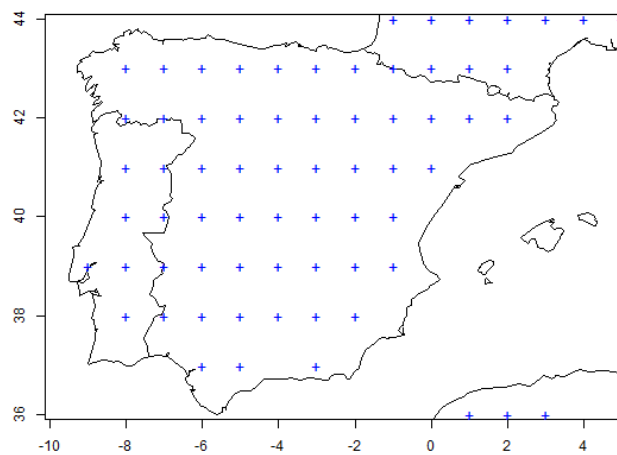




CALIBRACIÓN Y COMBINACIÓN DE PREDICCIONES ESTACIONALES EN LA PENÍNSULA IBÉRICA



Eroteida Sánchez García
José Voces Aboy
Ernesto Rodríguez Camino



Calibración y combinación de predicciones estacionales en la Península Ibérica

Autores: Eroteida Sánchez García, José Voces Aboy, Ernesto Rodríguez Camino (2014)

Agradecimientos: Este trabajo se realizó gracias a la ayuda inestimable de Paco Doblás-Reyes (junto con todo su equipo del IC3 de Barcelona), y de Caio Coelho, que nos han introducido en el mundo de la predicción estacional, siempre tan amables y colaboradores.



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, Alimentación y Medio Ambiente
Agencia Estatal de Meteorología
Madrid, 2014

Catálogo de Publicaciones de la Administración General del Estado:
<http://publicacionesoficiales.boe.es/>

NIPO: 281-14-013-4

Agencia Estatal de Meteorología (AEMET)
C/ Leonardo Prieto Castro, 8
28040 Madrid
<http://www.aemet.es/>

 @Aemet_Esp

Resumen

En este estudio se analizan cuatro modelos operativos de predicción a escala estacional en el área de la Península Ibérica para las variables temperatura a 2 metros y precipitación, presentando los resultados de la verificación de ambas variables por trimestres y utilizando una colección de índices de verificación adecuados tanto para las predicciones deterministas como probabilísticas.

Se verifican en primer lugar las salidas directas de los siguientes modelos dinámicos: *European Centre for Medium-Range Weather Forecasts (ECMWF) system 4*, *Météo-France system 3*, *Met Office system 3* y *National Centers for Environmental Prediction (NCEP) system version 2*. A continuación se aplica el método bayesiano basado en el enfoque denominado *forecast assimilation* (Coelho 2005, Coelho et al. 2006) que combina información histórica (climatológica) con las predicciones obtenidas a partir de los cuatro modelos, tanto individualmente como combinando los cuatro a la vez.

En general, los resultados muestran valores de los índices de verificación relativamente bajos como consecuencia de la baja predecibilidad a escala estacional en latitudes medias. Los resultados muestran generalmente consistencia entre los modelos siendo -y como es de esperar en nuestras latitudes- mejores los valores de los índices de verificación para la temperatura que para la precipitación. Se aprecian también ventanas de oportunidad con valores de los índices de verificación claramente mejores asociados a ciertas épocas de año, ciertas variables y en algunos casos limitadas a ciertos modelos.

INDICE

| | |
|---|----|
| 1. Introducción | 5 |
| 2. Datos | 5 |
| 3. Metodología | 6 |
| 4. Verificación | 8 |
| 5. Resultados | 8 |
| 6. Conclusiones | 15 |
| 7. Referencias | 16 |
| | |
| ANEXO I - Descripción de las posibles configuraciones al aplicar el FA | 17 |
| ANEXO II - Número de modos retenidos en el MCA | 20 |

1. Introducción

El carácter caótico de la atmósfera limita la predecibilidad determinista de su evolución hasta un máximo de 10-15 días. Sin embargo, más allá de este alcance las condiciones atmosféricas mantienen cierta predecibilidad pero solamente en un sentido estadístico, y por tanto, los pronósticos deben expresarse en términos probabilísticos (Murphy and Winkler, 1984).

Dos de las principales fuentes de incertidumbre en las predicciones climáticas son la falta de un perfecto conocimiento de las condiciones iniciales del sistema climático y la imposibilidad de modelar el sistema climático de un modo perfecto (Curry and Webster 2011; Knutti 2010; Slingo and Palmer 2011). Para solventar la primera fuente de incertidumbre, se utiliza la técnica del ensemble, es decir, se genera un conjunto de predicciones independientes a partir de un modelo dinámico, modificando ligeramente las condiciones iniciales (Gneiting and Raftery 2005; Palmer 2000). Con respecto a la segunda, entre otras técnicas, se puede usar la combinación de diferentes modelos (Doblas-Reyes et al. 2009). Coelho et al. (2004) usa un método bayesiano para combinar las predicciones obtenidas del modelo dinámico del Centro Europeo (ECMWF) con un modelo estadístico a la vez que calibrarlo para el índice Niño3.4. Stephenson et al. (2005) generaliza este índice combinando más de un modelo y más de una variable (DEMETER; Palmer et al. 2004), mostrando que se obtienen mejores resultados que considerando cada sistema de predicción individualmente.

También hay que tener en cuenta que la predecibilidad –siempre de carácter probabilístico- a escala de tiempo estacional depende fuertemente de ciertos modos de variabilidad, de las regiones, de la época del año y de las variables. Las predicciones estacionales de carácter operativo se circunscriben frecuentemente a las principales variables climáticas: temperatura y precipitación (Kirtman y Pirani 2008). La escasa predecibilidad que exhibe la atmósfera en las latitudes medias, y en el SW de Europa en particular (Doblas-Reyes 2010), obliga a que habitualmente se empleen enfoques probabilísticos robustos basados en terciles.

Desde un punto de vista operativo, el objetivo principal de este estudio es conocer la capacidad predictiva de los modelos considerados en función de la época del año y la variable, para mejorar las actividades de predicción estacional operativa en la Península Ibérica. Para ello es fundamental evaluar la pericia de los pronósticos estacionales e identificar las ventanas de oportunidad.

En primer lugar, se consideran las salidas directas de cuatro modelos dinámicos operativos a escala estacional: *European Centre for Medium-Range Weather Forecasts (ECMWF) system 4*, *Météo-France system 3*, *Met Office system 3*, y *National Centers for Environmental Prediction (NCEP) system version 2*. En segundo lugar, se aplica el método bayesiano, descrito en Stephenson et al. (2005), con diferentes parámetros, para tratar de mejorar las predicciones generadas a partir de las salidas directas de los modelos dinámicos.

En la sección 2, se describen las fuentes de datos de las observaciones y modelos de predicción estacional utilizados. Los métodos de calibración y combinación se tratan en la sección 3, y en la sección 4 la verificación de los citados sistemas de predicción. Los resultados obtenidos y su análisis se presentan en la sección 5. Por último, las conclusiones y las futuras líneas de investigación se encuentran en la sección 6. En los Anexos I y II se muestran diferentes detalles sobre el estudio llevado a cabo.

2. Datos

La base de datos de observaciones utilizada para hacer el estudio se ha obtenido del proyecto EU-FP6 ENSEMBLES (<http://ensembles-eu.metoffice.com>), concretamente la base de datos en puntos de rejilla E-OBS version 6.0, que contiene los datos diarios de la precipitación total y de la temperatura en la superficie terrestre con una resolución horizontal de 0.5° de latitud y 0.5° de longitud, desde 1950 hasta la actualidad. Estos datos se basan en datos diarios de estaciones disponibles en el sitio web del ECA&D (<http://www.ecad.eu>) junto con datos adicionales obtenidos por los proyectos STARDEX y ENSEMBLES (datos restringidos). A partir de los datos originales se han calculado los valores medios trimestrales de temperatura y de precipitación acumulada con 1° x 1° de resolución y sus anomalías, para evaluar la calidad de las predicciones de los diferentes modelos. También se

utilizan en el método bayesiano para calibrar y combinar las predicciones estacionales (ver sección 3).

Se usan las anomalías de la precipitación acumulada y de la temperatura media trimestral de los hindcasts (predicciones a tiempo pasado) de los siguientes modelos acoplados océano-atmósfera: *European Centre for Medium-Range Weather Forecasts (ECMWF) system 4*, *Météo-France system 3*, *Met Office system 3*, y *National Centers for Environmental Prediction (NCEP) system version 2*.

El *European Centre for Medium-Range Weather Forecasts (ECMWF) system 4 (S4)* está compuesto por la versión Cy36r4 del modelo *Integrated Forecast System (IFS)* a TL255 de resolución (80 km) para la componente atmosférica acoplado con la configuración ORCA1 del *Nucleus for European Modelling of the Ocean (NEMO)* para la componente oceánica. El IFS tiene 91 niveles e incluye la estratosfera. Las condiciones iniciales del océano provienen de un sistema de asimilación basado en un análisis variacional multivariante con ajuste del bias. Las condiciones iniciales del suelo y de la atmósfera se obtienen del ERA Interim y del modelo operativo del ECMWF, y del modelo de superficie HTESSEL (Kim et al. 2012, Molteni et al. 2011). El tamaño del ensemble es de 15 miembros y el periodo de integración es de 7 meses desde 1981 hasta 2010.

El *Météo-France system 3 (MF3)* está formado por el modelo ARPEGE (*Action de Recherche Petite Echelle Grande Echelle*) versión 4 como componente atmosférica (Batté and Déqué 2011), acoplado con ORCA desarrollado por LOCEAN para la componente oceánica. Las condiciones iniciales del océano son tomadas de MERCATOR. El modelo atmosférico tiene una resolución horizontal TL127 (160 km) y 91 niveles con la estratosfera no totalmente cubierta. El tamaño del ensemble es de 11 miembros y el periodo de integración es de 7 meses desde 1981 hasta 2010.

El *Met Office system 3 (UKMO3)*, tiene una componente atmosférica con una resolución espacial de $2.5^\circ \times 3.75^\circ$ y 85 niveles en la vertical. El modelo oceánico tiene una resolución general de 1.25° , aumentando suavemente desde 30° N/S hasta 0.33° en el Ecuador, con 75 niveles en la vertical. Las condiciones iniciales del océano se toman del sistema de análisis del océano del *UK Met Office*. El ensemble tiene 15 miembros y cubre el periodo de 1987 a 2010.

El *National Centers for Environmental Prediction system version 2 (CFSv2)* tiene una componente atmosférica con una resolución espacial de 100 km y 64 niveles en la vertical (Kim et al. 2012, Saha et al. 2013, Yuan et al. 2011). La componente oceánica es el *Geophysical Fluid Dynamics Laboratory Modular Ocean Model (MOM4)* version 4 con una resolución horizontal de 0.5° , refinada hasta 0.25° entre 10° N y 10° S, y 40 niveles en la vertical. Aunque el ensemble para el hindcast tiene 28 miembros, inicializados en diferentes días y horas, sólo se han usado para este estudio los 12 miembros más próximos a la fecha considerada. El periodo del *hindcast* cubre desde 1982 a 2010.

3. Metodología

El *Forecast Assimilation (FA)* es un método bayesiano que se ha usado para combinar los cuatro sistemas de predicción analizados en este trabajo. Éste método ha demostrado mejorar al *Simple Multi-Model (SMM)*, y a otros métodos de combinación (Lage et al. 2013), calibrando y combinando predicciones de diversas fuentes con información empírica a priori (histórica) (Stephenson et al. 2005). Una característica útil del FA es que permite desplazar espacialmente los patrones pronosticados para corregir los sesgos de las predicciones de los modelos combinados. En otras palabras, el procedimiento tiene en cuenta las dependencias de los puntos inter-grid. El método FA es una aproximación probabilística consistente que puede usarse para combinar información histórica (climatológica) con los pronósticos de la media del ensemble de los modelos dinámicos. Como cualquier otro método bayesiano, tiene su base en la teoría de la probabilidad y por tanto puede proporcionar predicciones probabilísticas bien calibradas.

En el anexo I se detallan y discuten las diferentes posibilidades para aplicar el método FA, en función de si se utiliza validación cruzada, de si se trabaja con valores estandarizados o sólo con las anomalías de las variables al realizar el análisis de covarianza máxima (MCA) y del periodo de referencia usado para el cálculo de la función *prior*. De la discusión en el anexo I se propone entre todas las alternativas posibles la denominada configuración FA3 que se basa en la utilización de validación cruzada al calcular el FA y el uso de valores estandarizados medios trimestrales de las dos

variables, precipitación y temperatura a 2 m, a la hora de calcular el MCA, y finalmente la utilización del periodo disponible en la serie de observaciones 1960 2010. Todos los resultados mostrados en la Sec. 5 se basan en la configuración FA3 descrita en el anexo I

En ausencia de modelos de predicción, la única información probabilística que se puede suministrar sobre una variable observable, y , se basa en la suposición de que los futuros valores que va a tomar se comportarán como hicieron en el pasado. Por ejemplo, la distribución de probabilidad de y_i en el instante t_i puede estimarse usando la función densidad de probabilidad climatológica, $p(y_i)$, estimada a partir de las observaciones históricas. En la teoría bayesiana, a $p(y_i)$ se la conoce como la función de distribución *prior* y encapsula el conocimiento a priori sobre la probabilidad de los posibles valores de y_i los cuales se conocen por la experiencia pasada, en nuestro caso por la climatología. Una función *prior* mejor se puede obtener a partir de algún modelo empírico. La base del método FA es que si se conocen los pronósticos, M , de un modelo determinado, se puede actualizar la función *prior* $p(y_i)$ para obtener la distribución condicional *posterior*, $p(y_f/x_f)$; es decir, la distribución de probabilidad de y_i dado el pronóstico $M(x_f)$, siendo x_f las variables de estado del modelo. Al utilizar probabilidades condicionadas se reduce la incertidumbre en los pronósticos sobre los futuros valores de y_i (Jolliffe and Stephenson 2003, chapter 9). La función de distribución *posterior* $p(y_f/x_f)$ se obtiene a partir de la función *prior* $p(y_i)$ aplicando el teorema de Bayes (Fig.1).

$$p(y_f / x_f) = \frac{p(x_f / y_f) p(y_f)}{p(x_f)}$$

posterior
likelihood
prior

Fig. 1.- Teorema de Bayes

En la figura 2 podemos ver una representación del espacio de trabajo en el método *Forecast Assimilation*.

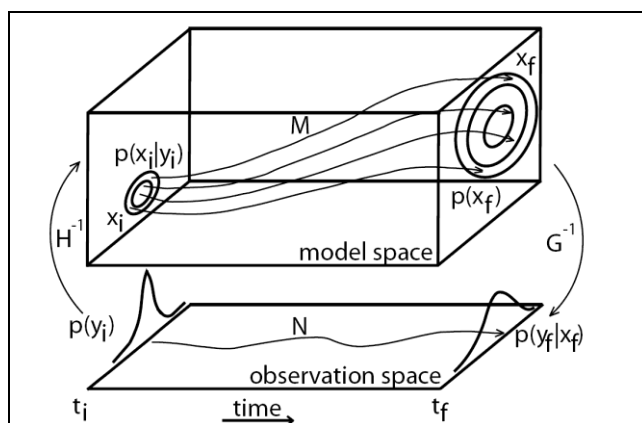


Fig. 2.- Espacio de trabajo en FA (Fuente: Stephenson et al. (2005))

En este trabajo se usa como función *prior* la climatología de la base de datos de observaciones E-OBS v6.0. Como pronósticos numéricos, M , se han utilizado cada uno de los cuatro modelos dinámicos detallados anteriormente, y también una combinación de los cuatro modelos, asignando el mismo peso a cada modelo. Se utiliza el análisis de covarianza máxima (MCA) para extraer los modos de covarianza entre las observaciones y predicciones (von Storch and Zwiers 1999). En el Anexo II se describe cómo se selecciona el número de modos retenidos al aplicar el MCA.

Las anomalías de los diferentes sistemas de predicción, calculadas como la diferencia entre el valor predicho por el sistema y su valor climatológico, son obtenidas eliminando el año del que se trate. De la misma forma se calculan las anomalías de las observaciones. Como periodo de trabajo se toma 1988 a 2008, ya que es el periodo común en el que se dispone de datos de los *hindcasts* de los cuatro modelos.

4. Verificación

La calidad de las predicciones de temperatura y precipitación obtenidas con los distintos sistemas de predicción considerados, ha sido evaluada desde un punto de vista determinista y probabilista. La significación estadística de todos los índices calculados se ha cuantificado mediante el cálculo del p-valor usando el método no paramétrico de *bootstrapping*. (Más detalles sobre este método se pueden encontrar en Wilks 2006).

El coeficiente de correlación entre los valores de las anomalías de observaciones y predicciones sobre el dominio que cubre la Península Ibérica, para las variables temperatura y precipitación, para los 12 trimestres del año y para los alcances (*lead-times*) 1, 2 y 3, es el único índice a nivel determinista calculado.

Desde el punto de vista probabilístico, se calculan los siguientes índices, también para las variables temperatura y precipitación, para los 12 trimestres del año y para los *lead-times* 1, 2 y 3: el *Ranked Probability Skill Score* (RPSS) para terciles y el *Relative Operating Characteristic (ROC) area*, para dos eventos (valores por encima/debajo del tercil superior/inferior). Una definición completa de estos índices se puede encontrar en Wilks (1995).

El *Ranked Probability Skill Score* (RPSS) es una generalización del *Ranked Probability Score* (RPS), que mide la suma de las diferencias al cuadrado en el espacio de las probabilidades acumuladas para una predicción probabilística multicategoría. El RPSS permite comparar el sistema de predicción en estudio con otro sistema de referencia. Valores positivos del RPSS indican más pericia que el sistema de referencia, generalmente la climatología.

Las curvas ROC miden la discriminación en la pericia de un sistema de predicción. Si la categoría en estudio es “valores por encima del tercil superior”, el área ROC indica la probabilidad de discriminar con éxito las observaciones con valores superiores al tercil superior de aquellas cuyos valores están por debajo del tercil superior. El área ROC toma valores entre 0 y 1; 0.5 indica no *skill*, 1 indica una discriminación perfecta, y 0 indica una perfecta mala discriminación. Es importante resaltar que las curvas ROC sólo miden la habilidad para discriminar entre dos posibles resultados, pero no dicen nada sobre la fiabilidad ni son sensibles al sesgo.

Para el cálculo de los terciles superior e inferior de las predicciones o *hindcast* de los distintos modelos dinámicos se usan todos los miembros del ensemble correspondiente, desde 1988 a 2008. Los terciles de la serie de observaciones también se calculan para el mismo periodo. En cuanto a los sistemas de predicción obtenidos mediante el método bayesiano, los terciles se obtienen suponiendo una distribución normal con media y desviación estándar obtenidas al aplicar el método.

5. Resultados

Se han calculado los tres índices descritos en la sección anterior, considerando todos los puntos de la rejilla del dominio que cubre la Península Ibérica (10°W-05°E, 36°N-44°N) (Fig. 3), por trimestres, para cada sistema de predicción analizado, para cada variable y para diferentes alcances (*lead-times*). Los valores obtenidos se presentan en formato de tabla (Tablas 1 a 4). La columna de la izquierda corresponde a los valores obtenidos para la precipitación y la columna de la derecha a los de la temperatura. En la fila superior de cada tabla aparecen los valores obtenidos para *lead-time* 1, en la fila central para *lead-time* 2 y en la inferior para *lead-time* 3. Por ejemplo, en el caso del modelo S4, para *lead time* 1, los valores tomados para comparar con las observaciones, para el trimestre FMA del año 1989, son los obtenidos con la pasada del modelo del 1 de enero de 1989. Para *lead time* 2, serían los obtenidos con la pasada del 1 de diciembre de 1988 y para *lead time* 3 la del 1 de noviembre de 1988.

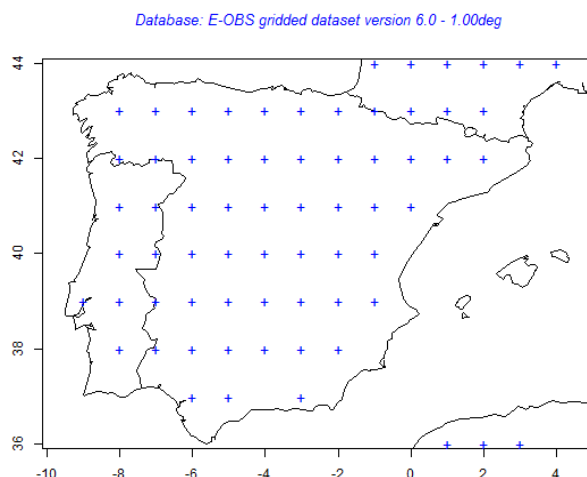


Fig. 3.- Puntos de la rejilla del área considerada en este estudio con datos en la base de observaciones E-OBS versión 6.0, para el periodo 1960-2010.

La Tabla 1 muestra los coeficientes de correlación entre observaciones y predicciones en el área de la Península Ibérica (Fig. 3) para las anomalías de precipitación acumulada y temperatura a 2 m., calculado por trimestres sucesivos separados un mes entre sí y para *lead-time* 1, 2 y 3. Se presentan los resultados de la salida directa de los cuatro modelos considerados y de la aplicación individualizada por modelos y combinado todos los modelos del algoritmo de calibración y combinación con configuración FA3 (descrita en el anexo I). La variable temperatura muestra valores significativos (con p-valor 0.05) superiores a 0.3 que se concentran en los meses estivales y a principios del otoño siendo esta ventana de oportunidad común a prácticamente todos los modelos. Fuera de esta ventana, hay valores significativos solamente en alguno de los modelos (UKMO3) y muy restringido a ciertos meses de invierno. Al aplicar el algoritmo FA de calibración se observa una cierta mejora –no general- que viene a coincidir con los meses y modelos que muestran una mayor correlación. Es de destacar que cuando el algoritmo se aplica de forma combinada a los cuatro modelos considerados no se observa una mejora con respecto al modelo que mejor se comporta, sugiriendo que quizás la mejor estrategia consistiría en eliminar en una hipotética combinación de modelos el (los) que peor se comporta(n). La precipitación –comparada con la temperatura- muestra valores más bajos, y en general sin significación, de la correlación. En este caso apenas puede apreciarse una mejora asociada a la aplicación del algoritmo FA. Es de destacar los valores altos de la pericia para la precipitación alcanzados en JJA para el *lead-time* 2 que provienen principalmente del modelo UKMO3 y que no parece tener una fácil interpretación a la vista del comportamiento en los *lead-time* 1 y 3.

La comparación del índice de correlación para diferentes alcances muestra algunos resultados notables. En primer lugar, los alcances más largos –que suponen predicciones para el mismo periodo a verificar pero realizadas en meses anteriores- no siempre van a asociadas con una degradación de la pericia. Por ejemplo, la pequeña ventana de oportunidad invernal para la temperatura asociada al modelo UKMO3 que aparece en DJF para *lead-time* 1, se traslada a JFM para *lead-time* 2 y a FMA para *lead-time* 3 sugiriendo la existencia de condiciones iniciales con más alta predecibilidad en los datos del mes de noviembre. Ciertas barreras de pericia asociadas a trimestres específicos o bien se trasladan hacia la derecha de la tabla al aumentar el *lead-time* o bien se diluyen o desaparecen al utilizarse condiciones iniciales anteriores con más alta predecibilidad. Por ejemplo, la barrera de pericia FMA para temperatura en el *lead-time* 2 se traslada un mes para el *lead-time* 3. También es de destacar que los beneficios asociados a la aplicación del algoritmo FA en el caso de la temperatura dependen mucho de cada modelo específico, por ejemplo, la aplicación al modelo UKMO3 siempre tiene efectos positivos globalmente hablando para la suma de todos los periodos y para todos los *lead-time*, hecho que no ocurre con los otros modelos. Esto nos lleva a la necesidad de considerar en el futuro posiblemente diferentes estrategias o modos de calibración para cada modelo. Es decir, muy posiblemente ciertos modelos tienen más posibilidades de mejorar al aplicarles los algoritmos de calibración y combinación aquí considerados y otros pueden estar ya muy optimizados y el margen de posible mejora es muy pequeño.

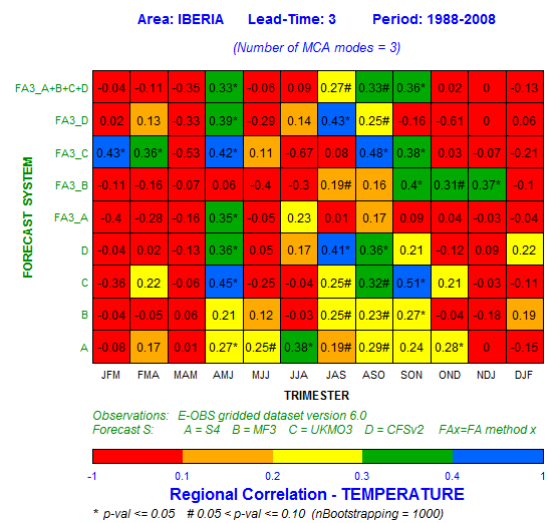
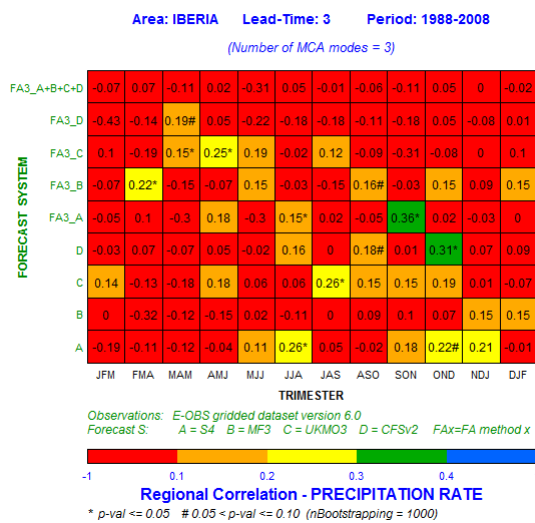
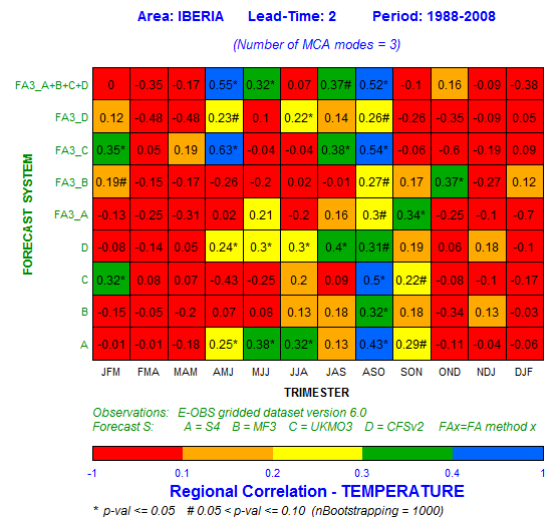
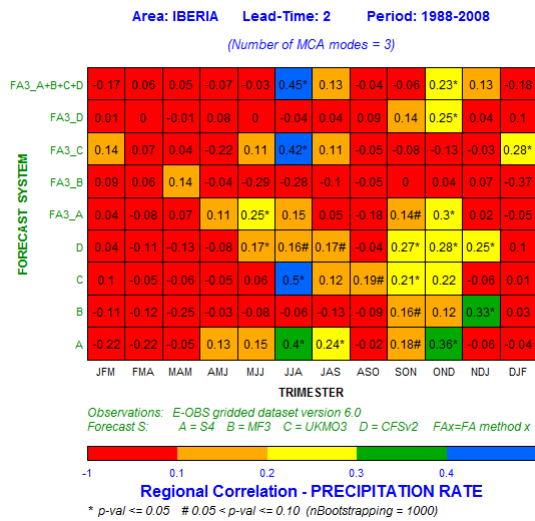
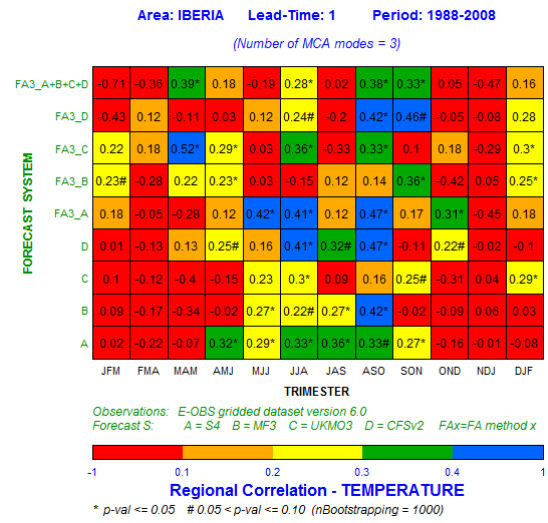
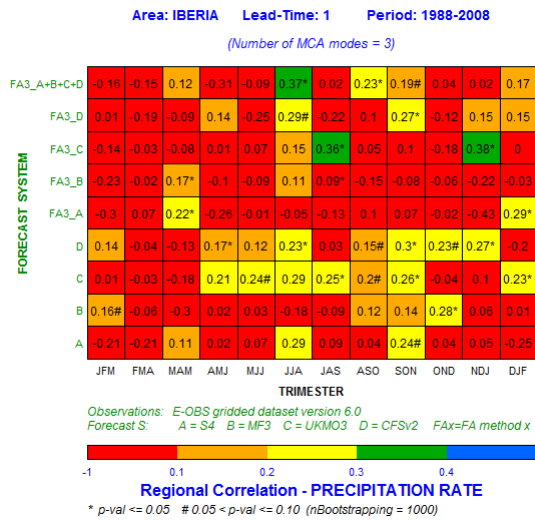


Tabla 1. Coeficiente de correlación en el área de la Península Ibérica, para las anomalías de precipitación acumulada y temperatura a 2 m., entre observaciones y predicciones, por trimestres y para lead-time 1, 2 y 3, con validación cruzada en el cálculo del FA (configuración FA3). En abscisas se representan los periodos de tres meses para los que se realiza la predicción desplazados de uno en uno para cada elemento de la tabla. En ordenadas se representan los diferentes modelos (S4, MF3, UKMO3, CFSv2) con salidas directas y con la aplicación del algoritmo FA3 (véase texto para su descripción).

La Tabla 2 muestra el *Ranked Probability Skill Score* (RPSS) para las anomalías de precipitación acumulada y temperatura a 2 m. en el área de la Península Ibérica (Fig. 3). En primer lugar conviene recordar que el RPSS no es simétrico (¡los valores negativos y positivos no tienen significados opuestos!). El rango de valores del RPSS se mueve desde 1 (predicción perfecta) hasta menos infinito, siendo los valores negativos indicativos de ausencia de pericia con respecto a la climatología. En consecuencia, nos centraremos solamente en los valores positivos (verdes y azules en las tablas). Al igual que cuando se medía la pericia con el coeficiente de correlación, lo primero que se observa es la prácticamente ausencia de pericia en la predicción de precipitación, mientras que la pericia medida por el RPSS para la predicción de temperatura muestra pequeñas diferencias entre modelos si bien en general no son significativas. Algunas de las características que se han detectado con el coeficiente de correlación también se observan con el RPSS, tales las ventanas de oportunidad para temperatura asociadas a los meses estivales y a principios del otoño –si bien se nota una fuerte dependencia de los modelos-, el desplazamiento de ciertas barreras o picos de pericia asociados a trimestres específicos al aumentar el *lead-time*. Se nota también una ausencia sistemática de mejora en el modelo S4 cuando se aplica el algoritmo de calibración y combinación, lo que nos devuelve al tema de que las distintas posibilidades de mejora de los diferentes modelos al aplicarles el algoritmo de calibración y combinación aquí analizado.

La Tabla 3 muestra el área ROC del tercil inferior y para las predicciones de anomalías de precipitación acumulada y temperatura a 2 m en el área de la Península Ibérica (Fig. 3). Recordemos que área ROC mide la discriminación de un sistema de predicción y toma valores entre 0 y 1, con valores inferiores a 0.5 indicando ausencia de pericia en la discriminación. La predicción de temperatura para el *lead-time* 1 muestra los valores más altos (superiores a 0.70) del área ROC para el tercil inferior de nuevo en los meses estivales (sobre todo en JJA) y principios de otoño, siendo estos valores además significativos. Uno de los modelos (UKMO3) muestra además una clara pericia en el periodo invernal (DJF). En general, se observa una mejora generalizada de todos los modelos al aplicar el algoritmo FA3, siendo esta mejora muy clara de nuevo para algunos modelos (UKMO3) y apenas perceptible para otros (p.e., S4). Al igual que con otros índices también se observa el desplazamiento de ciertas barreras o picos de pericia asociados a trimestres específicos al aumentar el *lead-time*. La predicción de precipitación si bien muestra en general una menor pericia que la de temperatura -en términos del área ROC para el tercil inferior- presenta valores significativos superiores a 0.5 en uno de los modelos (UKMO3) para el *lead-time* 1. La aplicación del algoritmo FA3 no se puede decir que mejore de forma generalizada la pericia siendo la variación muy dependiente del modelo y de la época del año.

La Tabla 4 muestra el área ROC del tercil superior para las predicciones de anomalías de precipitación acumulada y temperatura a 2 m en el área de la Península Ibérica (Fig. 3). Comparando con el comportamiento de otros índices –y singularmente con el área ROC del tercil inferior- se observa para la temperatura y el *lead-time* 1 un efecto positivo de la aplicación del algoritmo FA3 en tres de los modelos y una degradación en uno de los modelos así como en los cuatro modelos calibrados y combinados conjuntamente. El resto de las características ya se han mencionado al describir las tablas correspondientes a los otros índices.

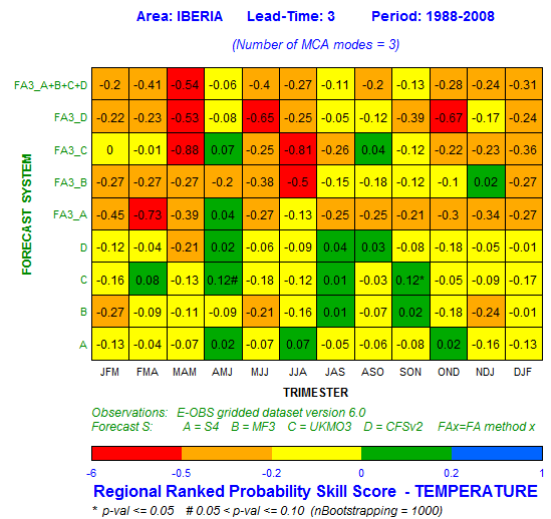
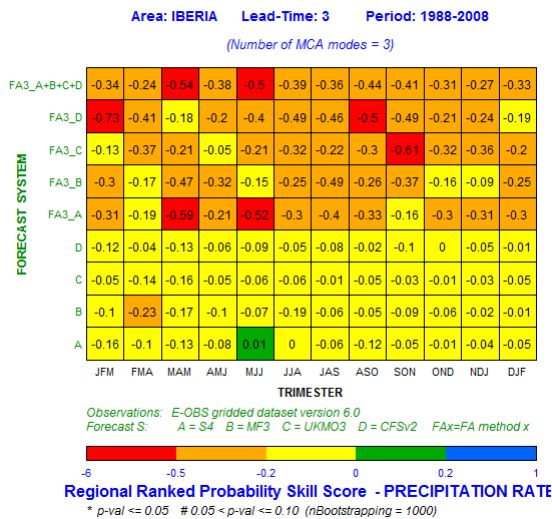
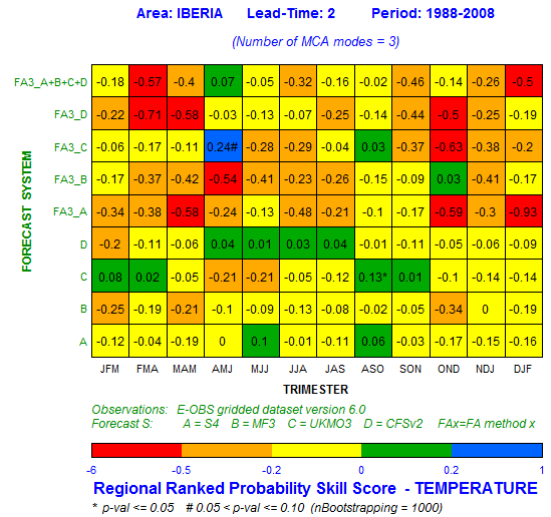
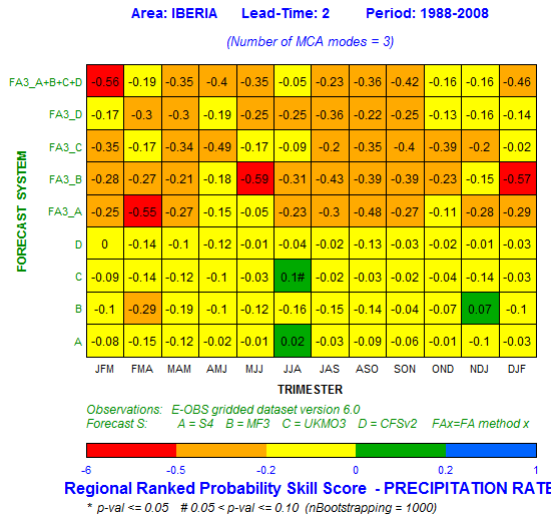
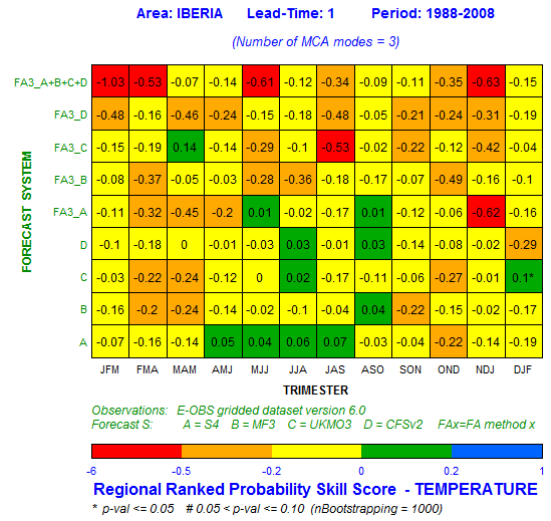
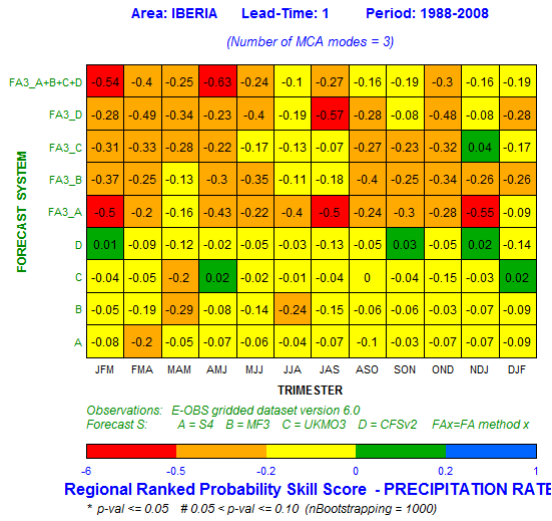


Tabla 2. Igual que Tabla 1, pero para el RPSS basado en terciles.

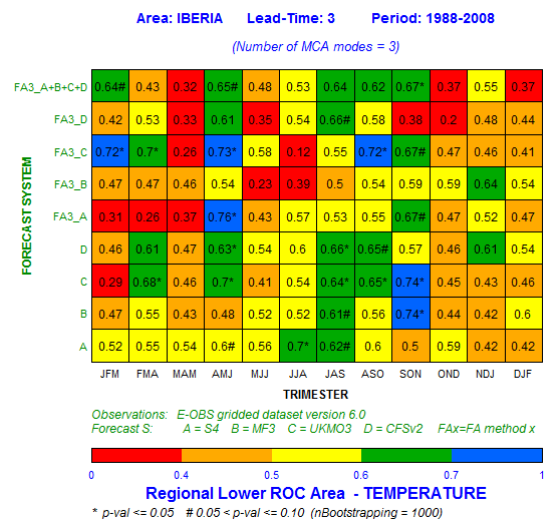
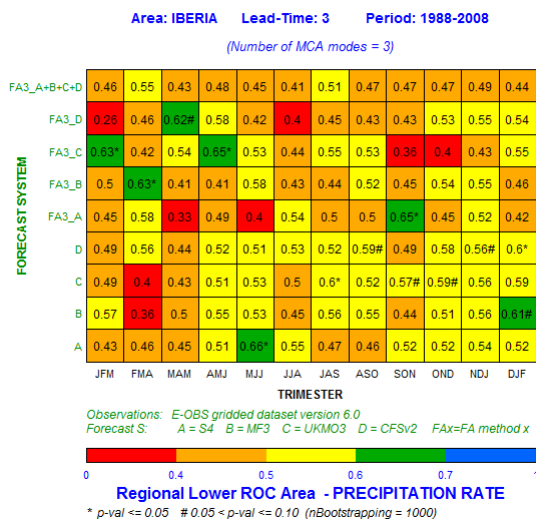
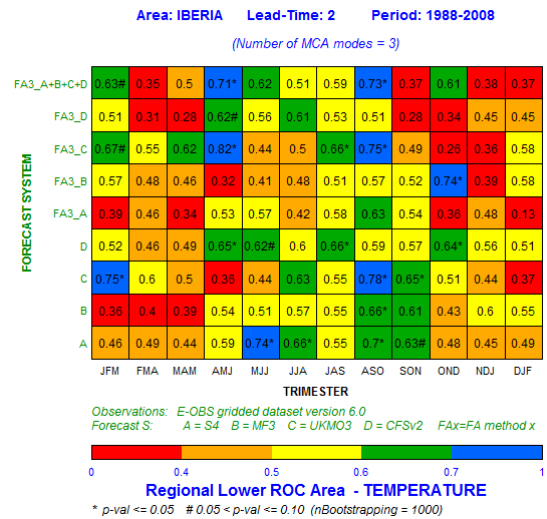
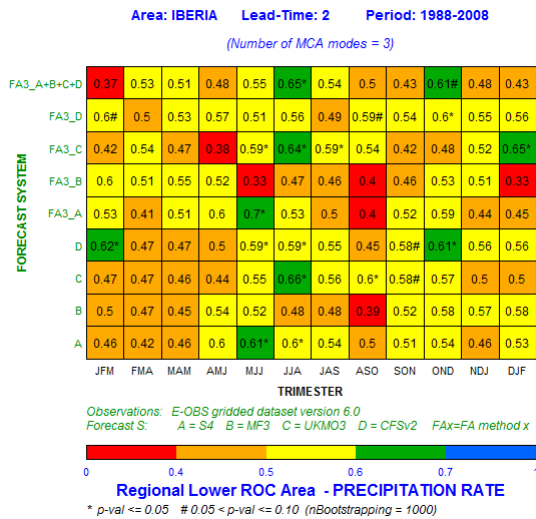
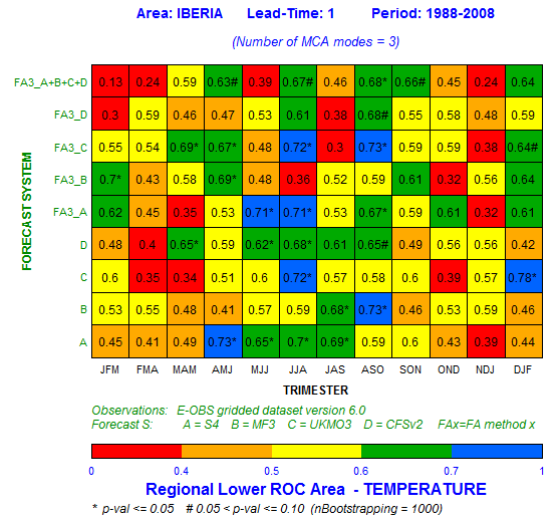
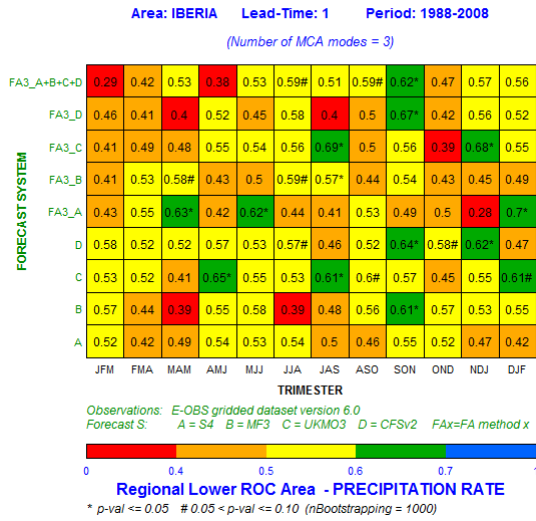


Tabla 3. Igual que Tabla 1, pero para el área ROC del tercil inferior.

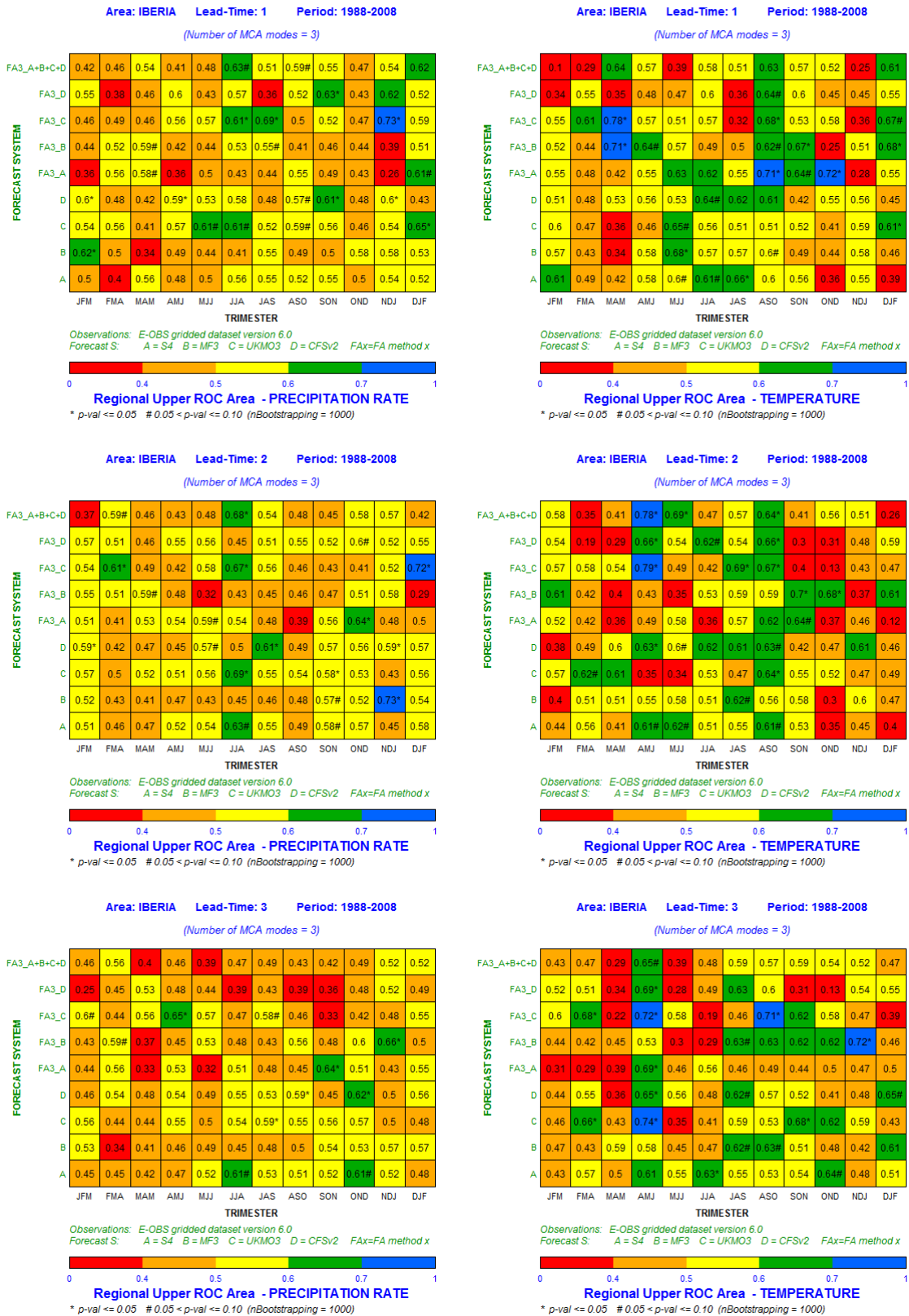


Tabla 4. Igual que Tabla 1, pero para el área ROC del tercil superior.

6. Conclusiones

Los resultados muestran, como era de esperar, valores de los índices de verificación que corresponden con una baja pericia de la predicción, consecuencia de la baja predecibilidad a escala estacional en latitudes medias. Sin embargo se pueden todavía extraer algunas conclusiones generales que pasamos a enumerar:

- Mejores valores en general para la temperatura que para la precipitación, como es de esperar en latitudes medias.
- Relativa consistencia entre modelos que permite identificar algunas ventanas de oportunidad para la predicción estacional asociados a ciertas épocas de año, ciertas variables y en algunos casos limitadas a ciertos modelos. Se puede destacar singularmente el verano para la temperatura posiblemente como relacionada con la tendencia general al calentamiento por el cambio climático.
- Ciertos índices muestran más pericia que otros, por ejemplo el área ROC, que proporciona una indicación de la discriminación, tiende a tener más pericia (relativa a la climatología) que otros índices que exploran otros aspectos de la predicción.
- Se detectan ciertas características relacionadas con la diferente calidad de los modelos en las diferentes épocas del año. Por ejemplo, el UKMO3 tiende a tener más pericia en invierno que el resto de los modelos, si bien el S4 muestra en general más pericia promediando sobre todas las estaciones.
- Los alcances más largos –que suponen predicciones para el mismo periodo a verificar pero realizadas en meses anteriores- no siempre van a asociadas con una degradación de la pericia, sugiriendo la existencia de condiciones iniciales con más alta predecibilidad en ciertos meses. Esto se traduce en que ciertas barreras o picos de pericia asociadas a trimestres específicos se trasladan hacia la derecha de la tabla al aumentar el *lead-time* al utilizarse condiciones iniciales anteriores pero con más alta predecibilidad.
- Los beneficios asociados a la aplicación del algoritmo FA dependen mucho de cada modelo específico, lo que nos lleva a la necesidad de considerar en el futuro posiblemente diferentes estrategias o modos de calibración para cada modelo. Es decir, muy posiblemente ciertos modelos tienen más posibilidades de mejorar al aplicarles los algoritmos de calibración y combinación aquí considerados y otros pueden estar ya muy optimizados y el margen de posible mejora es muy pequeño. Posiblemente haya que revisar la utilización del mismo algoritmo con la misma configuración (aquí se ha elegido la opción FA3) para todos los modelos.

7. Referencias

- Coelho, C. A. S. (2005) Forecast calibration and combination: Bayesian assimilation of seasonal climate predictions. PhD thesis, Department of Meteorology, University of Reading. 178 pp.
- Coelho C.A.S., D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes y G. J. van Oldenborgh (2006). Towards an integrated seasonal forecasting system for South America. *J. Climate*. 19, No. 15, 3704-3721.
- Doblas-Reyes, F. (2010). Seasonal prediction over Europe. Proceedings of the ECMWF Seminar on Predictability in the European and Atlantic regions, 6 to 9 September 2010. (http://www.ecmwf.int/publications/library/ecpublications/_pdf/seminar/2010/Doblas_Reyes.pdf)
- Jolliffe, Ian T., Stephenson, David B. (2003). Forecast Verification. A Practitioner's Guide in Atmospheric Science. ISBN 0-471-49759-2.
- Kirtman, B. y A. Pirani. (2008). WCRP Position Paper on Seasonal Prediction: Report from the First WCRP Seasonal Prediction Workshop, June 4–7, 2007, Barcelona, Spain. WCRP Informal Report No. 3/2008, ICPO Publication No. 127.
- Stephenson, D. B., Coelho, C. A. S., Doblas-Reyes, F. J. y Balmaseda, M. (2005). Forecast Assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus* 57A: 253–264.
- Troccoli, A., Harrison, M., Anderson, D.L.T., Mason, S.J. Seasonal climate: Forecasting and Managing Risk. 2008, XIV, 467 p.
- Weller R.A., Anderson J.L., Arribas A., Dickinson R.E., Goddard L., Kalnay E., Kirtman B., Koster R.D., Richman M.B., Saravanan R., Waliser D., Wang B. (Committee on Assessment of Intraseasonal to Interannual Climate Prediction and Predictability) (2010). Assessment of Intraseasonal to Interannual Climate Prediction and Predictability. National Research Council. National Academies Press. ISBN: 0-309-15184-8, pp 192.
- Wilks, Daniel S., Statistical methods in the atmospheric sciences (2006). ISBN 13:978-0-12-751966-1.
- WMO. New Attachment II-8 to the Manual on the GDPFS (WMO-no.485), Volume I: Standardised Verification System for Long-Range Forecasts.CBS-DPFS/ET-LRF/Final Report, p59-84

ANEXO I

Descripción de las posibles configuraciones al aplicar el FA

A continuación se detallan y numeran diferentes posibilidades a la hora de aplicar el método FA, en función de: a) utilización o no de validación cruzada; b) utilización de valores estandarizados o de las anomalías de las variables al realizar el análisis de covarianza máxima (MCA); y (c) periodo de referencia usado para el cálculo de la función *prior*.

1. Configuración FA1:

- Se hace validación cruzada al calcular el FA.
- Se trabaja con las anomalías de los valores medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo escogido para calcular la función prior es el mismo que el de los *hindcast* de los modelos dinámicos (1988 - 2008).

2. Configuración FA2:

- Se hace validación cruzada al calcular el FA.
- Se trabaja con los valores estandarizados medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo escogido para calcular la función prior es el mismo que el de los *hindcast* de los modelos dinámicos (1988 - 2008).

3. Configuración FA3:

- Se hace validación cruzada al calcular el FA.
- Se trabaja con los valores estandarizados medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo de referencia usado para el cálculo de la función *prior*. es 1960 - 2010-

4. Configuración FA4:

- Se hace validación cruzada al calcular el FA.
- Se trabaja con las anomalías de los valores medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo de referencia usado para el cálculo de la función *prior*. es 1960 – 2010.

5. Configuración FA5:

- No se hace validación cruzada al calcular el FA.
- Se trabaja con las anomalías de los valores medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo escogido para calcular la función prior es el mismo que el de los *hindcast* de los modelos dinámicos (1988 - 2008).

6. Configuración FA6:

- No se hace validación cruzada al calcular el FA.
- Se trabaja con los valores estandarizados medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo escogido para calcular la función prior es el mismo que el de los *hindcast* de los modelos dinámicos (1988 - 2008).

7. Configuración FA7:

- No se hace validación cruzada al calcular el FA.
- Se trabaja con los valores estandarizados medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo de referencia usado para el cálculo de la función *prior*. es 1960 – 2010.

8. Configuración FA8:

- No se hace validación cruzada al calcular el FA.
- Se trabaja con las anomalías de los valores medios trimestrales de las dos variables, precipitación acumulada y temperatura a 2 m a la hora de calcular el MCA.
- El periodo de referencia usado para el cálculo de la función *prior*. es 1960 - 2010

Evaluación de las ocho configuraciones consideradas para realizar el FA

Para decidir qué configuración utilizar a la hora de realizar el FA, se ha comparado, para cada modelo dinámico, la media de los coeficientes de correlación en el área de la Península Ibérica, para las anomalías de precipitación acumulada y temperatura a 2 m., entre observaciones y predicciones, por trimestres y para los *lead-time* 1, 2 y 3. Después de analizar los valores obtenidos, de entre las configuraciones numeradas FA1 al FA4, se ha decidido tomar la FA3 que, en general, es con la que se obtienen los mejores resultados, aunque la configuración F4 presenta resultados bastante similares. Las configuraciones FA5, FA6, FA7 y FA8 se han aplicado solamente como caso de estudio para comprobar la influencia que tiene el hecho de que se aplique el método FA sin hacer validación cruzada (es decir, sin eliminar los datos del año con el que estamos trabajando). Aunque como es de esperar los valores de los coeficientes de correlación obtenidos son bastante mayores que los obtenidos cuando se aplica FA con validación cruzada (FA1, FA2, FA3 y FA4), no es razonable hacer el ajuste de la recta de regresión usando los datos de un determinado año, si luego, a partir de ella, se va a predecir el valor de la variable para ese año. En la tabla 5 se muestran los resultados obtenidos al aplicar las ocho configuraciones con el modelo S4.

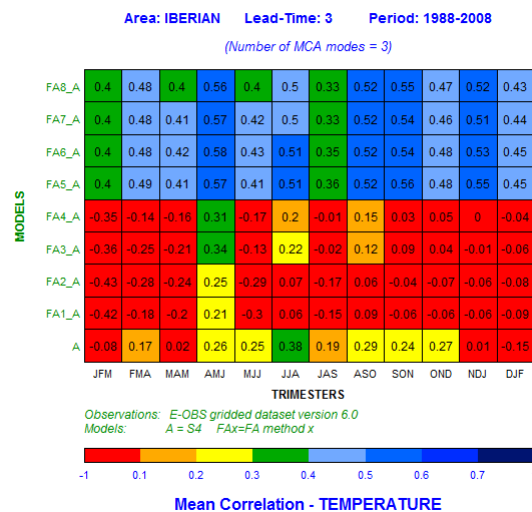
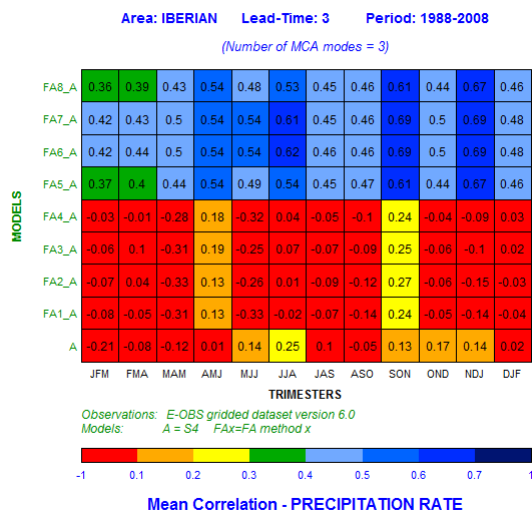
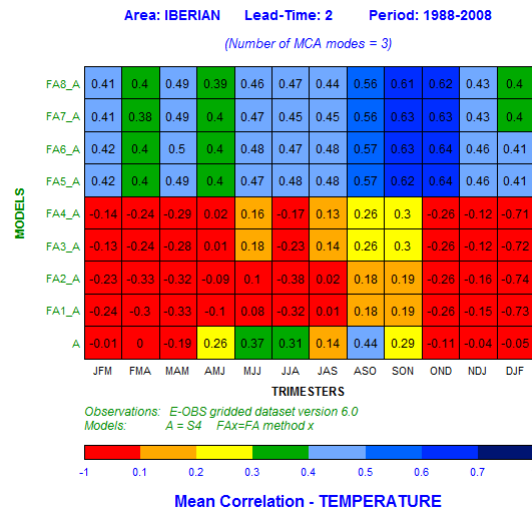
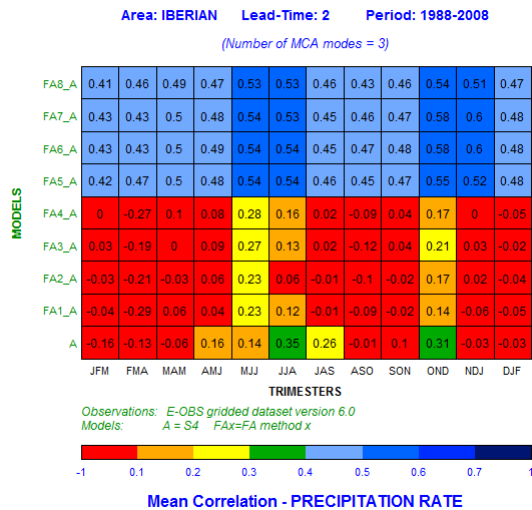
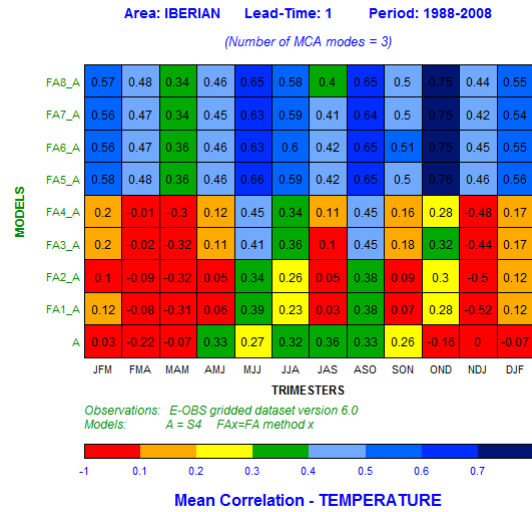
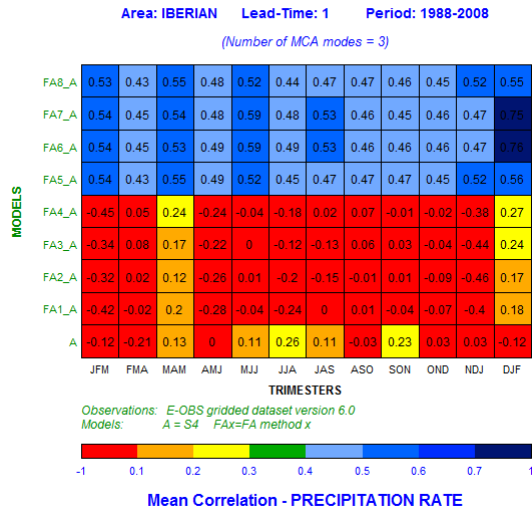


Tabla 5. Media de los coeficientes de correlación en el área de la Península Ibérica, para las anomalías de precipitación acumulada y temperatura a 2 m., entre observaciones y predicciones, por trimestres para lead-time 1, 2 y 3, para el modelo S4 y las ocho configuraciones analizadas al aplicar el FA sobre dicho modelo.

ANEXO II

Número de modos retenidos en el MCA

La figura 3 muestra la varianza explicada (en %) en función del número de modos retenidos en el análisis de covarianza máxima (MCA), utilizado para reducir el número de dimensiones del problema a la hora de aplicar el método del FA, para la variable estandarizada temperatura media trimestral a 2 m, considerando sólo el modelo S4. Resultados similares se han obtenido para la precipitación y los otros modelos dinámicos. En vista de los resultados, se retienen 3 modos en el MCA.

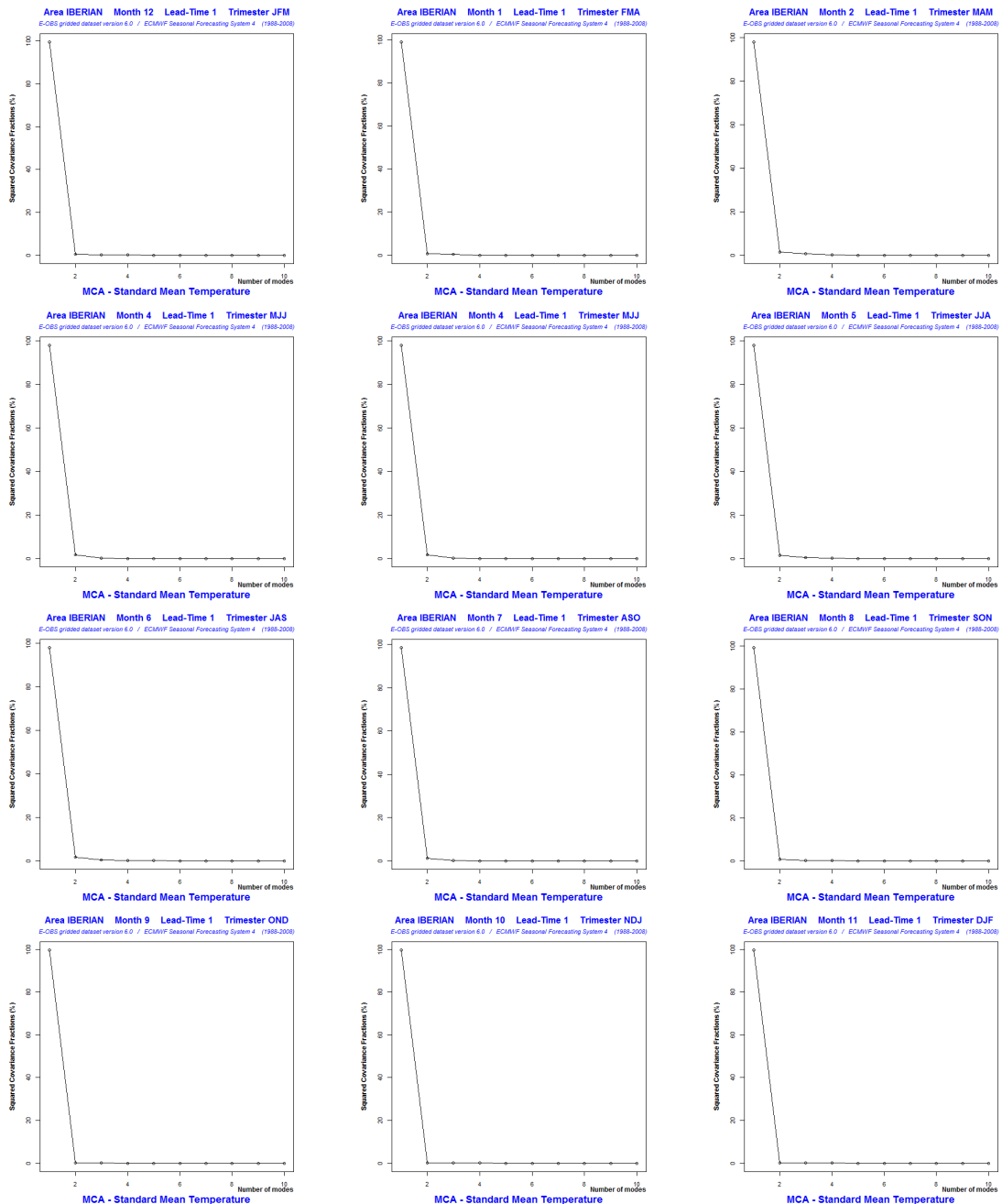


Figura 4. Varianza explicada (en %) en función del número de modos retenidos en el análisis de covarianza máxima (MCA), utilizado para reducir el número de dimensiones a la hora de aplicar el método del FA, para la variable estandarizada temperatura media trimestral a 2 m, considerando sólo el modelo S4.