rotada es de  $0.05^\circ \times 0.05^\circ$ , lo que equivale a unos 5 km x 5 km. Es conveniente señalar que aunque el análisis se ejecute en esta rejilla, la resolución de las estructuras de precipitación analizadas es inferior, ya que viene dada por la densidad real de estaciones y el modelo utilizado para describir la matriz de covarianza de los errores del *first guess*, B.

También hay que tener en cuenta que, como solo se han utilizado las observaciones seleccionadas del Banco Nacional de Datos Climatológicos, la rejilla generada solo tiene sentido en el territorio de España. La serie de datos pluviométricos en rejilla generados se ha puesto a disposición de los usuarios solo en los puntos del territorio español peninsular y de Baleares en dos formatos, ascii y netcdf.

## 3 Modelado de las estadísticas de los errores del *first guess* y de las observaciones

Como se ha mencionado, las estadísticas de los errores del *first guess* y de las observaciones son las que determinan las propiedades de filtrado e interpolación del análisis. Para modelarlas, se han asumido una serie de hipótesis y se han elegido unas funciones analíticas con una serie de parámetros, que se han calibrado utilizando el método empleado por Hollingsworth y Lönnberg (1986). Este método está basado en el uso de las series temporales de diferencias entre observaciones y *first guess*, también llamados incrementos de observación o innovaciones:  $[f_o(r_k) - f_b(r_k)]$ .

### 3.1 Modelos utilizados para representar las estadísticas de los errores del *first guess* y de las observaciones

En este análisis de precipitación, asumimos que los errores de las observaciones pluviométricas tomadas en las distintas estaciones, además de no estar sesgados, no están correlacionados entre sí. Ello da lugar a que R sea una matriz diagonal en la que los elementos no nulos son las varianzas de los errores de las observaciones,  $\sigma_o^2$  que suponemos también uniformes. Estos errores de las observaciones incluyen dos componentes, la de la propia precisión de los instrumentos y la de la representatividad de la medida en el análisis que se va a generar.

Para asignar valores a cada uno de los elementos  $B_{ij}$ , que es la covarianza de los errores del *first guess* entre dos puntos i, j, se necesita modelar las varianzas de los errores en cada uno de estos puntos y la función que describe la correlación de estos errores en los mismos, también llamada función de estructura del análisis. La primera versión que se desarrolló del análisis de superficie SPAN, asume para diferentes variables tales como la temperatura o la humedad, una función de correlación homogénea pero anisótropa que da cuenta de la influencia de la topografía. Además, trabaja con la hipótesis de separabilidad de las dependencias en la distancia horizontal (r), y en el desplazamiento

vertical  $\Delta z$ . De esta forma, la función de correlación se expresa como el producto de dos funciones gaussianas: una representa la dependencia en la horizontal y otra la anisotropía inducida por la orografía:

$$\rho(r, \Delta z) = \rho_h(r) \rho_v(\Delta z) \quad (7)$$

$$\rho(r, \Delta z) = e^{-0.5r^2/L_h^2} e^{-0.5\Delta z^2/L_z^2} \quad (8)$$

siendo  $L_h$  y  $L_z$  unos parámetros fijos que controlan la longitud de escala de la función de correlación, en la horizontal y en la vertical respectivamente (Rodríguez et al., 2003).

En este trabajo se ha adoptado para la precipitación las mismas hipótesis, pero la función analítica que expresa la dependencia en la distancia horizontal se ha sustituido por la suma de dos funciones autorregresivas de tercer orden (TOAR), que representan respectivamente las componentes de corta y larga escala de la correlación:

$$\rho_h(r) = a(1 + cr + \frac{c^2 r^2}{3})e^{-cr} + (1 - a)(1 + \frac{cr}{3} + \frac{c^2 r^2}{27})e^{-cr/3} \quad (9)$$

Donde  $a$  es la componente de amplitud y  $c$  el parámetro de escala. Esta función analítica se ha empleado para sustituir el modelo gaussiano en las variables de temperatura y humedad en el análisis de superficie SPAN (Martín and Navascués, 2004), y como se expone en la sección 3.3, presenta un mejor ajuste a los datos empíricos. Otros autores la han utilizado como función de correlación en el análisis objetivo de diferentes variables (Mitchell et al., 1990, Brasnett, 1999, Cansado et al., 2004).

Respecto a la varianza del error del *first guess*, en el modelo que se ha desarrollado se ha implementado una variación espacial de la misma que tiene en cuenta la distinta variabilidad diaria de la precipitación existente en las diferentes regiones españolas. Además, esta varianza incluye una amplificación con la precipitación promedio, debido a utilizar un valor nulo de precipitación en cada punto de rejilla como *first guess*, y que por tanto está sesgado.

Para tener en cuenta la dependencia estacional de las características de las estructuras de precipitación, se ha permitido también que los errores del *first guess* tuvieran una variación mensual. Para ello se han ajustado todos los parámetros que los definen de forma separada para cada mes del año.

Con todo ello, los parámetros que determinan las propiedades del análisis de precipitación incluyen las influencias de la distancia horizontal, la orografía, y la variación espacial, tanto de la precipitación promedio como de su variabilidad diaria. Todas estas dependencias presentan variaciones estacionales, que también son recogidas por los parámetros del análisis.

### 3.2 ***Método para sintonizar los modelos de los errores del first guess y de las observaciones***

El método que se ha empleado para sintonizar los parámetros que definen los errores del *first guess* y de las observaciones sigue el propuesto por Hollingsworth y Lönnberg (Hollingsworth and Lönnberg, 1986). A partir de las innovaciones de la red de radiosondeos, estos autores lo utilizaron para estimar estos errores en el sistema de análisis global por Interpolación Estadística tridimensional de las variables atmosféricas del modelo del Centro Europeo de Predicción a Plazo Medio (ECMWF). Esta metodología ha sido utilizada en multitud de ocasiones posteriormente. El método trabaja con la hipótesis de que los errores del *first guess* están espacialmente correlacionados, mientras que los errores de observación no. También asume que los errores de las observaciones y del *first guess* no están correlacionados entre sí. De esta forma, la covarianza de los errores del *first guess* puede estimarse calculando covarianzas en función de la distancia, a partir de pares de incrementos de observación o innovaciones de estaciones con diferente separación entre sí. La extrapolación de esta relación entre la covarianza y la separación entre las estaciones a una distancia cero proporciona el reparto de la varianza de los incrementos de observación entre una parte correlacionada, la varianza

del error del *first guess*, y una parte no correlacionada, la varianza del error de las observaciones. Una descripción más detallada del método puede encontrarse, por ejemplo, en Daley (1991).

La metodología ha sido aplicada en este trabajo en tres pasos para calibrar el modelo escogido para representar los elementos de B y R:

1. Generación de datos empíricos de varianzas en cada estación y covarianzas de innovaciones entre pares de estaciones.
2. Sintonización de las componentes del modelo de correlación o función de estructura dependientes de la distancia horizontal y de la separación en la vertical.
3. Estimación de las varianzas del error de las observaciones y del *first guess*.

Para la generación de los datos empíricos de partida se han utilizado las series temporales de las observaciones de precipitación acumulada en 24 horas durante 65 años. Con ellos se han calculado las varianzas en cada estación, y las covarianzas entre pares de estaciones de las innovaciones (recordemos que el *first guess* es un campo de precipitación nula). A continuación, se han obtenido las correlaciones empíricas entre cada par de estaciones para cada uno de los meses del año.

### 3.3 *Componente de la función de estructura dependiente de la distancia horizontal*

Los valores de los parámetros  $a$  y  $c$  de la función analítica (9), que representa la dependencia con la distancia horizontal de la correlación de los errores del *first guess*, se han sintonizado de forma independiente para cada uno de los meses del año. Para cada mes, los datos empíricos de las correlaciones entre pares de estaciones tales que su diferencia de altitud sea igual o inferior a 300m, se han agrupado en clases de distancia horizontal. La utilización de pares de estaciones con similar altitud permite filtrar la anisotropía inducida por la orografía, que posteriormente será modelizada. En cada una de las clases de distancia horizontal se ha calculado el promedio de los datos empíricos de correlación entre pares de estaciones de esa clase. De esta forma, se obtiene un conjunto de puntos de correlación en función de la distancia horizontal, que son los que se utilizan para ajustar el modelo analítico de función de correlación.

Como se ha mencionado, la función analítica escogida para ajustar la nube de puntos de las correlaciones de cada clase de distancia horizontal es la suma de dos funciones autorregresivas de tercer orden (TOAR), descrita por la ecuación (9). Estas funciones TOAR caen asintóticamente a cero y la distribución de los datos empíricos presenta una larga cola, con lo cual se ajustan mejor que las funciones gaussianas, como la que aparece como primer factor del segundo miembro de la ecuación (8). En la ecuación (9), la componente horizontal de corta escala está representada por el primer sumando, mientras que el segundo término da cuenta de la dependencia de larga escala.

En la práctica, la función que se ajusta a los datos empíricos es la siguiente:

$$g(r) = \alpha \rho_h(r) = \alpha \left[ a(1 + cr + \frac{c^2 r^2}{3})e^{-cr} + (1 - a)(1 + \frac{cr}{3} + \frac{c^2 r^2}{27})e^{-cr/3} \right] \quad (10)$$

es decir, la función descrita por (9) escalada por un factor  $\alpha$ . Este factor da cuenta del reparto de la varianza de las innovaciones entre la varianza de los errores del *first guess* y la varianza de los errores de las observaciones, ya que su valor es el de la extrapolación de la covarianzas de las innovaciones, que contienen los errores de *first guess* y observaciones, a la distancia nula.

Para ajustar los datos empíricos de las correlaciones de los errores de *first guess* promediados por intervalos de distancia horizontal entre estaciones a la función analítica expresada por (10), se ha utilizado la rutina E04GYF de la librería NAG (Gill and Murray, 1978). La rutina E04GYF se basa en un algoritmo quasi-Newtoniano para encontrar el mínimo de una suma de cuadrados de  $m$  funciones no lineales o residuos en  $p$  variables, que son los parámetros que definen la función  $g(r)$ . Los tres

parámetros que determinan el mejor ajuste a la función (10) son la intersección con el eje de ordenadas  $\alpha$ , que nos indica que porcentaje de la varianza total corresponde al error del *first guess*, la componente de amplitud  $a$ , y el parámetro de escala  $c$ .

La Tabla 1 muestra los valores obtenidos para  $a$ ,  $c$  y  $\alpha$  para los distintos meses del año. En ella se puede observar que las propiedades de filtrado del análisis, con las funciones de estructura que se han sintonizado a partir de los datos empíricos, varían a lo largo del año. La componente de larga escala de la correlación ( $1-a$ ), es siempre mayor que la de pequeña escala,  $a$ . Sin embargo,  $a$  tiene un mayor valor en los meses de julio y agosto, y menor de noviembre a enero.

La Figura 4 muestra los valores de correlaciones promedio por cada clase de distancia (puntos) obtenidos para el mes de diciembre, y el ajuste de los mismos a la función (10) (curva verde), así como sus componentes de corta y larga escala (curvas azul y rosa respectivamente).

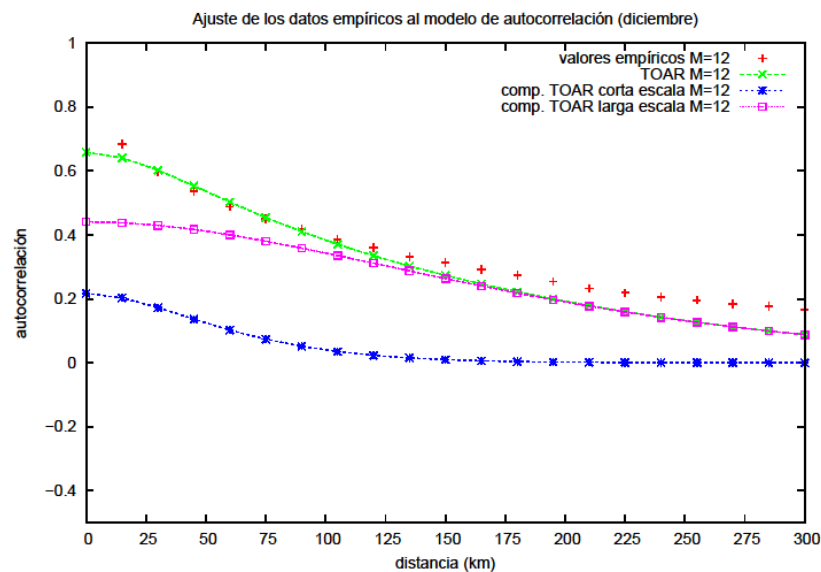


Figura 4: Ajuste de los datos de correlaciones de las innovaciones para el mes de diciembre (puntos en rojo) a la suma de funciones TOAR (verde) con la que se modela la función de correlación de los errores del *first guess*. En azul y rosa las funciones TOAR que representan las componentes de corta y larga escala de la función de correlación de los errores del *first guess*.

Un parámetro para caracterizar las correlaciones en las distancias más cortas, que es muy sensible al comportamiento espectral del análisis en las escalas más pequeñas, es la longitud de escala característica del análisis, o longitud de microescala,  $L$ , que se define como (Daley, 1991):

$$L = \left(-\frac{2\rho_h}{\Delta\rho_h}\right)_{r=0} \quad (11)$$

donde  $\Delta$  es el operador laplaciano.  $L$  es un indicador de cómo es de abrupta la función de correlación cerca del origen, es decir, cómo disminuye con la distancia desde su valor igual a 1 en el origen. La cuarta columna de la Tabla 1 presenta los valores de  $L$ , que han sido calculados teniendo en cuenta los valores de los parámetros  $a$  y  $c$  obtenidos en los diferentes meses del año. Las longitudes de escala características oscilan entre 40 km en julio, y 69 km en diciembre. Estos valores están relacionados con el diferente carácter de la precipitación en las distintas estaciones del año, más convectiva y de pequeña escala en verano. Es por tanto en esta estación en la que el análisis va a producir estructuras de mayor resolución, siempre que esté presente en las observaciones y se cuente con una densidad de ellas suficiente.

Meses	$a$	$c$	$L$ (km)	$\alpha$
M=1	0.3349	0.0399	67.9	0.6541
M=2	0.3467	0.0416	64.3	0.6456
M=3	0.3539	0.0426	62.4	0.6571
M=4	0.3818	0.0463	55.8	0.6390
M=5	0.4056	0.0503	50.1	0.5969
M=6	0.4077	0.0518	48.6	0.5470
M=7	0.4570	0.0596	40.4	0.5316
M=8	0.4331	0.0562	43.7	0.5579
M=9	0.3683	0.0456	57.3	0.6122
M=10	0.3496	0.0437	61.0	0.6439
M=11	0.3273	0.0400	68.2	0.6676
M=12	0.3301	0.0392	69.5	0.6579

Tabla 1: Valores de los parámetros de ajuste  $a$ ,  $c$  y  $\alpha$ , así como las correspondientes longitudes de escala característica  $L$  en los diferentes meses del año (filas).

La Figura 5 muestra las funciones de correlación obtenidas para los meses de diciembre (en rojo) y julio (en verde), en las que se puede observar tanto el menor valor de la longitud de escala característica como el mayor peso de la componente de corta escala en verano. En esta figura también se ha representado la función gaussiana que tendría la misma longitud de escala característica que la función TOAR en julio (curva azul). Estas curvas se solapan cerca del origen, pero la función gaussiana cae a cero mucho más rápidamente y por tanto produce un peor ajuste de los valores de correlación significativo que aún persisten a distancias más largas.

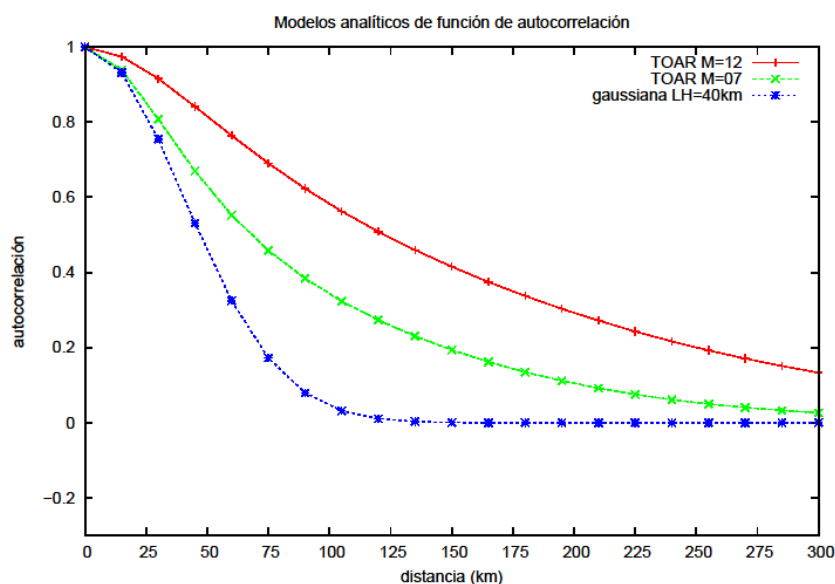


Figura 5: Funciones de correlación de los errores del *first guess* obtenidas en los meses de diciembre (rojo) y julio (verde). Gaussiana con la misma longitud de escala característica que la función TOAR obtenida para el mes de julio (azul).

Para analizar la posible dependencia geográfica de la función de correlación de los errores del *first guess*, se ha repetido este procedimiento de ajuste utilizando otros dos conjuntos de estaciones: el formado únicamente por aquellas situadas al este de  $-2^{\circ}\text{E}$ , para representar la zona mediterránea, y el constituido por las estaciones que se encuentran al oeste de  $-2^{\circ}\text{E}$ , para la región atlántica. La Figura 6 muestra los valores obtenidos para la longitud de escala característica a lo largo del año para estos dos

conjuntos de datos, así como para el formado por la totalidad de las estaciones pluviométricas (naranja). Se observa que, salvo en los meses de mayo, junio y agosto, los valores de  $L$  en la zona mediterránea son inferiores a los de la zona atlántica, sobre todo en el otoño. Estos resultados indican una mayor componente de corta escala de las estructuras de precipitación en el área mediterránea, si bien en el análisis que se ha implementado se han utilizado como funciones de correlación las obtenidas del ajuste a los datos empíricos de todas las estaciones disponibles en la península ibérica y Baleares.

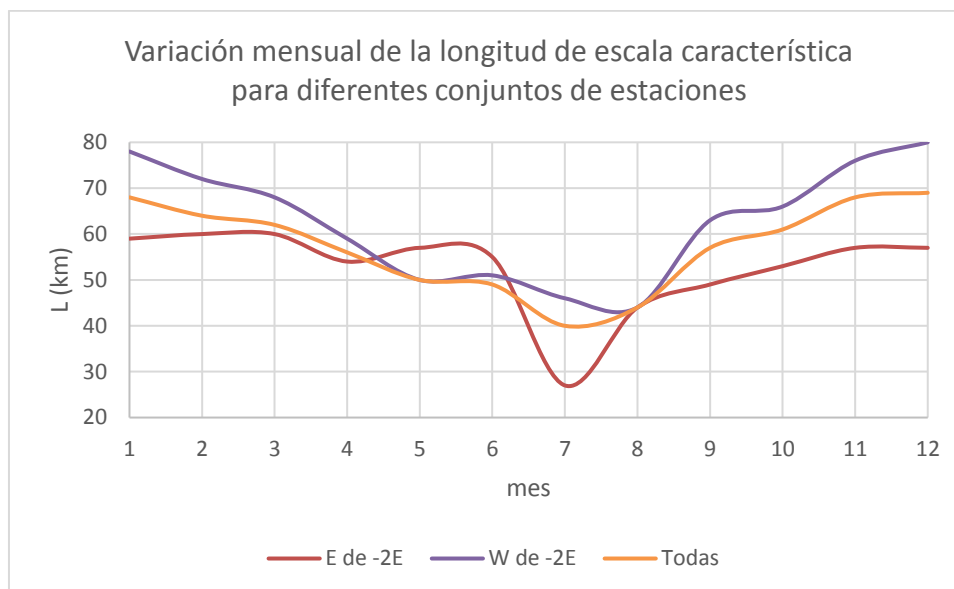


Figura 6: Variación mensual de la longitud de escala característica obtenida por el ajuste al modelo escogido de función de correlación de los errores del *first guess* de tres conjuntos de datos diferentes: el formado por las correlaciones entre todas las estaciones disponibles (naranja), el constituido por las correlaciones entre las estaciones situadas al este de  $-2^{\circ}\text{E}$  (marrón), y el formado por las correlaciones entre las estaciones que se encuentran al oeste de  $-2^{\circ}\text{E}$  (azul).

### 3.4 Modelado de la anisotropía debida a la orografía

Para estimar la dependencia de las correlaciones de los errores del *first guess* con la distancia vertical, el conjunto de datos de correlaciones de innovaciones se ha dividido en cuatro clases de acuerdo a la diferencia de altitud ( $\Delta z$ ) entre cada par de estaciones:

Clase 1:  $\Delta z \leq 300$  m

Clase 2:  $300 \text{ m} < \Delta z < 600$  m

Clase 3:  $600 \text{ m} < \Delta z < 900$  m

Clase 4:  $\Delta z \geq 900$  m

En cada una de estas clases de distancia vertical se han creado los mismos intervalos de distancia horizontal utilizados como se ha dicho anteriormente para la clase 1, en los que se promediaron los valores de las correlaciones de las innovaciones entre pares de estaciones. Dividiendo el valor de correlación promediado en el intervalo de distancia horizontal correspondiente entre el de la clase vertical 1 obtenemos un nuevo conjunto de datos de correlaciones normalizadas en función de la distancia horizontal, para cada clase vertical 2, 3, y 4 (ver Figura 7). En esta figura se observa cómo disminuye la correlación conforme aumenta la distancia en la vertical. También se aprecia que las curvas obtenidas están estratificadas; para cualquier clase de distancia en la vertical, la correlación en la vertical apenas varía para los pares de estaciones situados a diferentes distancias en la horizontal. Este hecho apoya la hipótesis de separabilidad en la formulación analítica de la correlación, como el producto de las componentes de distancia horizontal y de diferencia de altitud.



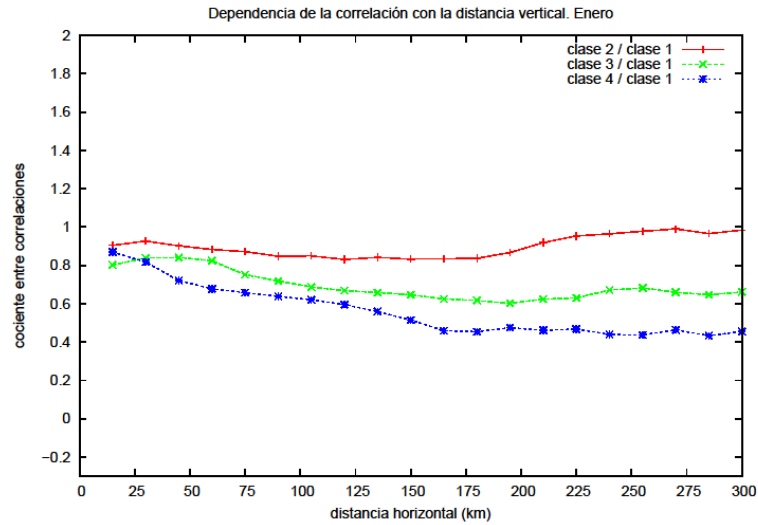


Figura 7: Dependencia de la correlación de los errores del *first guess* con la diferencia de altitud: en cada intervalo de distancia horizontal se representan las correlaciones empíricas entre pares de estaciones de distintas clases de distancia en la vertical (clases 2, 3 y 4 correspondientes a curvas roja, verde y azul) normalizadas por la correlación de la clase 1 de distancia vertical.

Las correlaciones normalizadas así obtenidas para cada uno de los intervalos de distancia horizontal, se han promediado hasta obtener un valor por cada clase vertical. El conjunto de estos valores para las clases verticales consideradas, forman los datos de partida para ajustar la dependencia en la vertical de las correlaciones a una función gaussiana:

$$\rho_v(\Delta z) = e^{-0.5\Delta z^2/L_z^2} \quad (12)$$

El ajuste se ha realizado para cada mes del año. Los valores de longitud de escala  $L_z$  que se han obtenido oscilan entre 1000-1700 m. Los más bajos/altos, asociados a una más rápida/más lenta disminución de la correlación en la vertical, ocurren en invierno/verano. Finalmente, el parámetro de longitud de escala en la vertical se promedió para todos los meses resultando ser de 1.3 km, siendo este valor el que se ha adoptado de forma constante a lo largo del año.

### 3.5 Varianzas del error del *first guess* y de las observaciones

En el apartado 3.3 se ha expuesto cómo el ajuste de las correlaciones empíricas de las innovaciones a una función analítica proporciona también el reparto de la varianza de estos incrementos de observación entre la varianza de los errores del *first guess* y de las observaciones.

La quinta columna de la Tabla 1,  $\alpha$ , presenta el porcentaje de la varianza de las innovaciones que corresponde a la varianza del error del *first guess*, que oscila a lo largo del año, entre un 53% en julio, y un 67% en noviembre. Estos porcentajes tan modestos están relacionados con un relativamente elevado error de las observaciones, principalmente debido a su componente de representatividad, que se pone aún más de manifiesto en verano.

En anteriores trabajos, para el análisis de la precipitación (Quintana et al., 2016) se habían adoptado valores constantes para las varianzas de los errores de *first guess* y de observación. Esta era una consideración muy grosera teniendo en cuenta la distribución espacial tan irregular que tiene la variabilidad diaria de la precipitación en España, que también presenta una variación mensual. En esta nueva versión del análisis, se ha generado un campo de varianzas del error del *first guess* que tenga en cuenta esta variabilidad espacial y estacional.

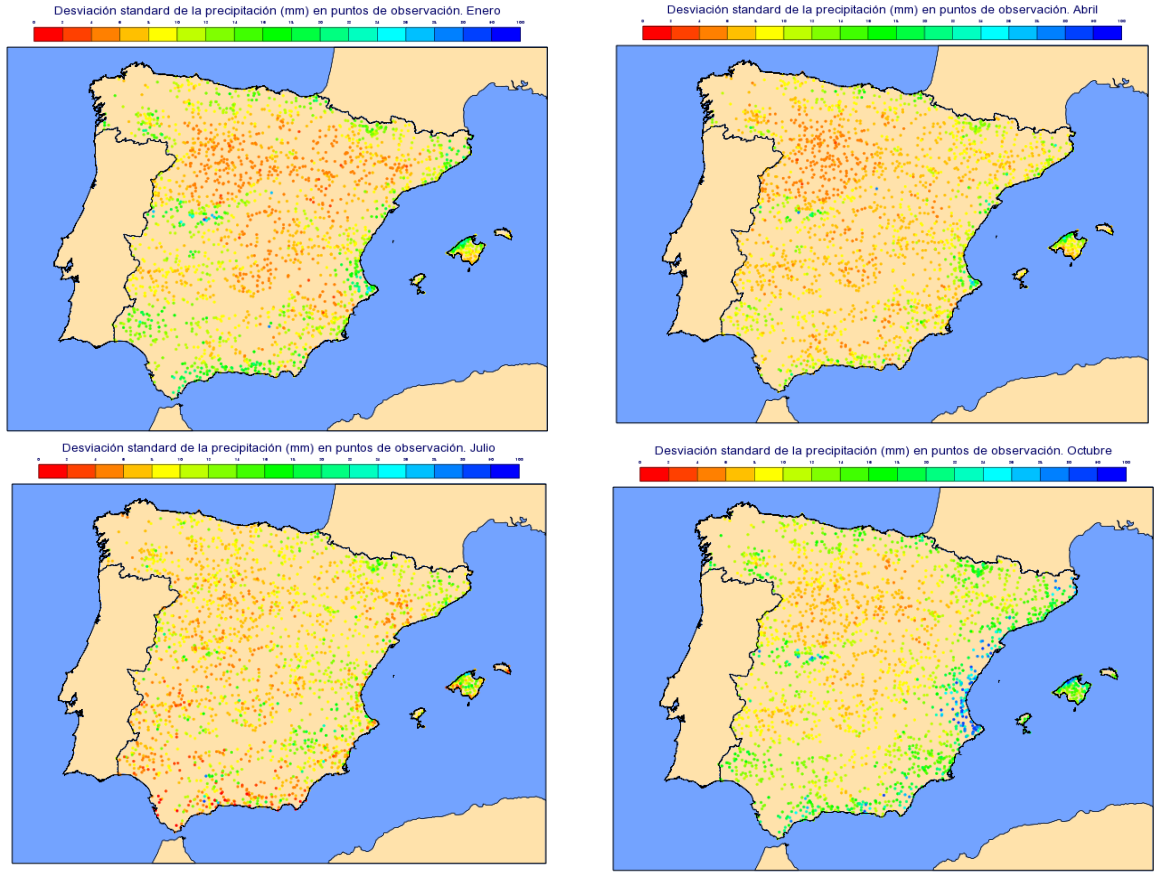


Figura 8: Desviación estándar de la precipitación (mm) en las estaciones pluviométricas para los meses de enero (superior izquierda), abril (superior derecha), julio (inferior izquierda) y octubre (inferior derecha).

La Figura 8 muestra la desviación standard de la precipitación acumulada en 24 horas en las estaciones pluviométricas utilizadas para diferentes meses del año. Esta figura permite apreciar la variabilidad espacial y estacional de la precipitación. Como se observa, los valores máximos se producen en Levante en el mes de octubre, aunque también son elevados en otros lugares montañosos, como en la sierra de Gredos en enero y octubre. Para obtener una estimación del error del *first guess* se procedido de la siguiente forma. En primer lugar, la varianza de las innovaciones en cada estación (que por utilizar un *first guess* de valor nulo es la varianza de la propia precipitación), se ha escalado con el valor del parámetro  $\alpha$  de ajuste de la función  $g(r)$  utilizada para la componente horizontal del modelo de correlación. A continuación, en cada estación este valor se ha amplificado sumando el cuadrado del promedio de la innovación,  $\overline{pcp}_{obs}^2$ , para tener en cuenta el sesgo del *first guess*:

$$\sigma_{fg_{obs}}^2 = \alpha \sigma_{pcp_{obs}}^2 + \overline{pcp}_{obs}^2 \quad (13)$$

A partir de estas varianzas amplificadas en cada estación, se ha generado un campo de varianza de los errores del *first guess* para cada mes en cada uno de los puntos de la rejilla con un algoritmo de análisis de Barnes. Las características de este sencillo método de análisis pueden consultarse en Daley (1991). De esta forma, la varianza del *first guess* en un punto de rejilla  $i$  es:

$$\sigma_{fg_i}^2 = \frac{1}{\sum_{obs=1}^{obs=N} W_{iobs}} \sum_{obs=1}^{obs=N} W_{iobs} \sigma_{fg_{obs}}^2 \quad (14)$$

siendo  $obs$  el índice que recorre todas las observaciones y  $W_{iobs}$  el peso gaussiano que tiene cada una de las observaciones con respecto al punto de rejilla  $i$ , y que en este análisis de Barnes viene dado por una función analítica:

$$W_{iobs} = e^{-0.5r^2/L_B^2} \quad (15)$$

con  $r$  la distancia en km entre el punto de rejilla y cada uno de los puntos de observación y  $L_B=30$  km. Esta distancia  $L_B$  ha sido escogida empíricamente teniendo en cuenta la distancia mínima entre estaciones en las zonas con menor densidad de datos.

La Figura 9 muestra los campos de desviación standard del *first guess* contruidos de esta forma para los meses de enero, abril, julio y octubre.

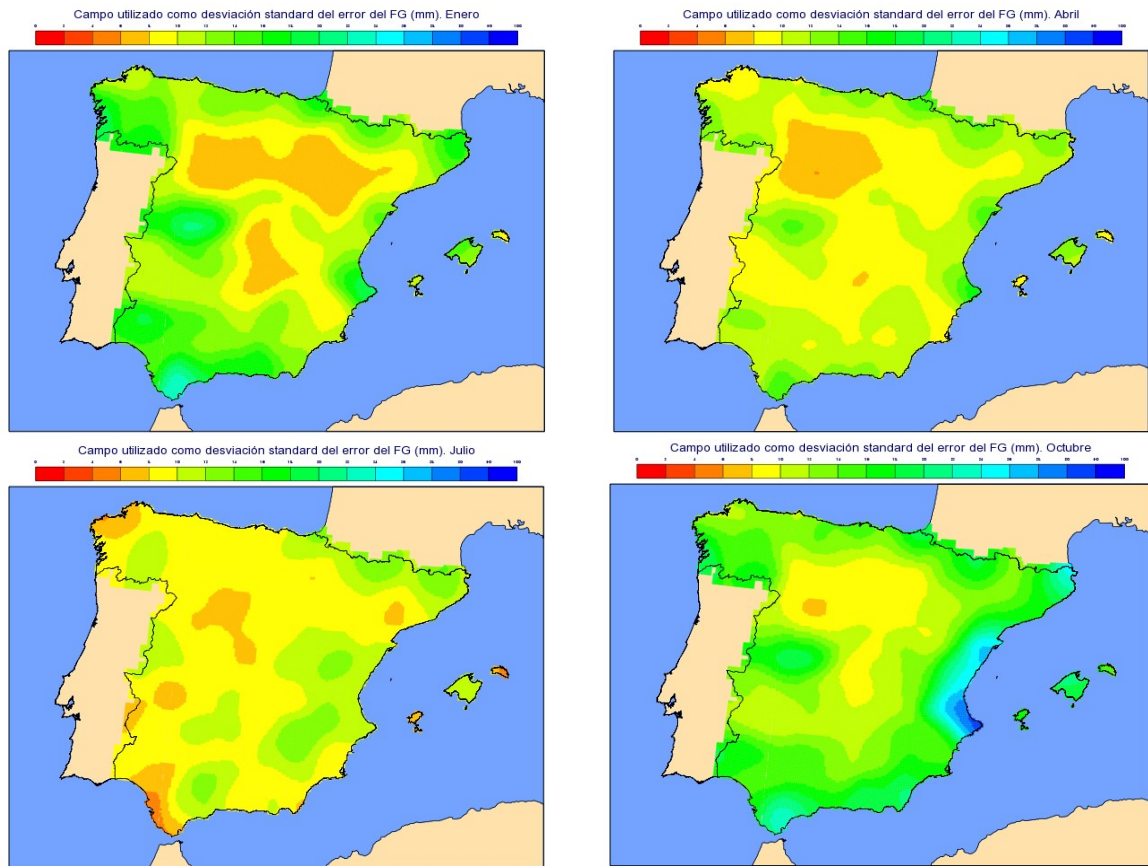


Figura 9: Campo de desviación standard del error del *first guess* (mm) utilizado en el análisis para los meses de enero (superior izquierda), abril (superior derecha), julio (inferior izquierda) y octubre (inferior derecha).

Tras construir el campo de varianza de los errores del *first guess*, se ha procedido a calibrar la varianza de los errores de las observaciones. A pesar de que los resultados del ajuste de los parámetros de la componente horizontal del modelo de correlación indican que los errores de observación son ligeramente mayores en los meses de verano, y que presentan variaciones en las diferentes áreas geográficas, en la versión que se ha desarrollado se ha implementado una varianza del error de observación uniforme y constante. Para estimarla, la varianza de las innovaciones calculada con todas las estaciones de cada mes se ha escalado con el factor  $(1 - \alpha)$ , y se ha obtenido el valor promedio de todos los meses del año. De esta forma, el valor adoptado para la desviación estándar del error de observación es  $\sigma_0 = 4$  mm.

## 4 Serie de precipitación diaria en rejilla desde 1951: Diagnósticos y validación

Una vez se han sintonizado las estadísticas de los errores del *first guess* y de las observaciones, que permiten construir los elementos de las matrices B y R, se ha procedido a ejecutar el análisis diario de precipitación desde el 1 de enero de 1951 hasta la actualidad. La evaluación de los campos generados se ha hecho de varias formas: analizando los propios diagnósticos del análisis, verificando los campos de precipitación obtenidos frente a observaciones independientes que no han sido utilizadas en el análisis, y obteniendo estadísticos que han sido comparados con los de las propias observaciones y con los calculados con la rejilla Spain02.

### 4.1 Diagnósticos del análisis

El paquete de análisis SPAN incluye el cálculo, proceso y archivo para cada una de las observaciones de las diferencias *observación menos first guess* (innovaciones) y *observación menos análisis* (residuos). A partir de estas series de innovaciones y residuos se calculan histogramas que permiten comprobar el comportamiento del análisis.

La Figura 10 muestra estos histogramas para los meses de enero y octubre. En ella se observa que, al ser el *first guess* un campo de precipitación igual a cero, la distribución de innovaciones está sesgada, con frecuencias no nulas únicamente para valores positivos de las innovaciones. Tras el análisis, los residuos se reparten de forma bastante simétrica alrededor del residuo igual a cero, al tiempo que la distribución se estrecha, indicando un análisis no sesgado a una distancia de las observaciones menor que el *first guess*.

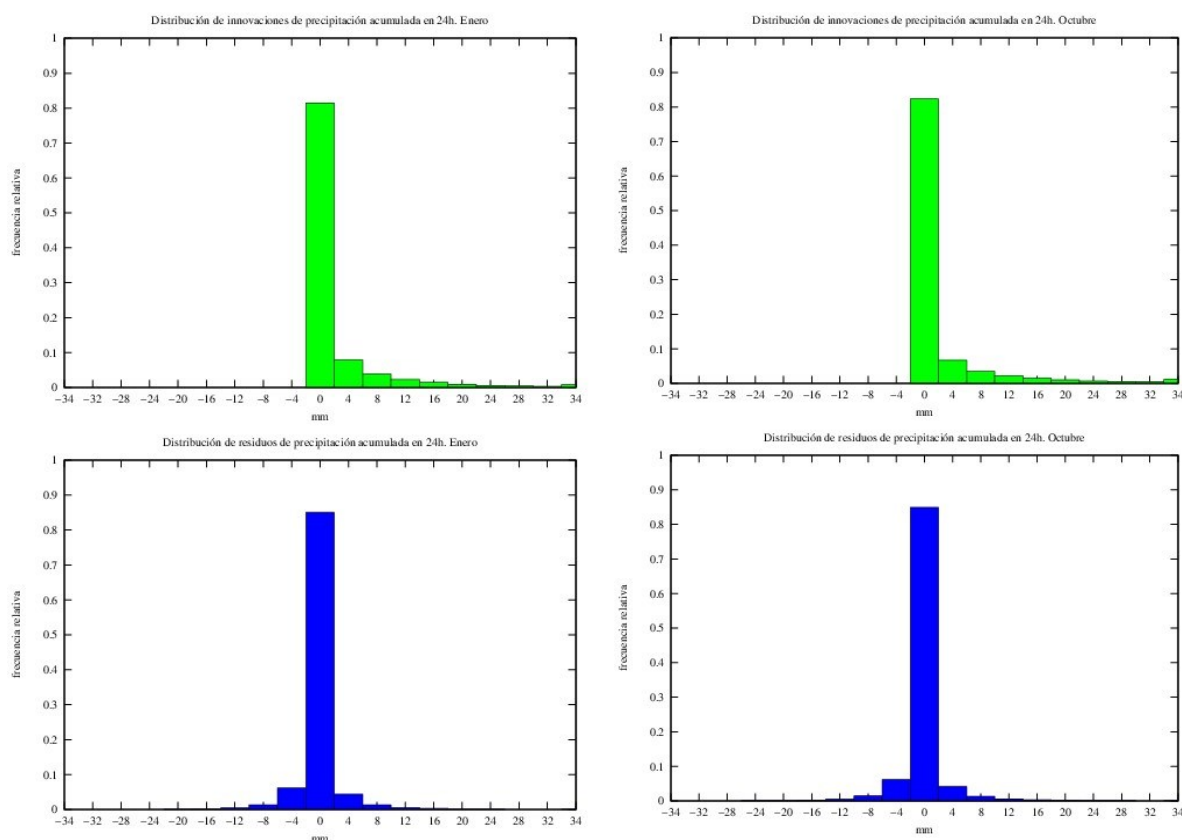


Figura 10: Histogramas de innovaciones y residuos del análisis en el mes de enero (izquierda) y octubre (derecha).

La Figura 11 presenta de otra forma el ajuste de las observaciones al *first guess* y al análisis para todo el periodo para el que se ha ejecutado el análisis. En ella se representa para cada mes del año, el promedio de las innovaciones y residuos, y la raíz cuadrada del promedio de sus cuadrados (rms). La gráfica muestra cómo la distancia inicial de las observaciones al *first guess* se reduce tras el análisis. También se observa que el análisis no presenta sesgo respecto a las observaciones que ha utilizado. La distancia de las observaciones al análisis es inferior al error de las observaciones que se ha asumido (4 mm). La variación a lo largo del año del bias del *first guess* (debido a la variación estacional de la precipitación) desaparece tras el análisis. La variación mensual del rms del análisis es también significativamente más pequeña que la del *first guess*.

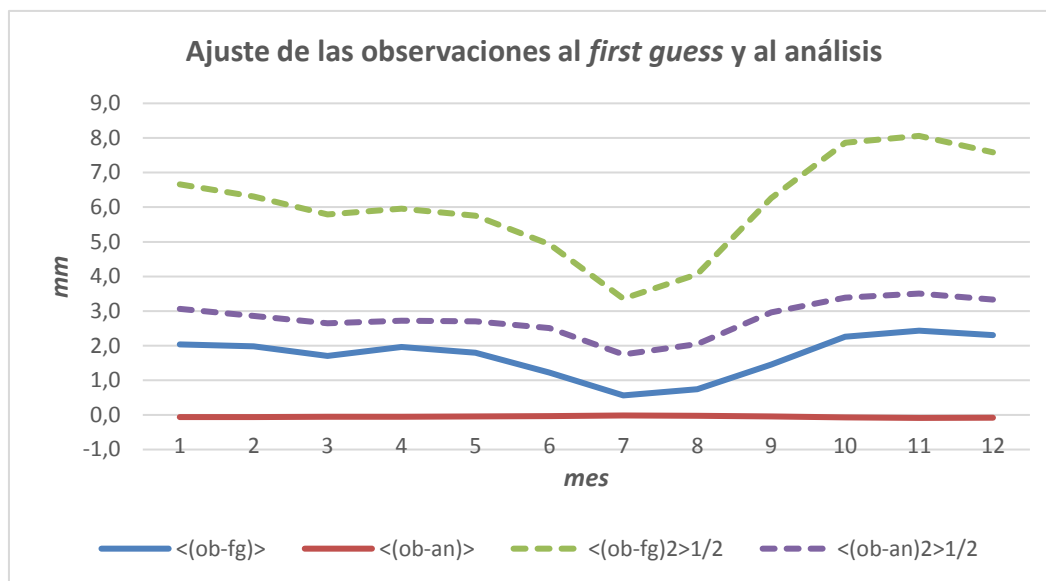


Figura 11: Variación mensual del promedio de innovaciones (curva azul continua), del promedio de residuos del análisis (curva roja continua), del rms de las innovaciones (curva verde discontinua) y del rms de los residuos (curva morada discontinua).

## 4.2 Validación frente a estaciones independientes

Como parte de la evaluación de los análisis de precipitación que se han generado, se ha incluido una verificación de los mismos frente a observaciones pluviométricas de precipitación acumulada en 24 horas de la red de estaciones de medida de contaminación atmosférica de fondo EMEP/VAG/CAMP, y del Banco Nacional de Datos Climatológicos, que no han sido utilizadas en el análisis. Esto es debido a que no se encontraban entre las seleccionadas para la generación de escenarios regionalizados de cambio climático, entre otros motivos, por la escasa longitud o falta de homogeneidad de las series. Sin embargo, estas series pueden ser utilizadas para validar los análisis diarios de precipitación frente a observaciones independientes.

Con este propósito, se han escogido un conjunto de 44 estaciones independientes distribuidas a lo largo de la España peninsular y Baleares. Adicionalmente, y con objeto de analizar el impacto de la baja densidad de observaciones utilizadas por el análisis en Galicia, se ha seleccionado un conjunto adicional de 20 estaciones independientes en esta Comunidad Autónoma. Para cada uno de los datos disponibles de las series de estos dos conjuntos de estaciones, se ha buscado el valor analizado del punto de rejilla más cercano, se ha comparado con el dato de observación, y se han calculado el coeficiente de correlación, el error promedio o bias, y la raíz cuadrada del promedio de los cuadrados del error diario o rmse (*root mean square error*). Los resultados obtenidos para el conjunto de España y para Galicia se muestran en las Figuras 12 y 13, respectivamente.

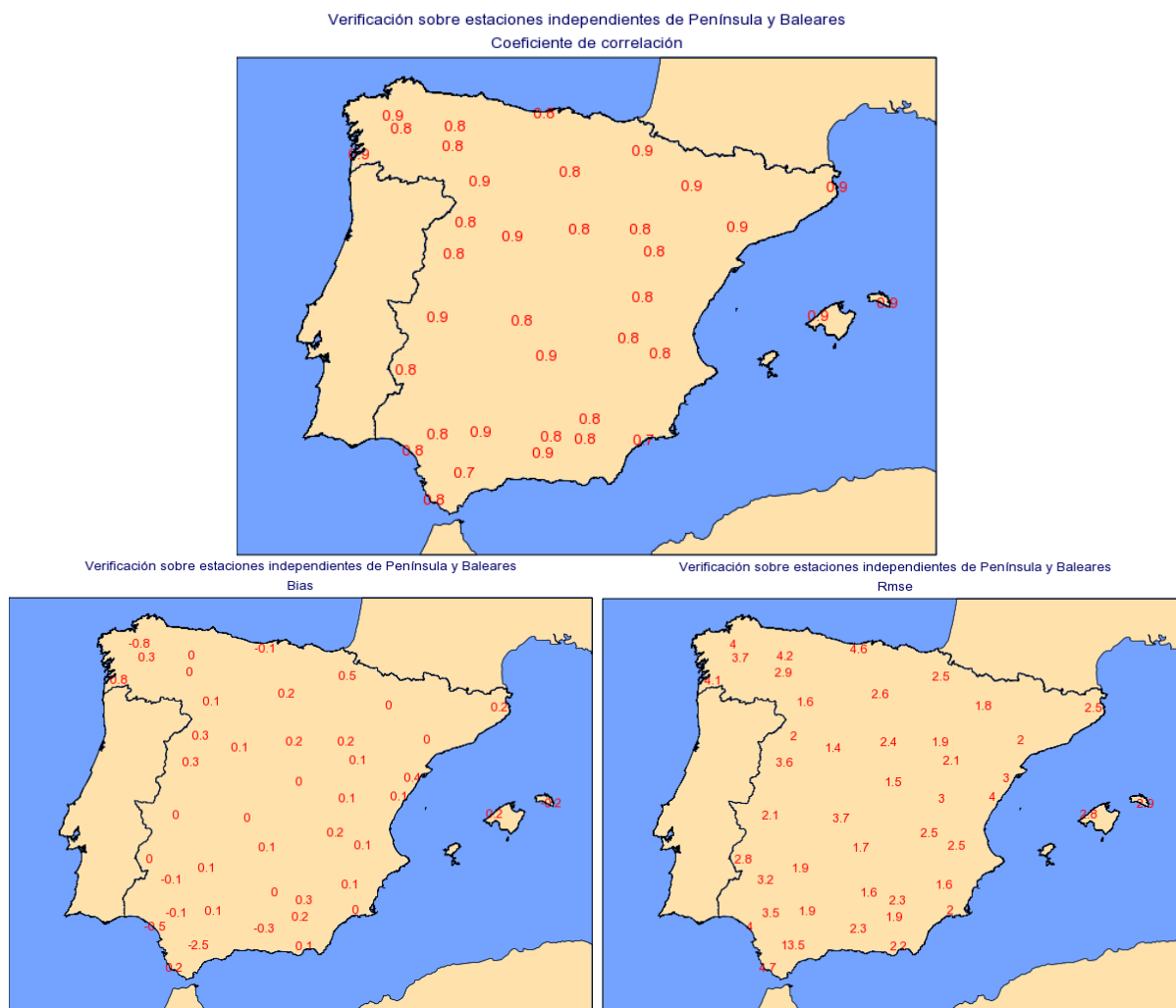


Figura 12: Verificación del análisis de precipitación frente a estaciones independientes: mapas con coeficiente de correlación (superior), bias (inferior izquierda) y rmse (inferior derecha).

Como se aprecia en la Figura 12, los resultados de la verificación indican valores muy elevados de correlación entre el análisis de precipitación diaria y las observaciones pluviométricas de estaciones independientes (0.8 o superior). En puntos aislados, como la sierra de Grazalema en el sur peninsular, la correlación disminuye hasta 0.7. El bias es generalmente pequeño, alcanzando el valor más alto también en Grazalema, donde el rmse asimismo es más elevado que en el resto de estaciones, en las que normalmente no alcanza los 4 mm. Los resultados obtenidos en Grazalema pueden explicarse por la alta variabilidad espacial de la precipitación en esta zona montañosa de España, unido a la ausencia de suficientes estaciones representativas en esta área, entre las seleccionadas para realizar el análisis.

En el conjunto de las 20 estaciones independientes seleccionadas en Galicia (Figura 13), la correlación con el análisis sigue siendo elevada, predominando los valores entre 0.8-0.9 salvo en zonas del interior oriental de la Comunidad y en la Costa da Morte, donde la correlación baja hasta valores de 0.7. Los valores de bias en valor absoluto y de rmse son superiores a los que se encuentran en el resto de la península ibérica, con las mayores subestimaciones de la precipitación por el análisis en áreas muy lluviosas entre las rías de Muros y Noia y de Arousa, aunque en esta zona la correlación alcanza 0.8. Hay que tener en cuenta que la estación más cercana a las mismas que ha sido utilizada por el análisis es Santiago de Compostela.



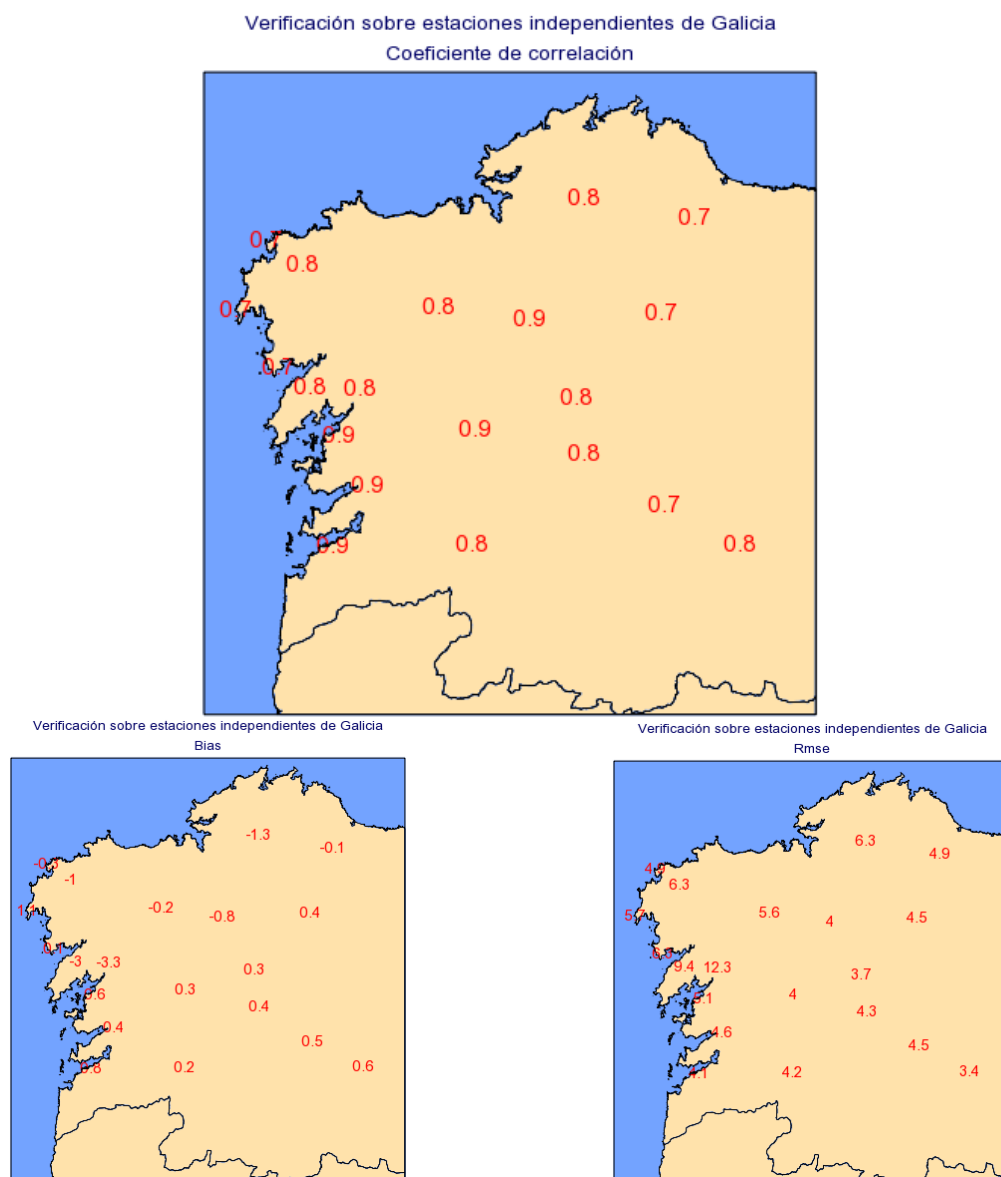


Figura 13: Verificación del análisis de precipitación frente a estaciones independientes en Galicia: mapas con coeficiente de correlación (superior), bias (inferior izquierda) y rmse (inferior derecha).

#### 4.3 Estadísticas climatológicas obtenidas con la precipitación en rejilla

A partir de la precipitación diaria en rejilla se han obtenido una serie de estadísticos climatológicos, cuya comparación con los obtenidos a partir del conjunto de estaciones pluviométricas con las que se ha generado, contribuye también a su validación. En este apartado se presentan como ejemplos los valores de los terciles superior e inferior de la precipitación en invierno (Figura 14), así como la clasificación en terciles de los trimestres de invierno, primavera, verano y otoño de un año concreto (Figura 15).

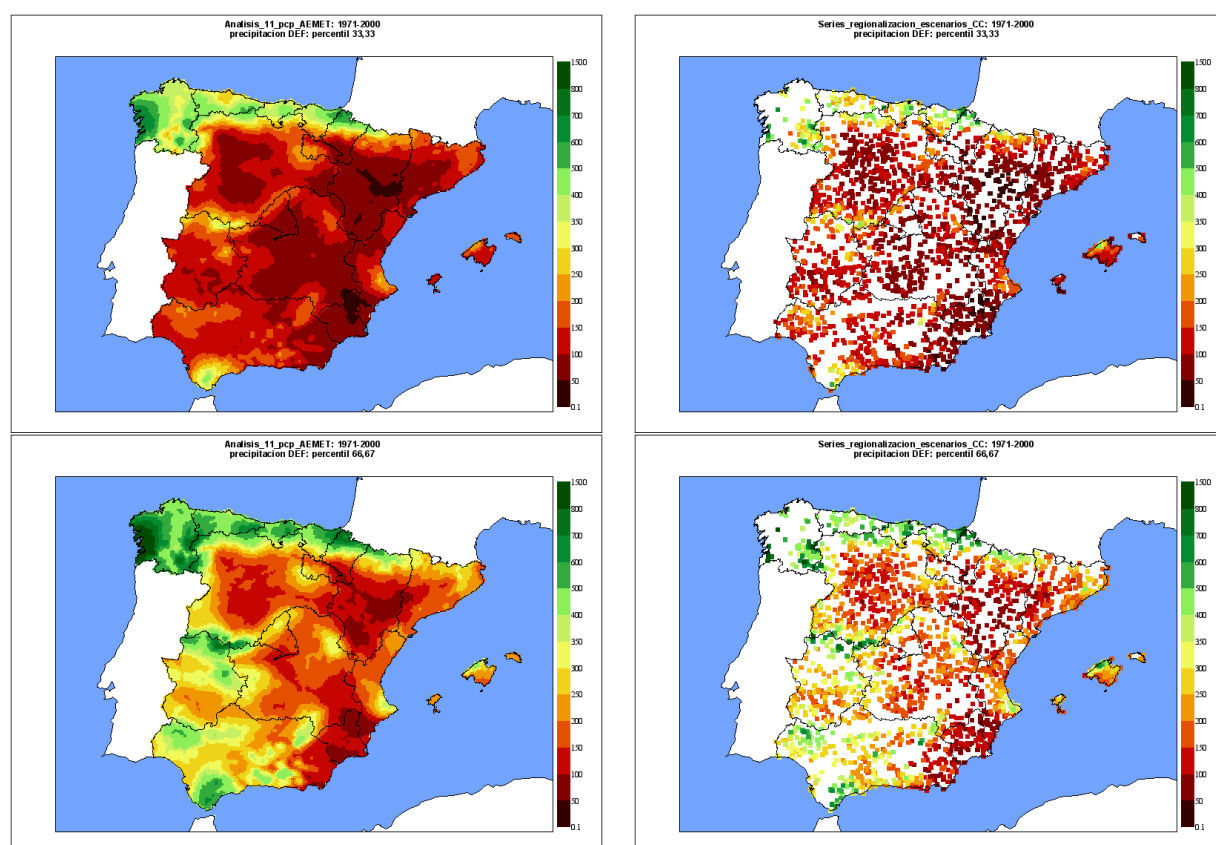


Figura 14: Terciles inferior (arriba) e superior (abajo) de la precipitación en invierno utilizando los datos de precipitación en rejilla (izquierda) y el conjunto de observaciones utilizado para generarla (derecha). Periodo de referencia 1971-2000.

La Figura 14 muestra un acuerdo cualitativo muy alto entre los valores de los terciles de precipitación utilizando los datos en rejilla y las observaciones, tanto en la estructura y distribución espacial como en la magnitud de estos valores, si bien esta es ligeramente inferior en algunas zonas, como la sierra de la Tramontana en Baleares. Como se muestra en la Figura 15, la utilización de los datos de la rejilla conduce a la misma clasificación de la precipitación en las distintas zonas geográficas que la que se obtiene utilizando los datos de observación.



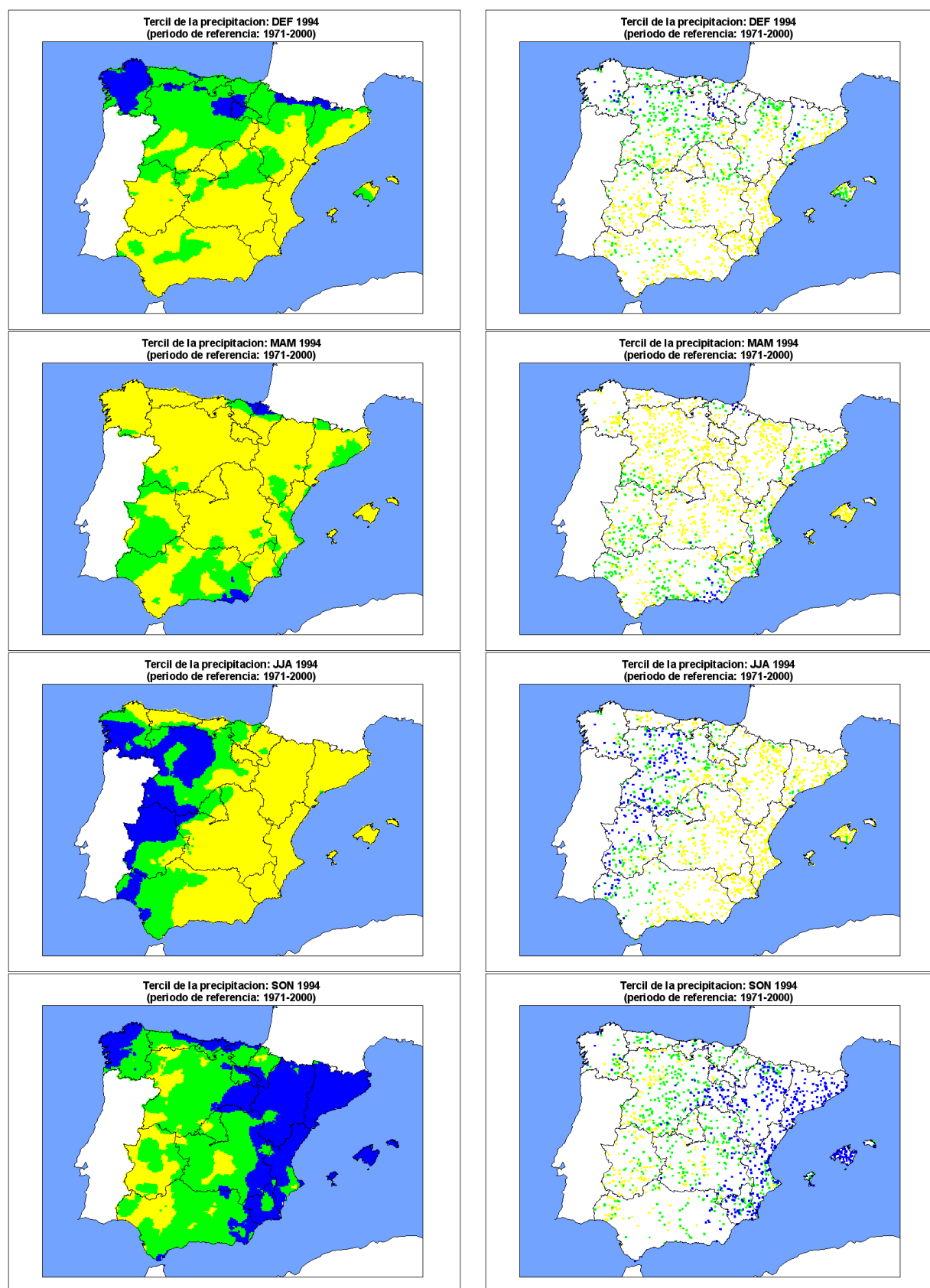


Figura 15: Tercil de la precipitación en el año 1994 (húmedo/normal/seco en azul/verde/amarillo) de los trimestres: diciembre-enero-febrero (DEF), marzo-abril-mayo (MAM), junio-julio-agosto (JJA) y septiembre-octubre-noviembre (SON), de acuerdo a los datos de precipitación en rejilla (izquierda) y al conjunto de observaciones utilizado para generarla (derecha). Período de referencia: 1971-2000.

Cuando se analizan los episodios más extremos las diferencias entre los valores producidos por la rejilla de precipitación y por las observaciones aumentan. La Figura 16 muestra los valores máximos de la precipitación a través del percentil 90 de los dos conjuntos de datos en el periodo de referencia 1971-2000 para los trimestres de invierno, primavera y otoño. Se puede apreciar que los patrones espaciales de los valores máximos de la precipitación se recogen muy bien por parte de la rejilla, así como la variación estacional de los mismos. Sin embargo, los valores absolutos de los máximos de la rejilla son inferiores a los producidos por las observaciones (ver por ejemplo, la zona de la sierra de la Tramontana en la isla de Mallorca). No se puede esperar que la rejilla reproduzca los valores pluviométricos puntuales recogidos en las estaciones. El análisis construye un valor de precipitación basado en todas las observaciones disponibles y en un *first guess* de valor nulo, teniendo en cuenta las características asignadas a los errores de estos dos tipos de estimaciones. En concreto, las funciones de estructura del análisis son las que determinan las escalas espaciales que se resuelven. Los máximos de las series pluviométricas ocurren en situaciones en las que los patrones de precipitación presentan un elevado gradiente espacial, y por tanto, son más filtradas por el análisis.

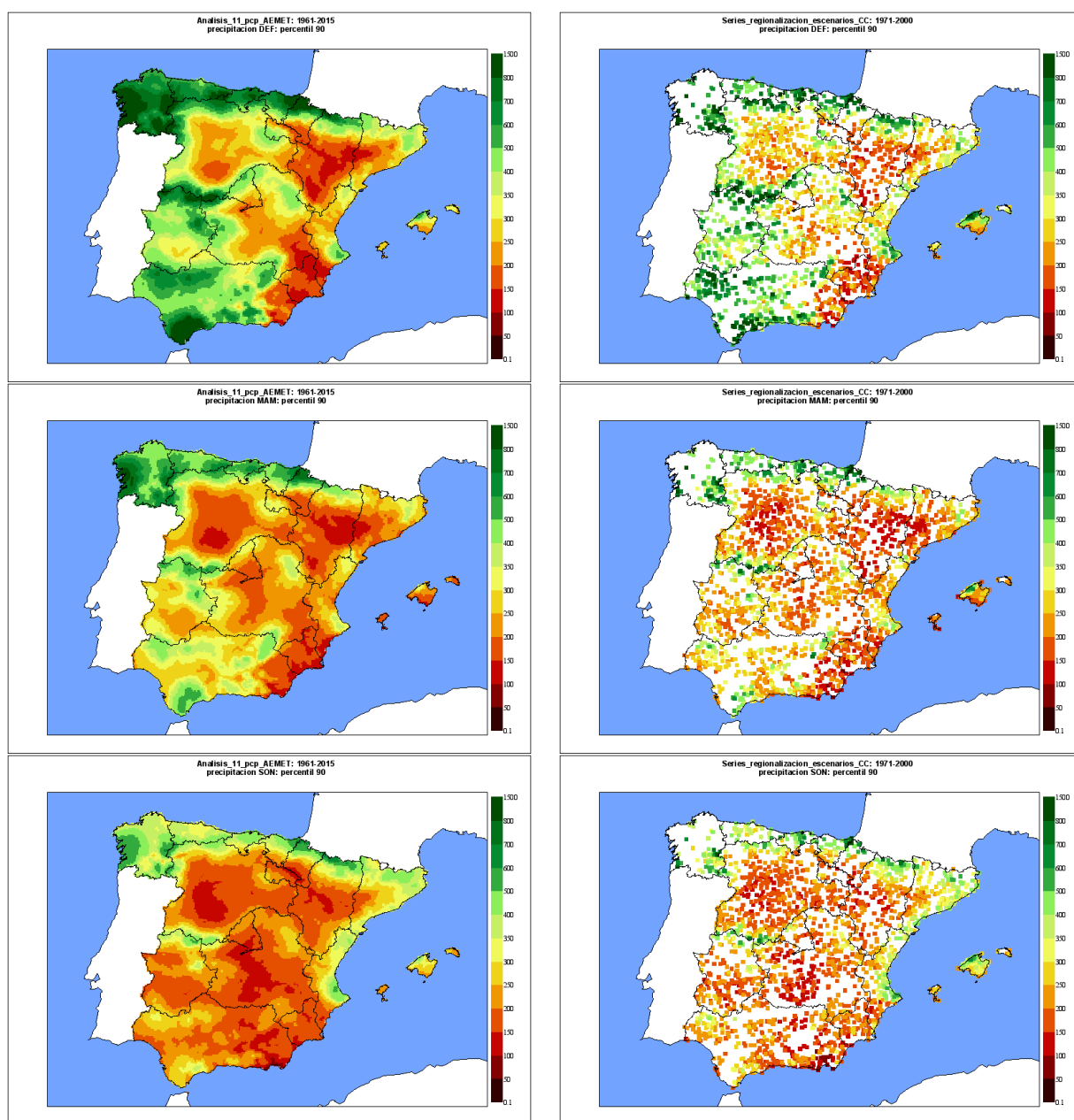


Figura 16: Valores máximos (percentil 90) de la precipitación acumulada en los trimestres de invierno (arriba), primavera (en medio), y otoño (abajo) utilizando el análisis (izquierda) y el conjunto de observaciones utilizado para generarlo (derecha). Periodo de referencia 1971-2000.

Además de con las observaciones que se han utilizado para el análisis, las estadísticas climatológicas obtenidas con la rejilla generada se han comparado con la rejilla ya existente Spain02. La Figura 17 muestra los mapas con la mediana de la precipitación, en invierno y en verano, calculada con la serie de análisis de precipitación descrito en este documento y con la de la rejilla Spain02. A pesar de que ambas rejillas difieren en el método de interpolación utilizado, en la resolución, y posiblemente no han utilizado exactamente el mismo conjunto de observaciones, los patrones espaciales son muy similares. Las diferencias más apreciables entre ambas se encuentran en Galicia, donde al menos en el análisis de precipitación que se describe, se ha utilizado un número muy reducido de estaciones pluviométricas. Por otro lado, como era de esperar, las estructuras de precipitación presentan una mayor definición en la serie de análisis que en la rejilla Spain02.

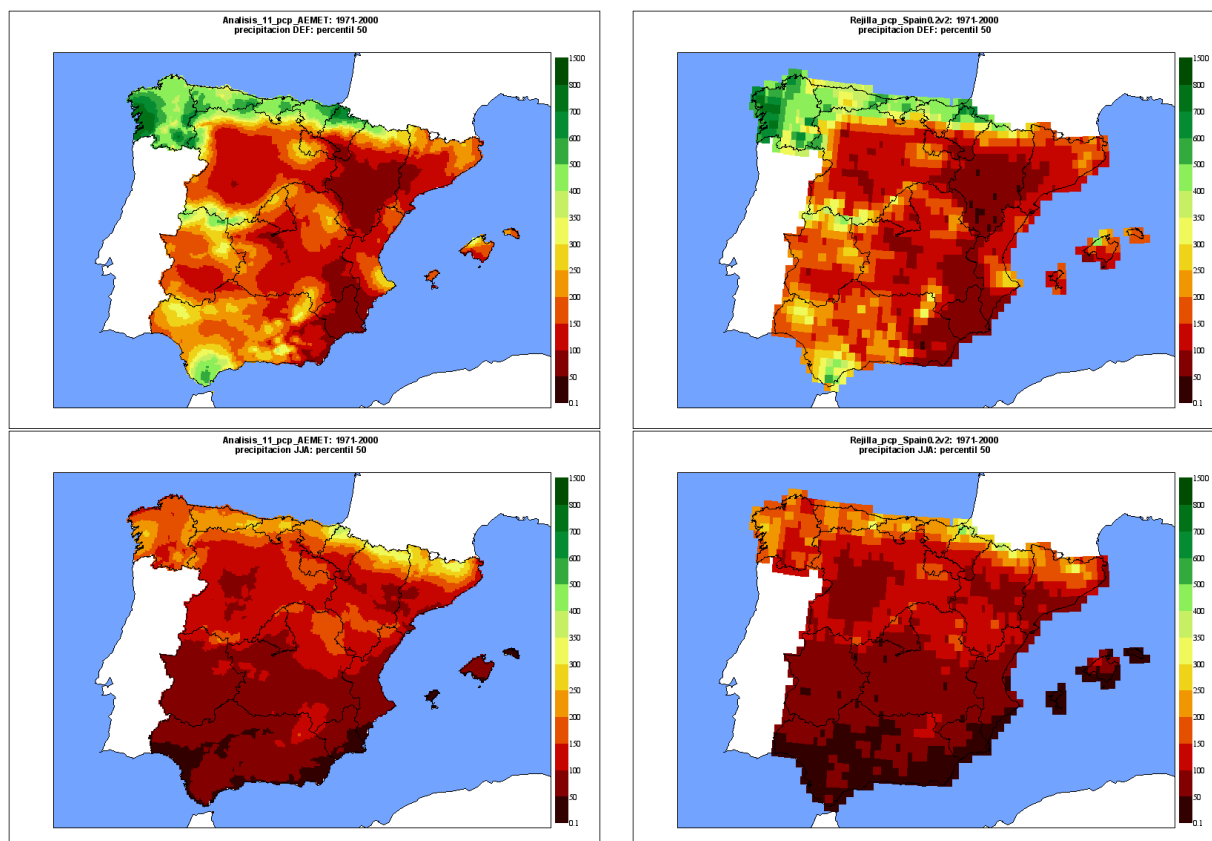


Figura 17: Mediana de la precipitación invernal/estival en el análisis y en la rejilla Spain02. Periodo de referencia 1971-2000.

## 5 Conclusiones

Se ha generado una serie de precipitación en rejilla a escala diaria, desde el 1 de enero de 1951 hasta la actualidad, con una finalidad climática. Para ello, se han utilizado las series temporales de observaciones pluviométricas de las estaciones del Banco Nacional de Datos Climatológicos que fueron seleccionadas para la generación de escenarios regionalizados de cambio climático con métodos empíricos, y que satisfacen un número de condiciones respecto a su homogeneidad y completitud. La geometría de la rejilla es la del modelo de predicción numérica HIRLAM operativo en AEMET que tiene una resolución aproximada de 5 km. El método utilizado para generar los campos diarios de precipitación es un análisis objetivo por Interpolación Estadística. Es un algoritmo lineal en el que los pesos asignados a las observaciones producen un error mínimo de interpolación.

Los parámetros del análisis se han calibrado en base a las estadísticas de las propias observaciones, en concreto a sus valores medios y varianzas, y correlaciones entre pares de estaciones, utilizando modelos analíticos para representar la dependencia espacial de las correlaciones tanto en la horizontal como en la vertical. La inclusión de la influencia de la diferencia de altitud, ha permitido simular la

variación de la precipitación con la orografía. Las varianzas y los valores medios de las observaciones pluviométricas en diferentes estaciones se han empleado para simular la variabilidad espacial de la precipitación. Todos estos parámetros se han ajustado de forma independiente para cada uno de los meses del año, representando así la variación estacional tanto la variabilidad diaria de la precipitación, como de sus características espaciales.

La serie pluviométrica en rejilla así creada, se ha evaluado de diferentes formas. En primer lugar, se ha examinado la distancia de las observaciones a la primera estimación del análisis, o *first guess*, que es un campo de precipitación cero en todos los puntos de la rejilla, y cómo se reduce en el proceso del análisis. Los histogramas de los valores *observación menos análisis* muestran una distribución no sesgada más estrecha que los correspondientes a los valores *observación menos first guess*.

La comparación de la precipitación en rejilla frente a observaciones simultáneas independientes, que no se han utilizado en el análisis, muestra valores muy elevados de correlación. En general, las diferencias sistemáticas entre ambos conjuntos de datos son de unas pocas décimas de mm/día y el rms de estas diferencias es inferior al error asumido para las observaciones (que incluye las componentes de precisión y de representatividad). Las mayores discrepancias que se observan, como se podría esperar, corresponden a zonas montañosas y a aquellas áreas en las que la densidad espacial de las observaciones utilizadas por el análisis es menor y existe una alta variabilidad espacial pluviométrica. Las zonas montañosas son también en las que se produce la mayor subestimación de la precipitación.

Las estadísticas climatológicas que se han calculado con los datos pluviométricos en rejilla, son consistentes con las obtenidas a partir de las observaciones empleadas por el análisis, tanto espacialmente como en cuanto a su variación estacional. La caracterización climática de años concretos que se obtiene con la rejilla, es prácticamente idéntica a la realizada con las observaciones directamente. No obstante, la magnitud de los extremos de la rejilla, medidos con el percentil 90, es inferior a la que presentan las observaciones. Las estadísticas climatológicas obtenidas con los análisis de precipitación son comparables a las calculadas con la rejilla Spain02, desarrollada hace años con un algoritmo diferente, y posiblemente con un conjunto ligeramente distinto de observaciones. Como era de esperar, por su mayor resolución, la rejilla presentada en este documento presenta mayor detalle espacial, apareciendo las principales diferencias en Galicia, la zona con menor densidad de observaciones del análisis.

Se sugiere utilizar la serie temporal de datos pluviométricos en rejilla a escala diaria que se ha generado como herramienta con fines climáticos, teniendo en cuenta los resultados de su evaluación y su incertidumbre, algunos de cuyos aspectos han sido mencionados en este documento. En AEMET, estos datos están mostrando ser muy útiles para analizar con detalle la influencia de los patrones de variabilidad climática en la precipitación de diferentes zonas de España, y para la generación y verificación objetiva de predicciones climáticas estacionales de mayor resolución. Una extensión de esta rejilla servirá próximamente para la regionalización estadística de proyecciones globales de cambio climático.

## 6 Reconocimientos

Leonor Martín Martín y Rosa García Marín han proporcionado las series de datos meteorológicos de las estaciones de la red EMEP/VAG/CAMP empleados para la validación de los análisis diarios de precipitación frente a estaciones independientes. Ana Isabel Ambrona Rodríguez ha suministrado los datos pluviométricos del Banco Nacional de Datos Climatológicos utilizados también para evaluar la rejilla generada. M<sup>a</sup> Jesús Casado Calle ha contribuido con sus sugerencias y comentarios a la revisión de este documento. Ernesto Rodríguez Camino e Isabel Martínez Marco han promovido y apoyado la realización de esta serie de datos pluviométricos en rejilla a escala diaria con fines climáticos. Las autoras agradecen muy especialmente sus aportaciones acerca de su diseño y desarrollo.

## 7 Referencias

- Brasnett, B., 1999: A global analysis of snow depth for numerical weather prediction. *J. Appl. Meteorol.* 38: 726-740. doi: 10.1175/1520-0450(1999)038<0726;AGAOSD>2.0.CO;2.
- Brunet, M., M.J., De Castro, M. Galán, P., López, J.A., Martín, J.M., Pastor, A., Petisco, E., Ramos, P., Ribalaygua, J., Rodríguez, E. y Torres, L., 2008: Generación de escenarios regionalizados de cambio climático para España, Agencia Estatal de Meteorología, Ministerio de Medio Ambiente y Medio Rural y Marino. ISBN: 978-84-8320-470-2, 157 páginas.
- Cansado, A., Martín, C., Navascués, 2004: Optimum Interpolation new snow depth analysis in HIRLAM, *HIRLAM Newsletter* 45, 248-252.
- Daley, R., 1991: Atmospheric data Analysis. Cambridge University Press, Cambridge, UK, 457pp.
- Dahlgren, P., Landelius, T., Kallberg, P., Gollvik, S., 2016: A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional High-Resolution Limited-Area Model (HIRLAM). *Q. J. R. Meteorol. Soc.* 142, 2119–2131, doi:10.1002/qj.2807.
- Dee, D.P. and Da Silva, A.M., 1998: Data assimilation in the presence of forecast bias. *Q.J.R. Meteorol. Soc.*, 124, 269-295. doi: 10.1002/qj.49712454512.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balsameda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van der Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healey, S.B., Hersbach, H., Holm, E.V., Isaksen, I., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.N., Vitart, F., 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597, doi: 10.1002.qj828.
- Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E., 1993: A meteorological estimation of relevant parameters for snow models. *Ann. Glaciol.*, 18, 65-71. [http:// dx.doi.org/10.1002/hyp.10125](http://dx.doi.org/10.1002/hyp.10125).
- Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B. and Mérindol, L., 2009: Reanalysis of 44 Yr of Climate in the French Alps (1958-2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation. *J. Appl. Meteorol. Climatol.*, 48(3), 429-449. <http://dx.doi.org/10.1175/2008JAMC1808.1>.
- Gill, P.E, Murray, W., 1978: Algorithms for the solution of the nonlinear least-squares problem *SIAM. J. Numerical Anal.* Vol 15, pp 977-992.
- Herrera, S., Gutiérrez, J.M., Ancell, R., Pons, M.R., Frías, M.D. & J. Fernández, 2012: Development and análisis o a 50-year high resolution daily gridded precipitation dataset over Spain (Spain0.2). *Int. J. Climatol.* 32: 74-84. doi:10.1002/joc.2256.
- Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones, M. New, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res. (Atmospheres)*, **113**, D20119, doi:10.1029/2008JD10201.
- Hollingsworth, A., Lönnberg, P., 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus* 38A: 111-136.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Am.Meteorol. Soc.* **77** 437–471. doi:10.1175/1520-0477(1996)077<0437:TNYP>2.0.CO;2

Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., Olsson, E., 2016: A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Q. J. R. Meteorol. Soc.* 142, 2132–2142, doi:10.1002/qj.2813.

Mahfouf, J.F., Brasnett, B., and Gagnon, S., 2007: A Canadian Precipitation Analysis (CaPa) Project: Description and Preliminary Results. *Atmosphere-Ocean* 45 (1) 2007, 1-17.

Martín, C., and Navascués, B., 2004: Estimation of structure functions for the analysis of two-meter temperature and relative humidity. *Proceedings of SRNWP/HIRLAM Workshop on Surface Processes, turbulence and mountain effects*. Norrköping, Sweden, 15-17 September 2004.

Mitchell, H., Charette, C., Chouinard, C., and Brasnett, B., 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Weather Rev.* 18: 1591-1614. doi: 10.1175/1520-0493(1990)118<1591:RISFTC>2.0.CO;2.

Navascués, B., Rodríguez, E., Ayuso J.J., and Järvenoja, S., 2003: Analysis of surface variables and parameterization of surface processes in HIRLAM. Part II: Seasonal assimilation experiment. *HIRLAM Technical Report 59*. Available from SMHI, S-60176 Norrköping, Sweden.

Navascués, B., Calvo, J., Morales, G., Santos, C., Callado, A., Cansado, A., Cuxart, J., Díez, M., del Río, P., Escribà, P., García-Colombo, O., García-Moya, J. A., Geijo, C., Gutiérrez, E., Hortal, M., Martínez, I., Orfila, B., Parodi, J. A., Rodríguez, E., Sánchez-Arriola, J., Santos-Atienza, I., and Simarro, J., 2013: Long-term verification of HIRLAM and ECMWF forecasts over Southern Europe: History and perspectives of Numerical Weather Prediction at AEMET, *Atmospheric Research*, 125–126, 20-33, doi: 10.1016/j.atmosres.2013.01.010.

Quintana-Seguí, P., Peral, M. C., Turco, M., Llasat, M. C., and Martín, E., 2016: Meteorological Analysis Systems in North-East Spain: Validation of SAFRAN and SPAN, *J. Environ. Inform.*, 27, 116– 130, doi:10.3808/jei.201600335.

Rodríguez, E., Navascués, B., Ayuso, J.J. and Järvenoja, S., 2003: Analysis of surface variables and parameterization of surface processes in HIRLAM. Part I: Approach and verification by parallel runs. *HIRLAM Technical Report 58*. Available from SMHI, S-60176 Norrköping, Sweden.

Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., García-Moya, J.-A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascués, B., Nielsen, N.-W., Odegaard, V., Rodríguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Sass, B.-H., Savijärvi, H., Schreur, B.-W., Sigg, R., The, H., Tijn, A., 2002: *HIRLAM-5 Scientific Documentation*. (Available from SMHI, S-60176 Norrköping, Sweden).