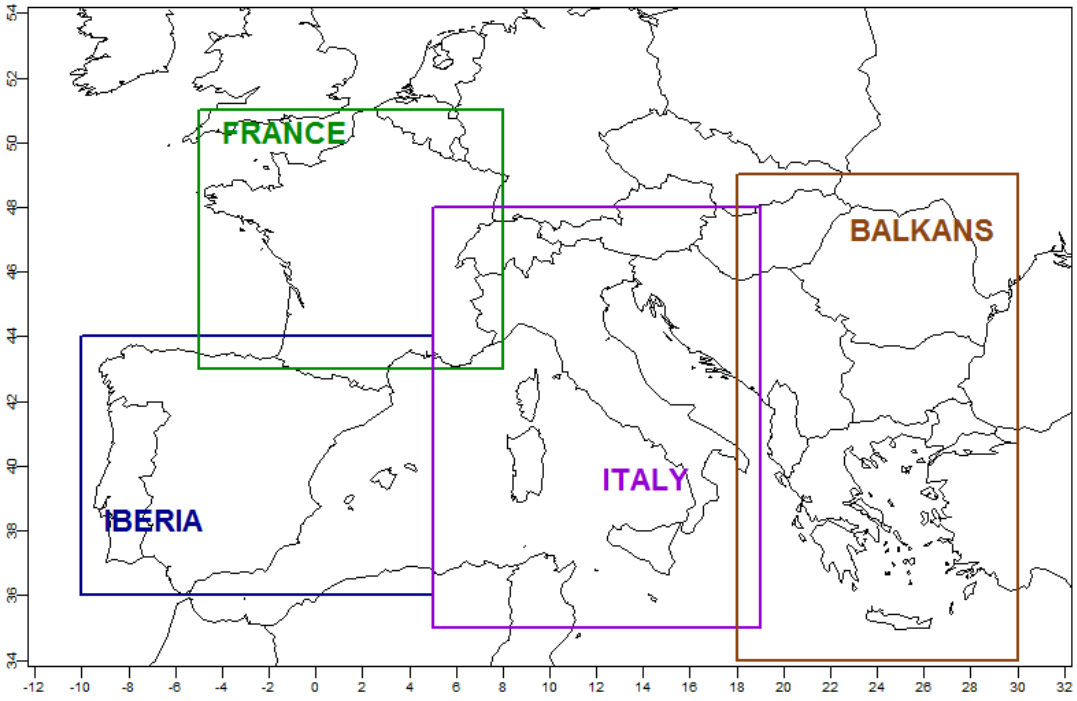
CALIBRATION AND COMBINATION OF SEASONAL FORECAST OVER SOUTHERN EUROPE



Eroteida Sánchez García
José Voces Aboy
Ernesto Rodríguez Camino

GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA, ALIMENTACIÓN Y MEDIO AMBIENTE
AEMet
Agencia Estatal de Meteorología

Edita:

© Ministerio de Agricultura, Alimentación y Medio Ambiente
Agencia Estatal de Meteorología
Madrid, 2014

Catálogo de Publicaciones de la Administración General del Estado:
https://cpage.mpr.gob.es

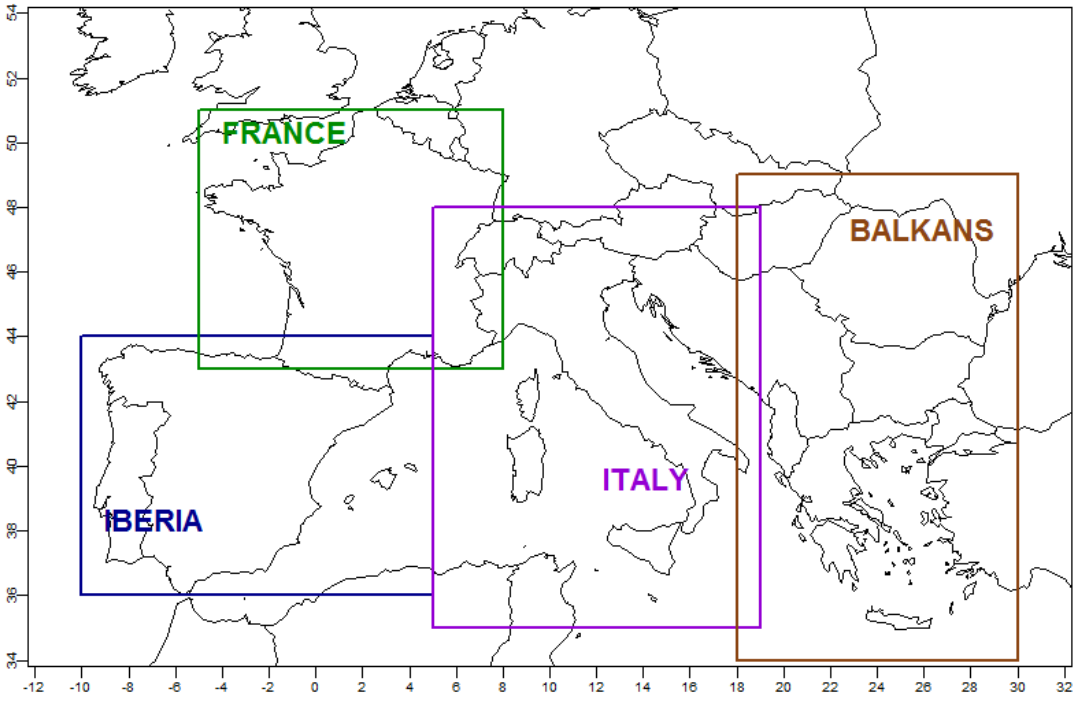
NIPO: 281-14-014-X
https://doi.org/10.31978/281-14-014-X

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

@Aemet_Esp

https://www.facebook.com/AgenciaEstataldeMeteorologia

CALIBRATION AND COMBINATION OF SEASONAL FORECAST OVER SOUTHERN EUROPE



Eroteida Sánchez García
José Voces Aboy
Ernesto Rodríguez Camino

GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA, ALIMENTACIÓN Y MEDIO AMBIENTE
AEMet Agencia Estatal de Meteorología

2014

Abstract

Verification of temperature and precipitation seasonal forecasts from four different operational systems - European *Centre for Medium-Range Weather Forecasts* (ECMWF) *system 4*, *Météo-France system 3*, *UK Met Office system 3* and *National Center for Environmental Prediction* (NCEP) system version 2 - for different seasons, lead times, variables and sub-regions over Southern Europe is computed based on available hindcasts. The impact of calibration and combination of seasonal hindcasts using different setups of a Bayesian scheme has also been discussed. Although results show relatively low skill as a consequence of the low predictability at seasonal scale over mid-latitudes, there is a noticeable consistency among models. As expected over Southern Europe, scores for temperature are better than for precipitation. We also show and discuss windows of opportunity associated to certain seasons/variables/models/regions.

This study has been done within the EUPORIAS (European Provision of Regional Impact Assessment on Seasonal and decadal timescale) project.

"EUPORIAS is financed by the European Commission through the 7th Framework Programme for Research, Grant Agreement 308291"

CONTENTS

1. Introduction

The chaotic features of the atmosphere limit the predictability of deterministic weather forecasts up to 10-15 days. Beyond this range, the predictability of atmospheric conditions has only sense from a statistical point of view and therefore forecasts must be expressed in probabilistic terms (Murphy and Winkler, 1984). The main sources of uncertainty of forecasts at seasonal time scales come from the insufficient knowledge of initial conditions for the climate system and the lack of accuracy of climate models (Curry and Webster 2011, Knutti 2010, Slingo and Palmer 2011). The first source of uncertainty is explored using ensemble techniques based on independent forecasts from slightly different initial conditions (Gneiting and Raftery 2005; Palmer 2000). The second source of uncertainty is estimated, among other techniques, by combining different climate model integrations (Doblas-Reyes et al. 2009). Some authors have proposed several approaches to combine and calibrate seasonal forecasts generated by seasonal forecasts based on different models and also on empirical algorithms (see, e.g., Palmer et al. 2004, Coelho et al. 2004, Stephenson et al. 2005). Predictability at seasonal time scale is highly dependent on particular atmospheric and oceanic modes of variability, regions, seasons and variables. Operational seasonal forecasts are frequently circumscribed to temperature and precipitation (Kirtman y Pirani 2008). The weak atmospheric predictability in mid-latitudes, and in particular over the Mediterranean region, has given preference to simple and robust seasonal forecast based on terciles (Doblas-Reyes 2010).

The main objective of this paper is to gain knowledge about the skill of the here considered models as a function of the season, variable and region in order to improve the operational seasonal forecast activities in the Mediterranean region. We evaluate for each region the skill of seasonal forecasts and identify windows of opportunity or circumstances with higher skill. These windows of opportunity may be linked to certain teleconnections, seasons, variables or specific forecast systems. The windows of opportunity can be produced by signals from several processes interacting constructively, but in many cases their reasons for such occurrence are still unclear. In this report we study and discuss the skill of state-of-the-art operational seasonal forecast models for different domains within the Mediterranean region, for different seasons and for different variables. For this particular study we have first considered direct outputs from the following four models: *European Centre for Medium-Range Weather Forecasts* (ECMWF) *system 4*, *Météo-France system 3*, *UK Met Office system 3* and *National Center for Environmental Prediction* (NCEP) *system version 2*. Then, we have also applied the Bayesian calibration and combination method described by Stephenson et al. (2005) with different settings in an attempt to improve the scores of direct model outputs from individual seasonal forecasting systems.

Section 2 describes data sources both from seasonal forecast models and observations. Calibration and combination methods are analyzed in Section 3. Verification scores here applied are summarized in Section 4. Results are presented in Section 5. Finally, conclusions and way forward are discussed in Section 6.

2. Data

The E-OBS gridded dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) has been used for the observational data (Haylock et al. 2008). The E-OBS (version 6.0) gridded data set provides daily surface temperature and precipitation at 0.5º x 0.5º latlon horizontal resolution (for the ENSEMBLES European domain) from 1950 up to now. The data set is based on daily station data available from the ECA&D website (http://www.ecad.eu) together with additional (restricted) data obtained by the STARDEX and ENSEMBLES projects. From the original data, three monthly averaged anomaly values of temperature and precipitation upscaled to 1º x 1º horizontal resolution were computed to verify seasonal models outputs. These derived data were also used by the Bayesian method applied for calibration and combination of seasonal forecasts (see Section 3).

Hindcasts of the following seasonal coupled atmosphere-ocean models have been used for their verification at seasonal time scales:

The *European Centre for Medium-Range Weather Forecasts* (ECMWF) *system 4* (S4) consists of Cy36r4 of the Integrated Forecast System (IFS) at TL255 resolution (80 km grid point resolution) coupled with the ORCA1 configuration of the Nucleus for European Modelling of the Ocean (NEMO). The IFS has 91 levels and includes the whole stratosphere. Ocean initial conditions come from an assimilation system based on an advanced multivariate variational analysis with bias adjustments. Atmosphere and land surface initial conditions come from a mixture of ERA Interim and ECMWF operations, and an offline run of the HTESSEL surface model (Kim et al. 2012, Molteni et al. 2011). The ensemble size is 15 members for the hindcast period 1981-2010.

The Météo-France system 3 (MF3) consist of the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) version 4 for the atmospheric component (Batté and Déqué 2011) coupled with ORCA, developed by LOCEAN, for the ocean model. The ocean initial conditions are prepared by MERCATOR. The atmospheric model has TL127 horizontal resolution with a Gaussian grid spacing of about 160 km and 91 levels with the stratosphere not fully resolved. The ensemble size is 11 members for the hindcast period 1981-2010.

The UK Met Office system 3 (UKMO3) has an atmospheric component with a spatial resolution of 2.5º x 3.75º grid and 85 vertical levels. The ocean model has a basic resolution of 1.25º, with meridian refinement to 0.3º at the equator and 75 vertical levels. Ocean initial conditions are taken from the Met Office ocean analysis system. The ensemble has 15 members for the hindcast period 1987-2008.

The National Center for Environmental Prediction system version 2 (CFSv2) has an atmospheric component with a spatial resolution of 100 km and 64 vertical levels (Kim et al. 2012, Saha et al. 2013, Yuan et al. 2011). The ocean component is the Geophysical Fluid Dynamics Laboratory Modular Ocean Model (MOM4) version 4 with horizontal resolution of 0.5º, refined at 0.25º between 10ºN and 10ºS, and 40 vertical levels. Although the ensemble for the hindcast has 28 members, initialized at different days and hours, we have here only used the 12 more recent members. The hindcast period ranges between 1982 and 2010.

3. Methodology

The Forecast Assimilation (FA) is a Bayesian method which has been used to combine the four forecast systems here analyzed. The FA method has proved to be competitive against the Simple Multi-Model (SMM) method -where all single models are equally weighted- and other combination methods (Lage et al. 2013). The FA method calibrates and combines predictions from several sources with prior (historical) empirical information (Stephenson et al. 2005). A useful feature of FA is that it allows predicted patterns to be shifted spatially in order to correct for model biases in coupled model predictions. In other words, the procedure accounts for inter-grid point dependencies. The FA method is a consistent probabilistic approach which can be used for combining historical (climatological) information with dynamical model ensemble mean forecasts. The FA method, as any other Bayesian method, is firmly based on rigorous probability theory and so can provide well-calibrated probability forecasts.

Different setups to apply the FA method are discussed in Annex I, depending on the application –or not- of cross-validation, the usage of standard values or anomalies values for temperature and precipitation in the maximum covariance analysis (MCA) method and the reference period taken for the computation of the prior function. According to this discussion, the FA3 setup is selected (using cross-validation, standard values and 1960-2010 as a reference period). All results shown in Section 5 are computed using the FA3 configuration. Annex II discusses the retained number of modes by the MCA.

With no access to a coupled model forecasts, M, the only possible probabilistic assessment about the observable variable, y, has to be based on the assumption that future values of y will behave like they did in the past. For example, the probability distribution of y_i at some time t_i can be estimated by using the climatological probability density function, $p(y_i)$, estimated from historical observations. In Bayesian theory, $p(y_i)$ is known as the prior distribution and encapsulates prior knowledge about likely possible values of y_i which is usually known from past experience, in our case from climatology. A more informative prior could be also provided by some empirical model. However, when a particular model forecasts, M, are known for the future, it is then possible to update the prior $p(y_i)$ to obtain the

conditional posterior distribution $p(y_f|x_f)$. In other words, this is the probability distribution of y_i given that the forecast $M(x_i)$ is known, being x_j the model state variables. Conditioning on forecasts helps to reduce the uncertainty about future values of y_i (Jolliffe and Stephenson 2003, their chapter 9). The prior probability density when combined with a set of numerical weather predictions yields a conditional posterior distribution posterior probability. The posterior distribution $p(y_f|x_f)$ is found from the prior $p(y_i)$ by making use of Bayes' theorem.

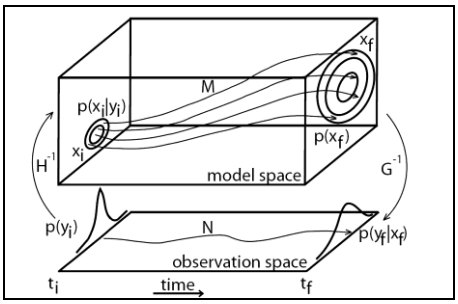


Fig. 1 - Conceptual Framework for forecasting.
Fuente: Stephenson et al.(2005)

In this paper we use as prior distribution the climatology from the E-OBS v6.0 observational database. As numerical forecasts, *M*, we have alternatively applied each of the four dynamical models here analyzed and also a combination equally weighted of the four models. In order to extract leading co-varying modes from model predictions and observational data, the maximum covariance analysis has been applied (von Storch and Zwiers 1999).

The anomalies of the different prediction systems, computed as the difference between the forecasted and climatological values for each system, are obtained by cross-validated forecasts on data not used in the estimation, i.e., the year to be forecast is removed from the data set. Cross-validation is also used for the calibration/combination FA method. We have used as common calibration period 1988-2008 covered by the hindcasts of the four systems here analyzed.

4. Verification

Seasonal forecasts of temperature and precipitation obtained with the different forecasting systems here considered are verified using both deterministic and probabilistic skill scores. Statistical significance of all computed scores has been quantified by the p-value estimated using a bootstrapping non-parametric method (Wilks 2006).

The correlation between the predicted and the observed mean value of anomalies over the different land domains within the Mediterranean region (see Fig.2) is the only deterministic skill score computed both for temperature and precipitation. The score was computed for 12 different three-month periods and for lead times 1, 2 and 3.

From a probabilistic point of view, the following skill scores have been also computed for the same variables (temperature and precipitation), for 12 different three-month periods and for lead times 1, 2 and 3: the Ranked Probability Skill Score (RPSS) for terciles, and the Relative Operating Characteristic (ROC) area and the Brier Skill Score (BSS) for two events (values above/below the upper/lower tercile). A complete definition of these scores can be found in Wilks (2006).

The Ranked Probability Skill Score (RPSS) is a generalization of Ranked Probability Score (RPS) based on a reference forecasting system. The RPS averages squared "error" in the cumulative probabilistic forecasts. Positive values of RPSS indicate more skill than the reference system, usually the climatology.

The ROC curves measure discrimination and skill. If the category of interest is above-normal, the score based on the ROC area indicates the probability of successfully discriminating above-normal observations from normal and below-normal observations. The ROC area ranges from 0% to 100%,

with a score of 50% representing no skill, 100% indicating perfect discrimination, and 0% indicating perfectly bad discrimination. It is important to stress that ROC curves are measuring only the discrimination ability between two possible results, but it is not informative about reliability as it is not sensitive to bias.

The Brier Score (BS) is the most common verification method for probabilistic forecasts, as it, has a mathematical structure similar to the Mean Square Error (MSE). BS measures the difference between the forecast probability of an event (p) and its occurrence (o), expressed as 0 or 1, depending on whether the event has occurred or not. As with RMSE, the BS is negatively orientated, i.e. the lower, the "better". The Brier Skill Score (BSS) is conventionally defined as the relative probability score compared with the probability score of a reference forecast.

All the ensemble members from 1988 to 2008 are taken into account in order to compute the lower and upper terciles of the hindcasts. The terciles of the observation data are also computed over the same period. We assume a normal distribution function to calculate terciles of the analysed Bayesian methods.

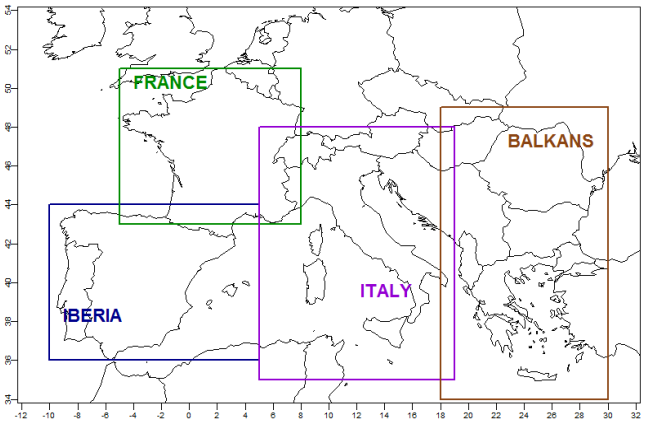


Fig. 2 – The selected four land domains over the Mediterranean region

5. Results

The four skill scores described in the previous section have been calculated taking into account all grid points on each selected Southern European region (see Fig. 3 to 6). The calculated value of each score over each selected domain is displayed using tables for anomalies of temperature and precipitation, for 12 different three-month periods and for lead-times 1, 2 and 3 (Table 1 to 24).

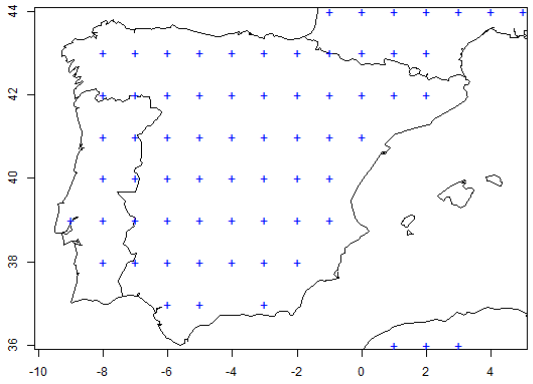


Fig. 3. Grid points over the Iberian Peninsula domain

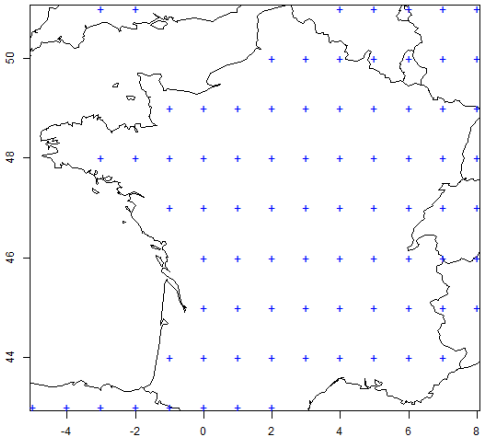


Fig. 4. Grid points over the France domain

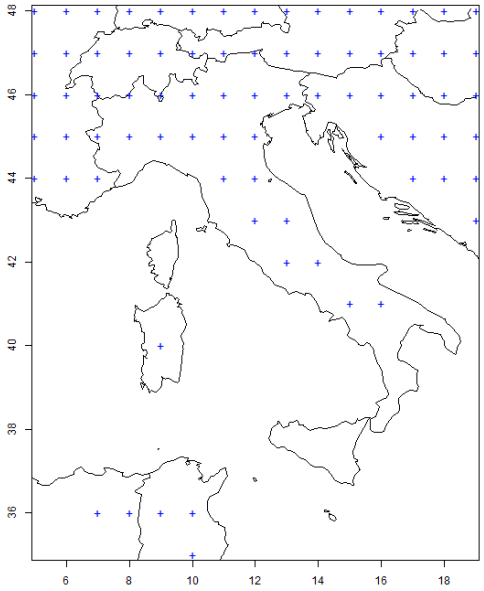


Fig. 5. Grid points over the Italy domain

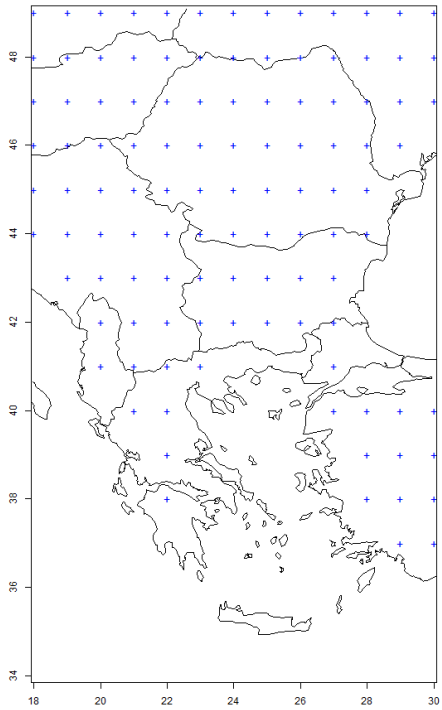


Fig. 6. Grid points over the Balkan domain

Tables 1, 2, 3 and 4 show the correlation coefficients between observations and seasonal forecasts in the four selected domains (Fig. 3, 4, 5 and 6), The correlation coefficients are referred to the precipitation and 2m temperature anomalies, computed for successive three-month periods and for lead times 1, 2 y 3. Tables 1-4 show results from: 1) direct model output from each of the four models here studied; 2) individual application to each model of the calibration and combination algorithm with FA3 configuration (see Annex I); and 3) combined application to the same algorithm to all four models. Temperature shows significant values (p-value 0.05) higher than 0.3 mainly centered in summer period and beginning of autumn for the Iberia and France domains. This window of opportunity is mostly common to all models here considered. The Italy and Balkan domains show a shift of this window towards spring and summer seasons. Apart from this clearly noticeable window, significant values appear only for certain models (UKMO3 for Iberia or MF3 for the rest of domains) and mainly restricted to certain months during autumn and winter. When the calibration and combination algorithm is applied, it is noticeable a slight improvement of the scores -although not general- which coincides

with models and periods showing higher correlation. It is important to underline the fact that the application of the algorithm combining the four models here considered do not produce better scores than the best scored model. This suggests that the best strategy would probably consist of eliminating the worst model(s) before the combined application of FA to several models. Precipitation shows generally lower skill than temperature and frequently non-significant values. In this last case, FA hardly improves skill. One should stress the high skill showed in JJA for lead-time 2 over Iberia mainly coming from the UKMO3 model. This result is not easily explained as scores for lead-times 1 and 3 do not behave accordingly.

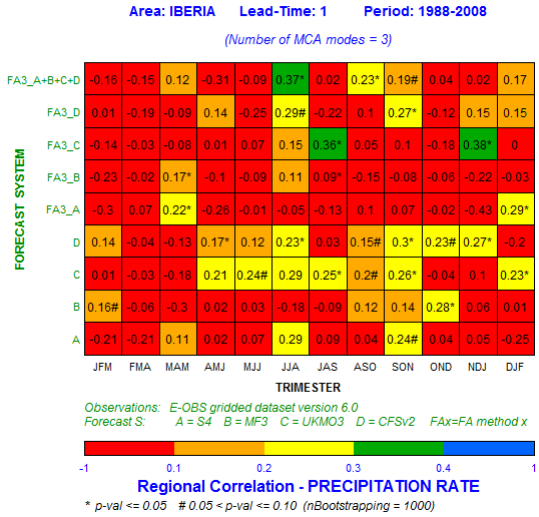
Comparison of correlation coefficients for different time horizons show some noticeable results. Firstly, longer lead-times are not always associated with skill degradation. Focusing, e.g., in the Iberian domain, the narrow winter window of opportunity for temperature associated to the UKMO3 model occurring in DJF for lead-time 1 moves to JFM for lead-time 2 and to FMA for lead-time 3. This behavior suggest a strong impact of initial conditions for the month of November. Secondly, certain skill barriers, linked to specific periods, either move to the right of Tables 1-4 when lead-time is increased or lessen or even disappear when previous initial conditions with higher predictability are used. The FMA temperature skill barrier for lead-time 2 shifts one month for lead-time 3. It is also worthwhile to point out that benefits associated with FA depends highly on each specific model. E.g., the application of FA to the UKMO3 model has always positive impact in terms of correlation coefficient both for all lead-times and three monthly periods. This fact has no correspondence with other models for the Iberian domain. However, when FA is applied over other domains no single model is able to improve skill, although there are differences among models and domains. For example, FA successfully improves scores when applied to MF3 over Italy for temperature, whereas FA deteriorates scores over the Balkan for all models. This suggests the need of different strategies or calibration setups for each model and for each domain. Probably some models have more potential for improvement through calibration and combination whereas others are already much optimized and their room for improvement is very limited, always depending on particular domains.

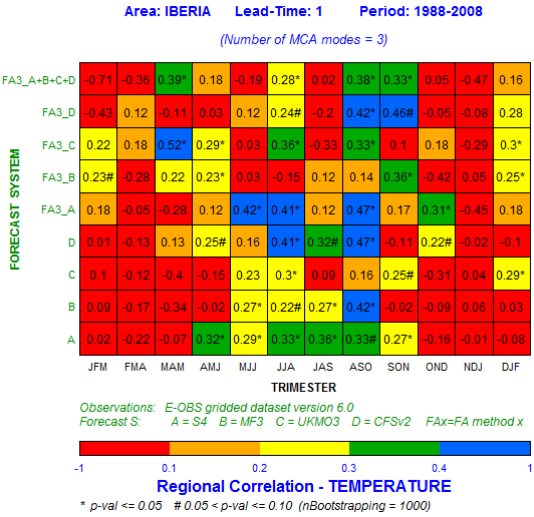
Tables 5, 6, 7 and 8 show the *Ranked Probability Skill Score* (RPSS) for accumulated precipitation and 2m temperature over the four domains here considered (Figs. 3 - 6). RPSS is not a symmetric skill score. It ranges from 1 (perfect forecast) to - ∞. Negative values indicate that the forecast is less accurate than climatology. Consequently, we focus only on positive values (green and blue in tables). Analogously to the correlation coefficient, visual inspection shows a general lack of skill for precipitation, whereas for temperature slight differences appear – mostly non-significant- among models and domains. Some of the features described for the correlation coefficient are also valid for RPSS, such as the occurrence of summer windows of opportunity for temperature with stronger dependency now on models and the appearance of a noticeable shifting towards spring over the most eastern domains.

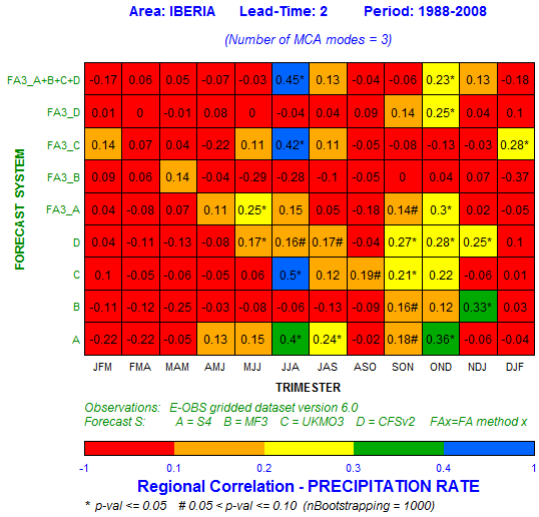
Tables 9, 10, 11 and 12 show lower tercile ROC area for accumulated precipitation and 2m temperature over the four domains (Figs. 3 - 6). Let's recall that this skill score measures the ability to discriminate of a forecast system. It ranges from 0 to 1 with values lower than 0.5 indicating no skill. Temperature seasonal forecasts for lead-time 1 show again the highest significant values (>0.70) during the summer season and late/early spring/autumn depending on the considered domain. Generally, skill is highest over Iberia and lowest over Balkans. In the winter period (DJF-JFM), skill of the UKMO3 model over Iberia and skill of the MF3 model for the rest of domains are noteworthy. Generally speaking, there is not a clear improvement when FA3 is applied, although again some models over certain domains show some improvement (e.g., UK over Iberia and Balkans, MF3 over France and Italy), whereas others (e.g., S4 over Balkans) practically does not show improvement in any case. Similarly to other scores, shifting of barriers or skill peaks associated to specific periods are also noticeable when lead-time increases. Although precipitation shows generally less skill than temperature in terms of lower tercile ROC area, it reaches significant values higher than 0.5 for one of the models (UKMO3) with lead-time 1 over Iberia. Again FA3 does not generally improve skill, being its impact much dependent on model and season.

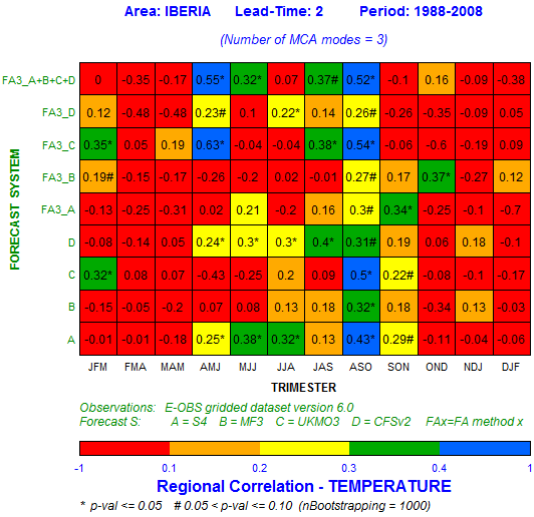
Tables 13, 14, 15 and 16 show upper tercile ROC area for accumulated precipitation and 2m temperature over the four domains (Figs. 3 - 6). Comparing against lower tercile ROC area, the improvement for temperature and lead-time 1 over Balkans is worth to mention, whereas it is not noticeable for other domains. In coincidence with other scores, FA3 does not generally improve skill except in particular cases.

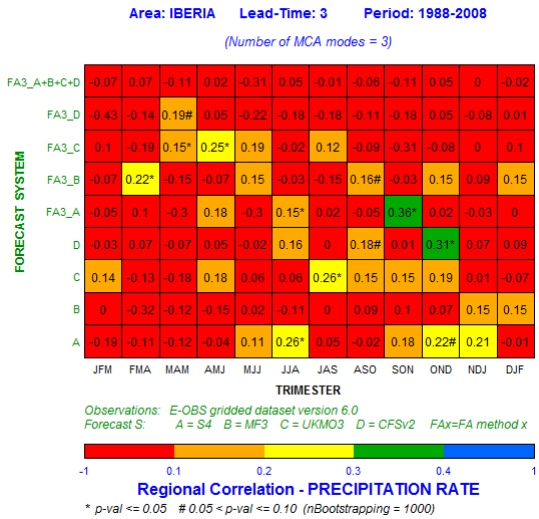
Tables 17 - 20 show lower tercile Brier Skill Score (BSS), and Tables 21 - 24 the corresponding for upper tercile BSS for accumulated precipitation and 2m temperature over the four domains (Figs. 3 - 6). Similarly to RPSS, the BSS is neither symmetric. It ranges from 1 (perfect forecast) to - ∞. Negative values indicate that the forecast is less accurate than climatology. Results, as expected, are rather similar to those obtained from RPSS. Skill, as measured by BSS, is slightly higher for temperature than for precipitation. It also appears the summer window of opportunity, which extends from April to October over France and Iberia, whereas it shifts toward spring (March-August) for lead-time 1 over Italy and Balkans. Nevertheless, there are still noticeable differences among models. With the exception of this window, positive values appear only scattered for some models and for some winter/autumn months. It is also noticeable for lead-time1 the occurrence of more positive values for the lower tercile than for the upper tercile over Iberia and France. Over Italy both lower and upper terciles BSS positive values appear with approximately the same frequency, and over Balkans positive values dominates for upper tercile.











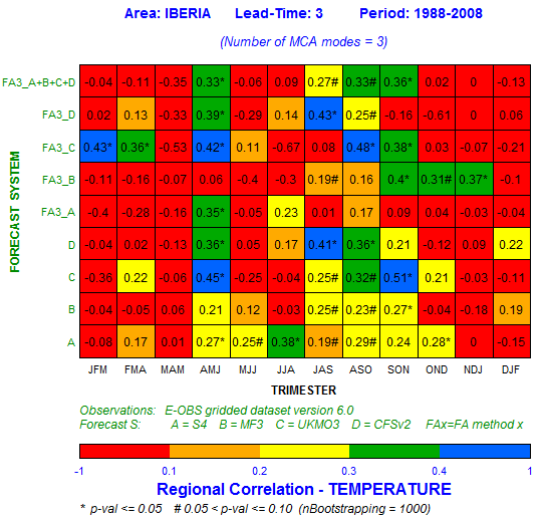
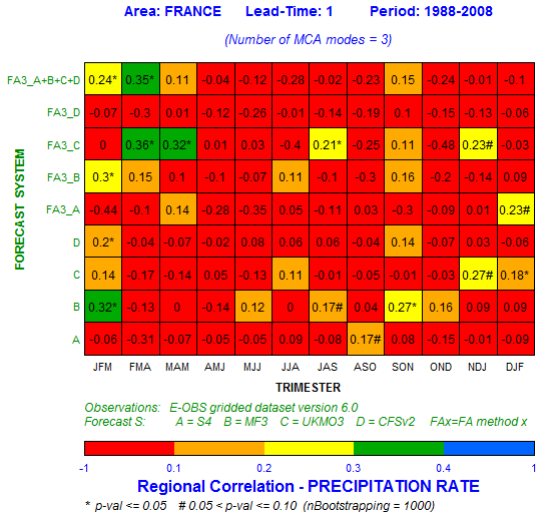
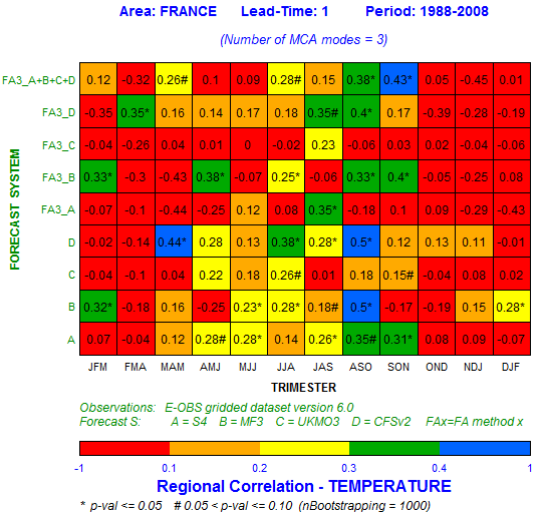
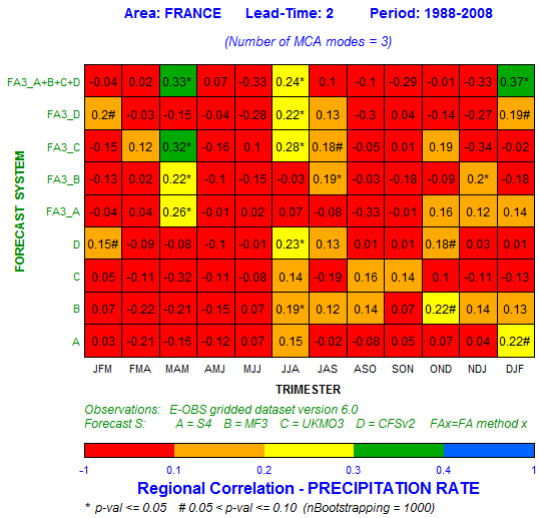
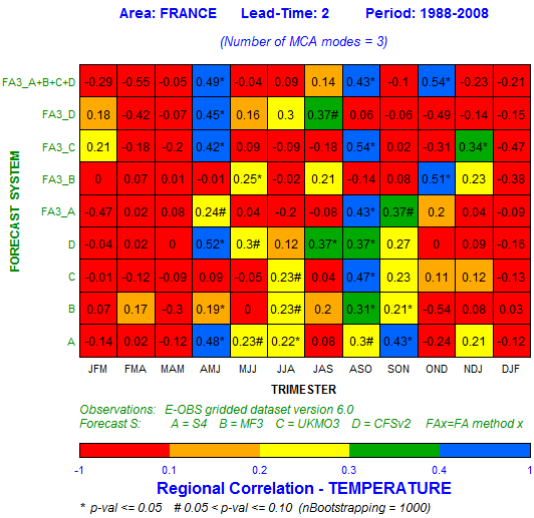


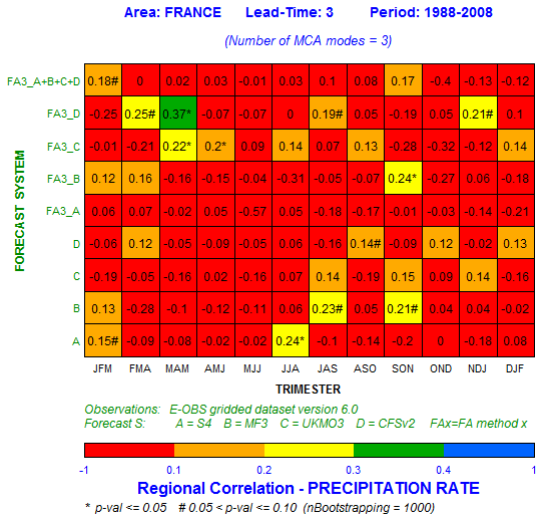
Table 1. Regional correlation coefficients between observations and forecasts computed for the anomaly values of temperature and total precipitation, for the 12 different three-month periods and for the lead-time 1, 2 and 3 over IBERIAN domain. The three-month periods for the seasonal forecasts are done, moved one by one for each column in the table, are shown in the X-axis. The direct output of the different models (S4, MF3, UKMO3 and CFSv2) and the FA3 algorithm outputs are represented in the Y-axis (see text for their description).











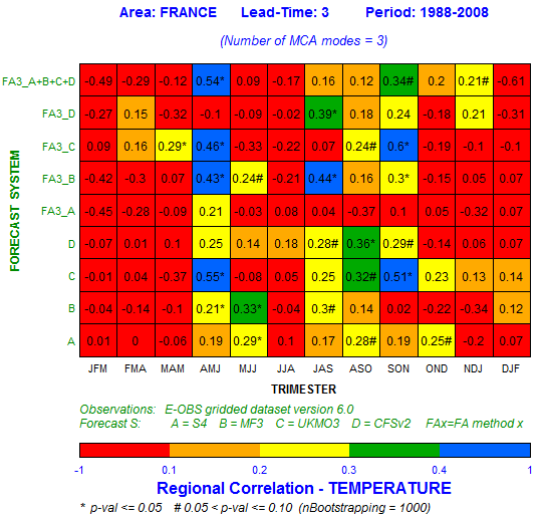
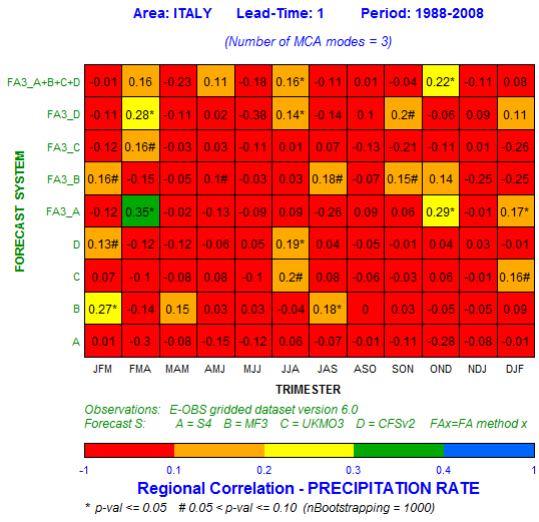
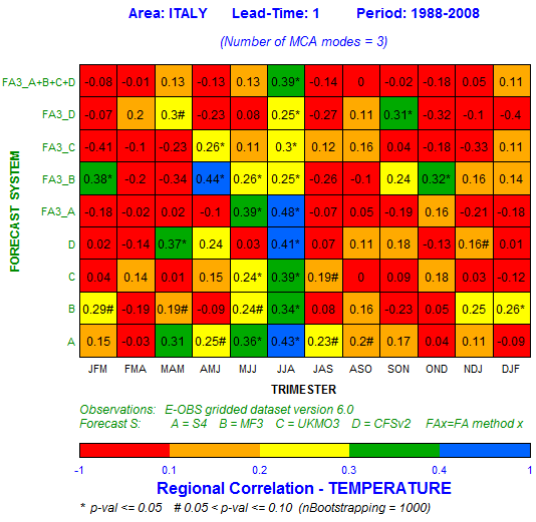
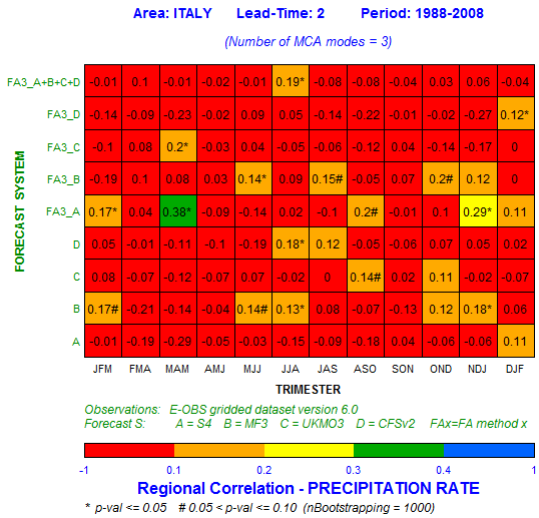
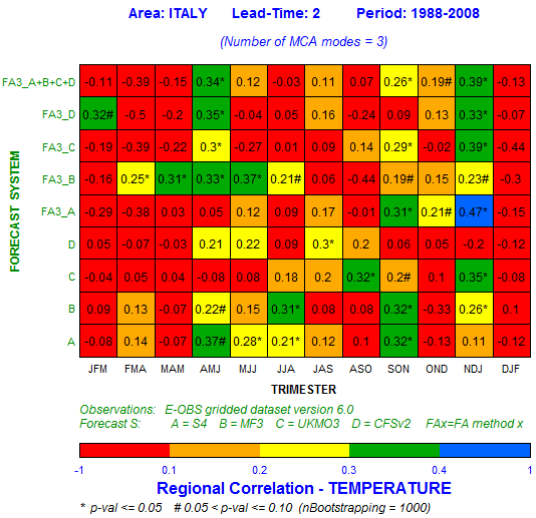


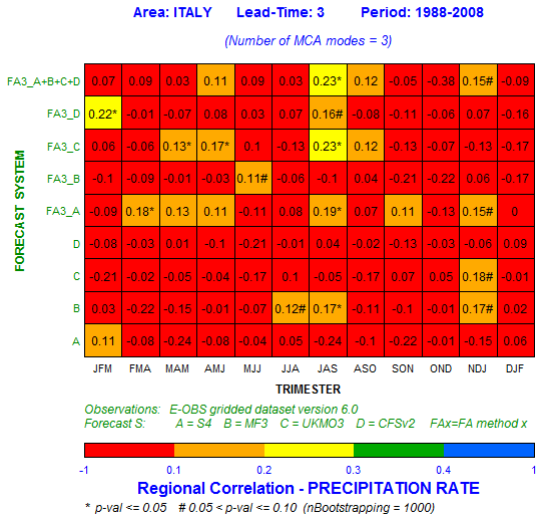
Table 2. The same as Table 1, but over the FRANCE domain.











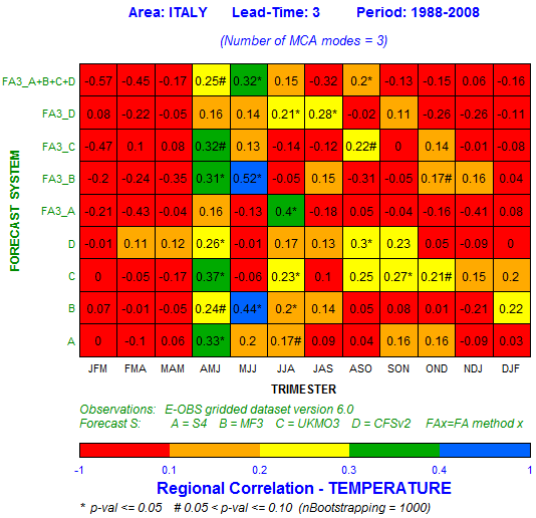
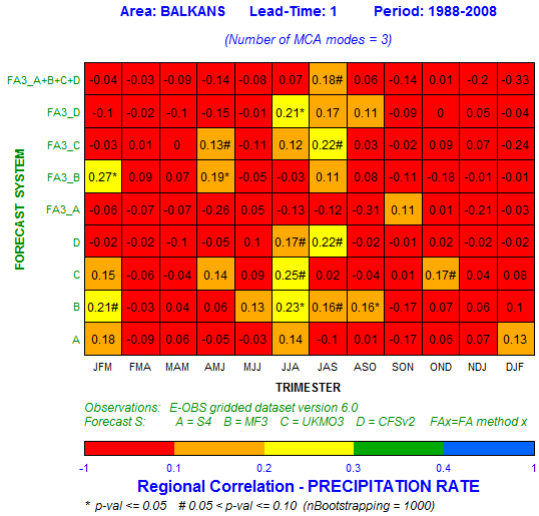
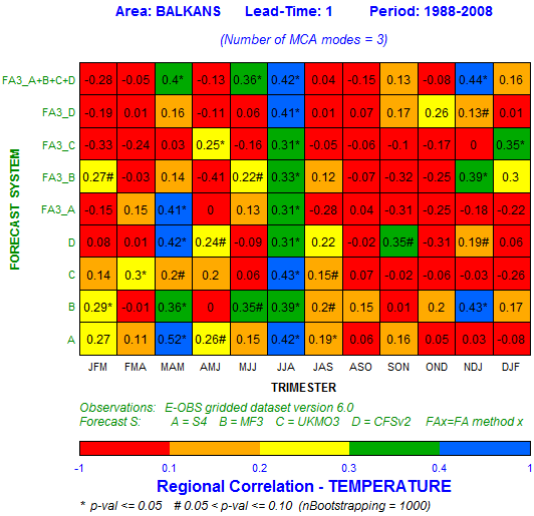
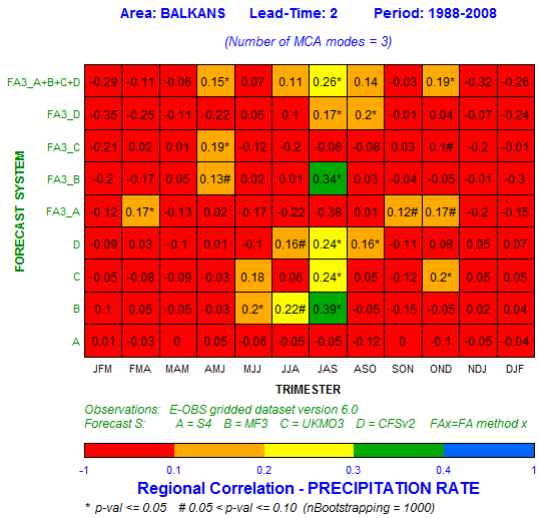
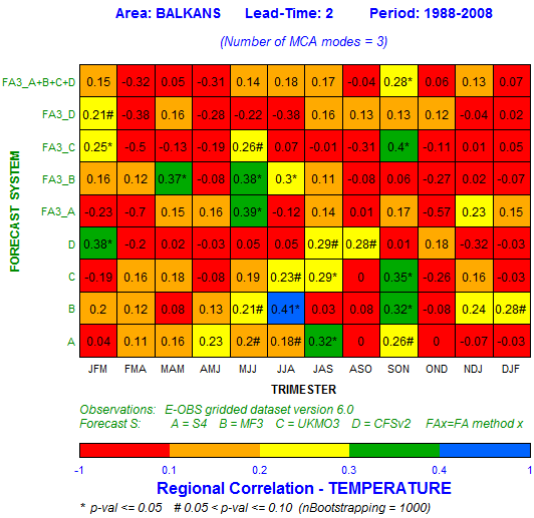


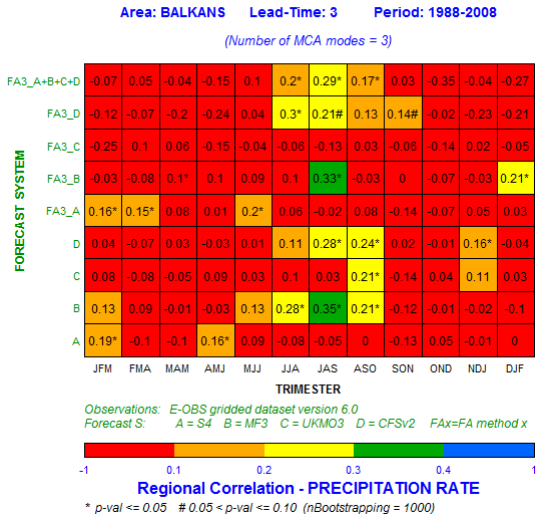
Table 3. The same as Table 1, but over the ITALY domain.











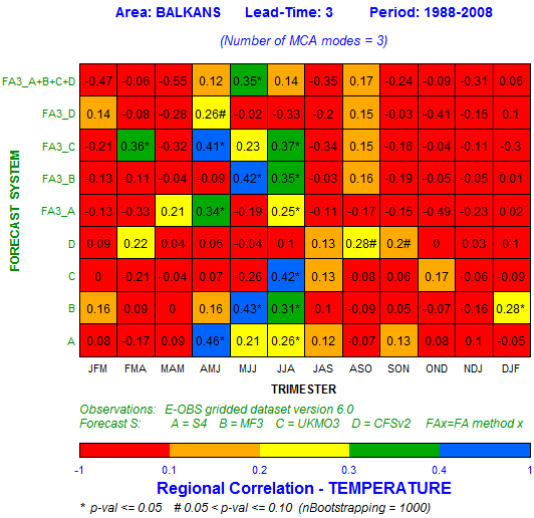
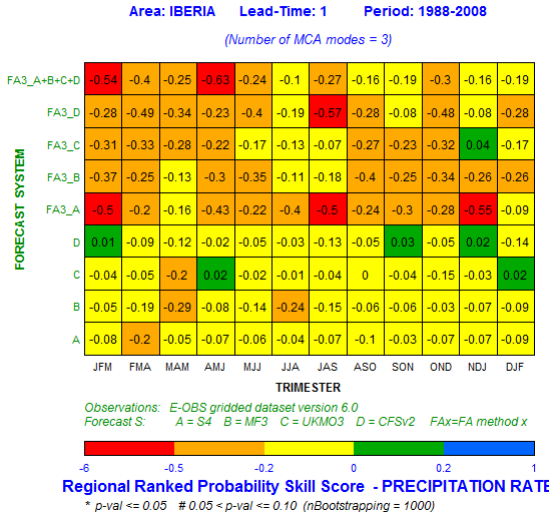
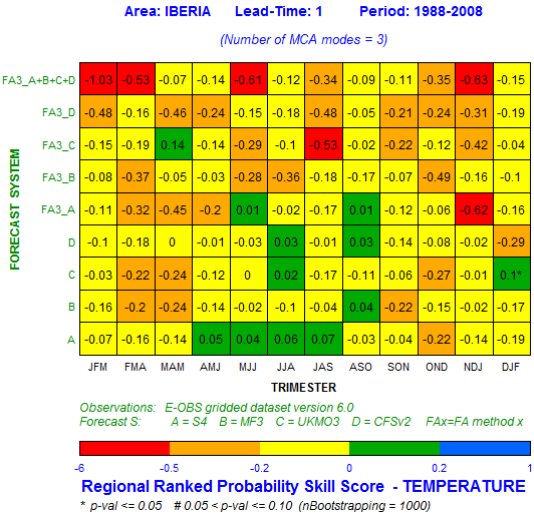
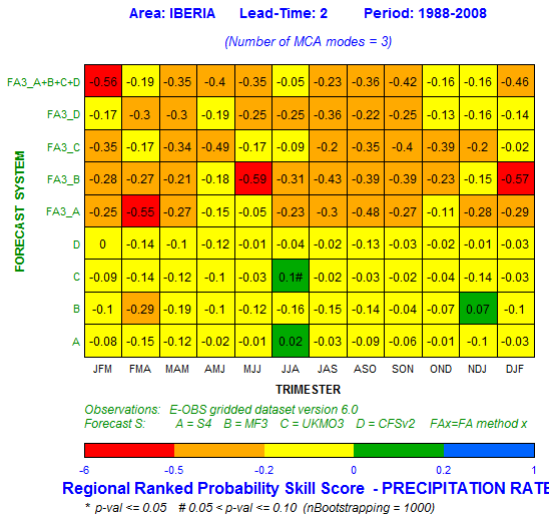
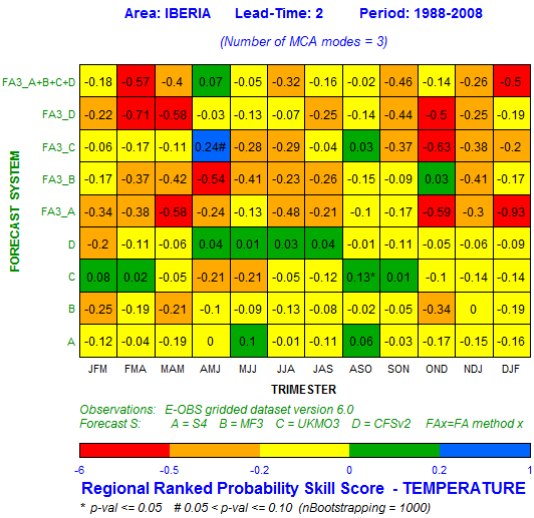


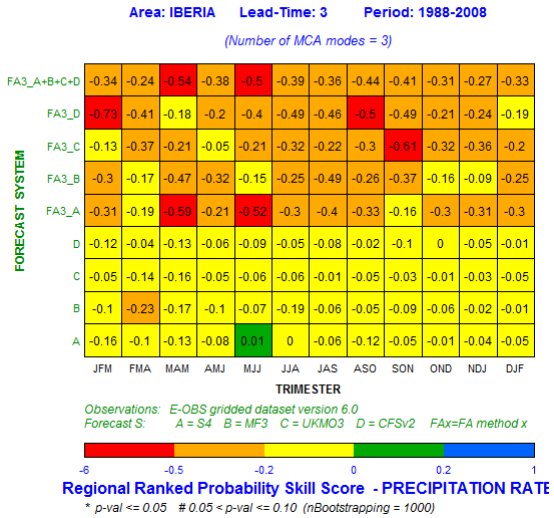
Table 4. The same as Table 1, but over BALKANS domain.











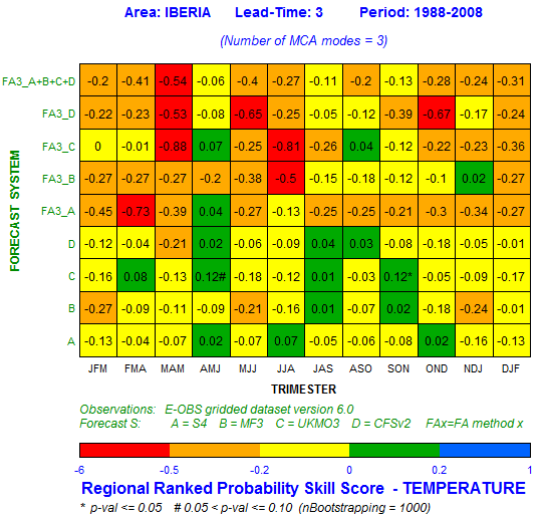
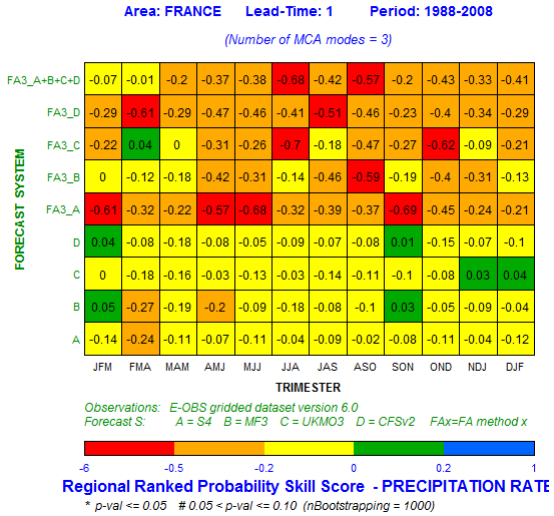
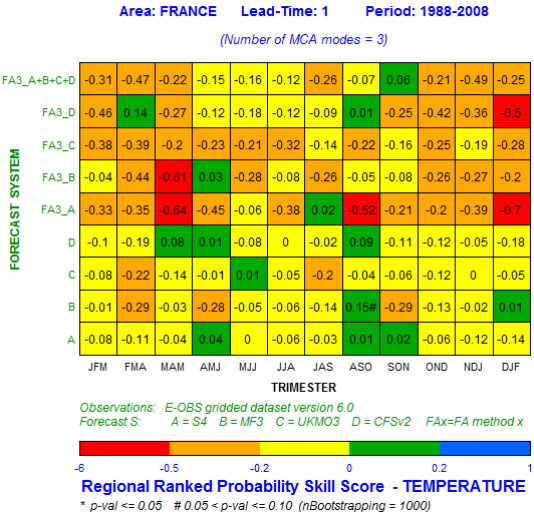
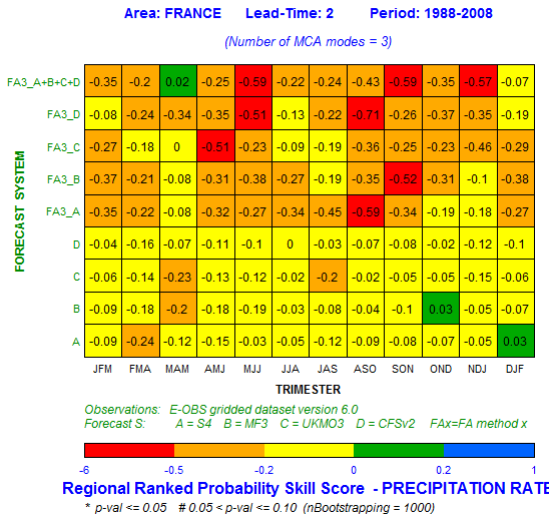
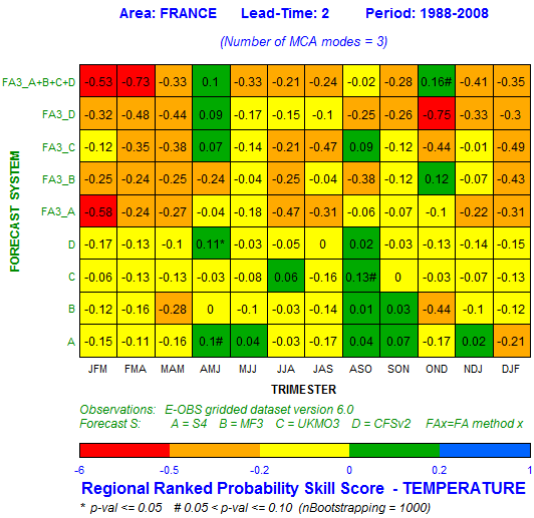


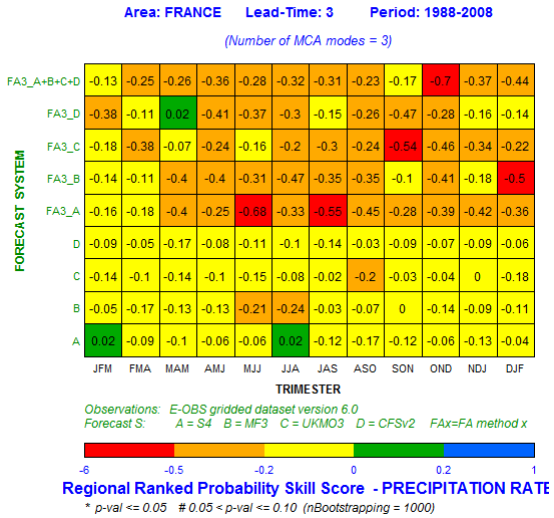
Table 5. The same as Table 1, but for the Ranked Probability Skill Score.











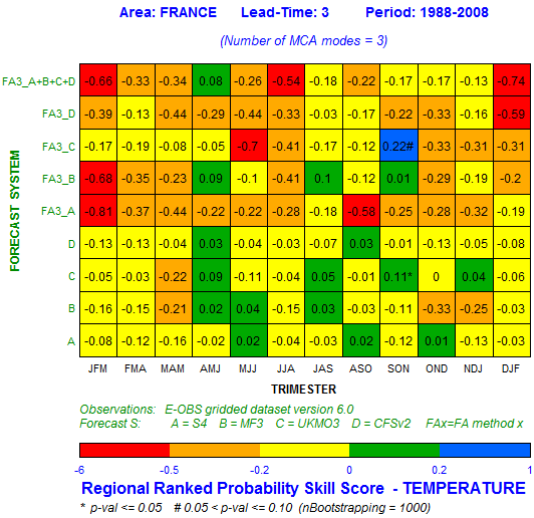
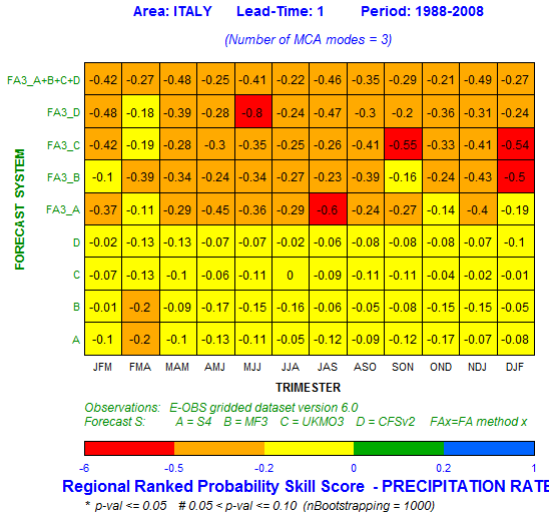
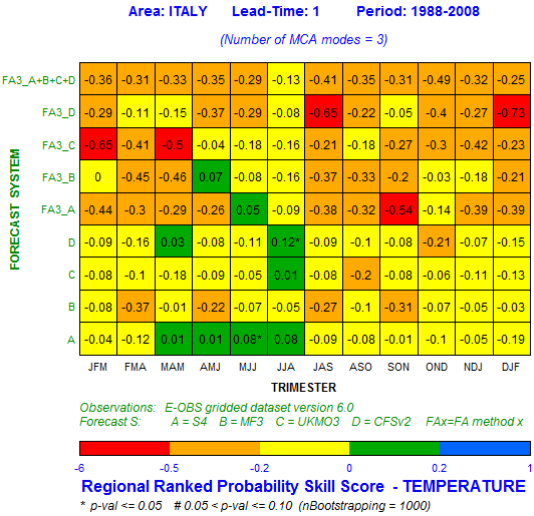
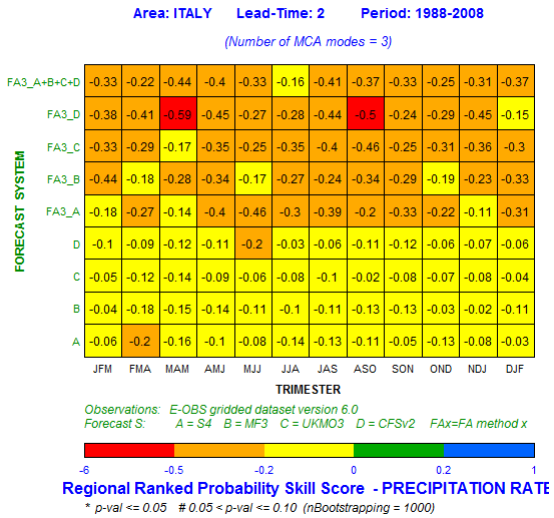
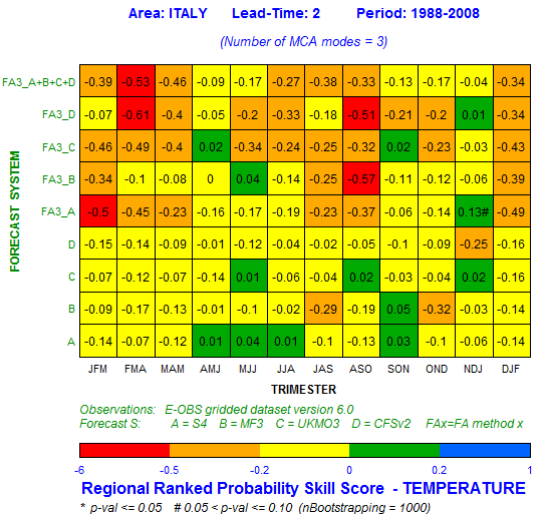


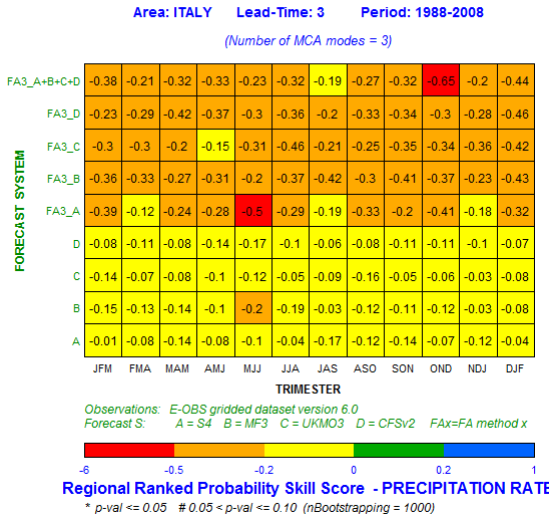
Table 6. The same as Table 1, but for the Ranked Probability Skill Score over FRANCE domain.











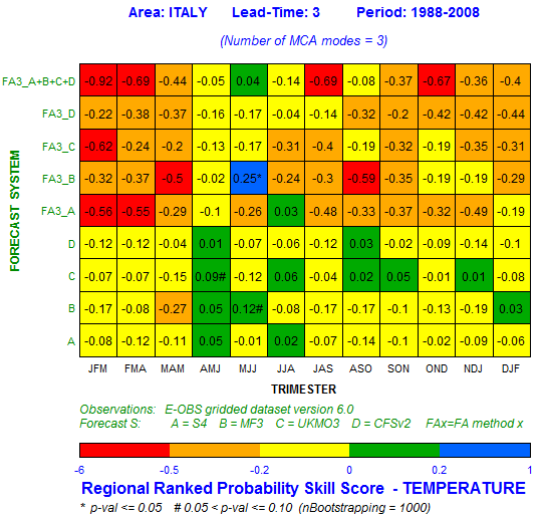
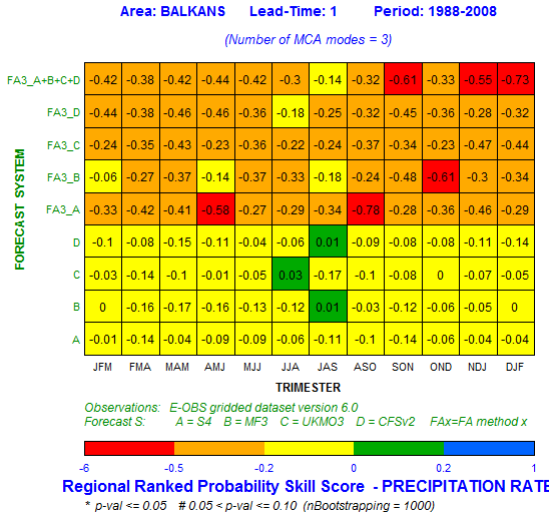
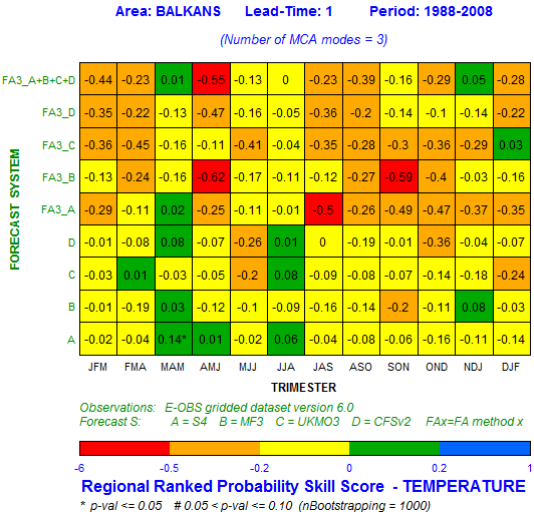
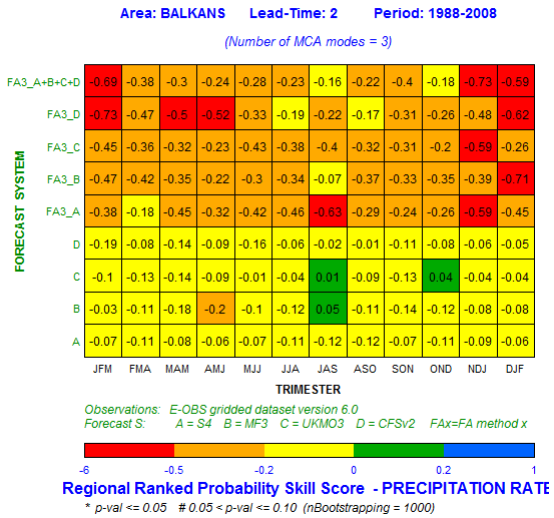
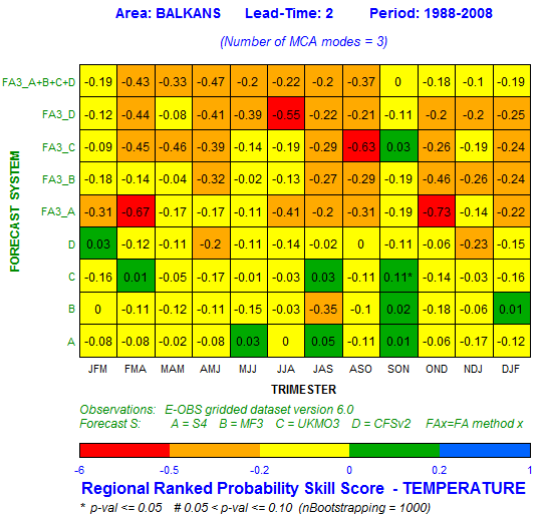


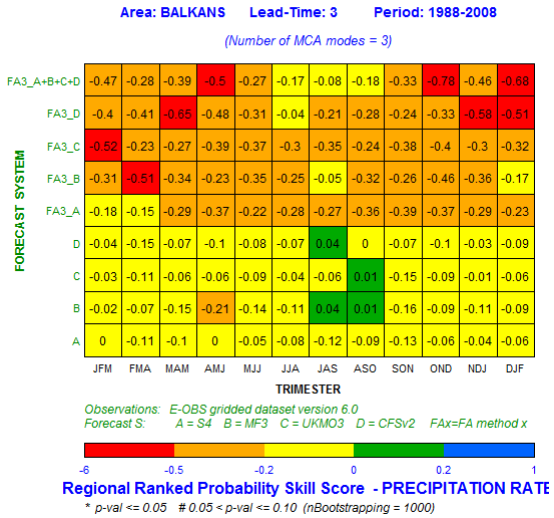
Table 7. The same as Table 1, but for the Ranked Probability Skill Score over ITALY domain.











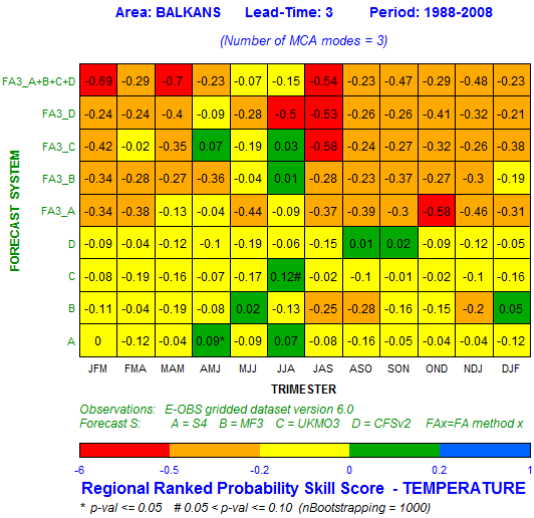
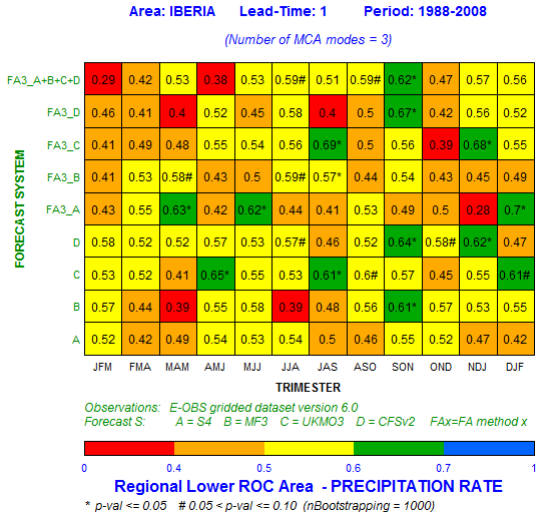
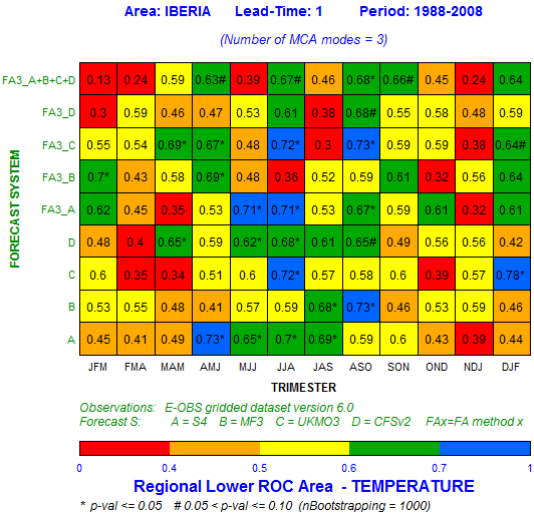
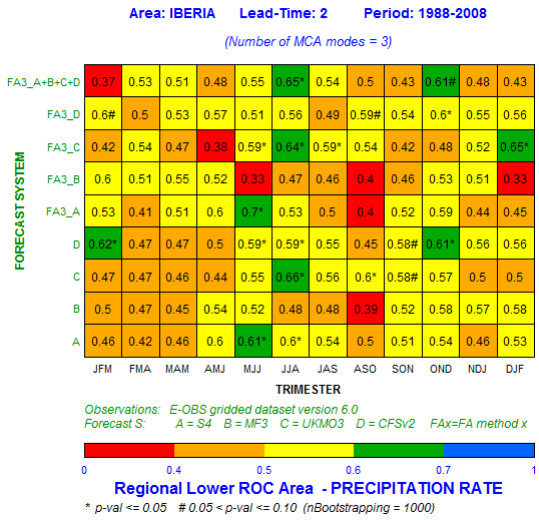
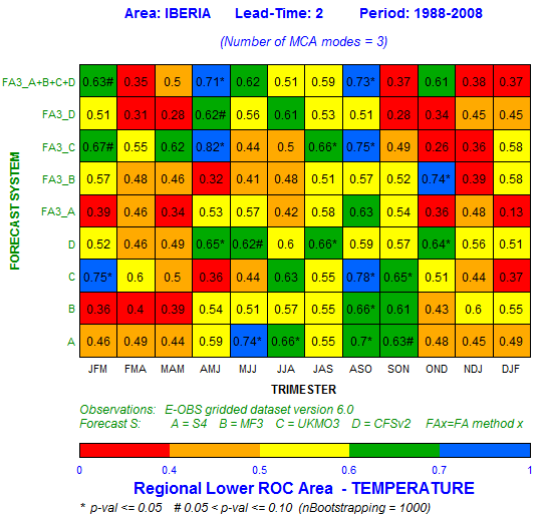


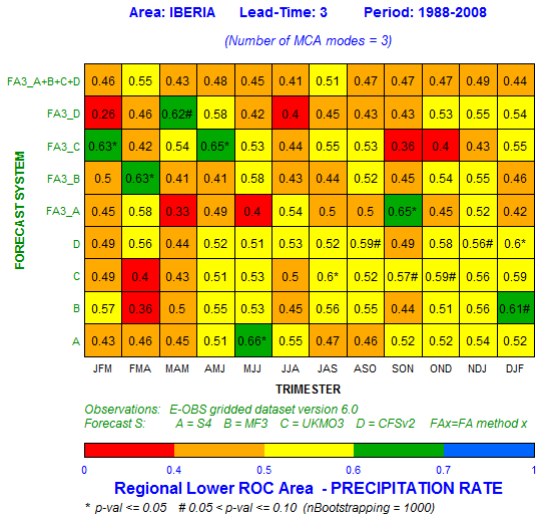
Table 8. The same as Table 1, but for the Ranked Probability Skill Score over BALKANS domain.











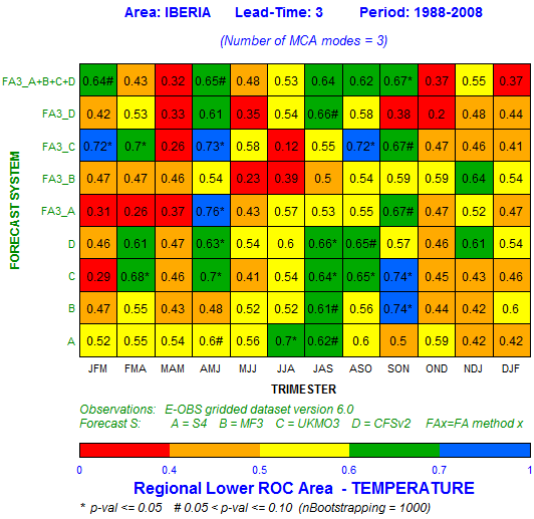
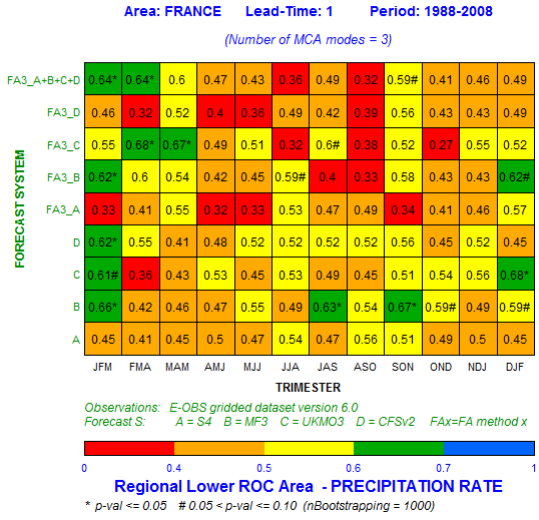
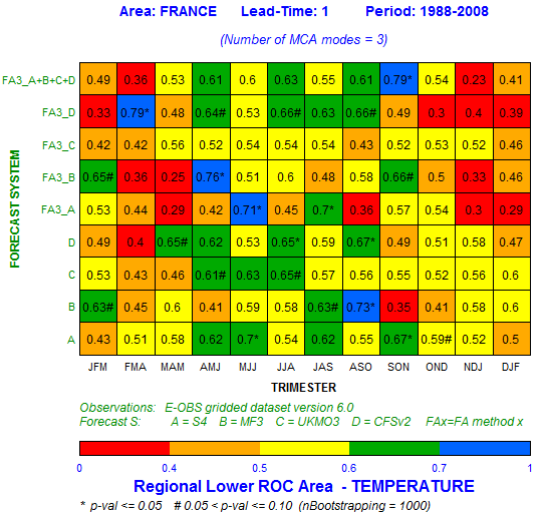
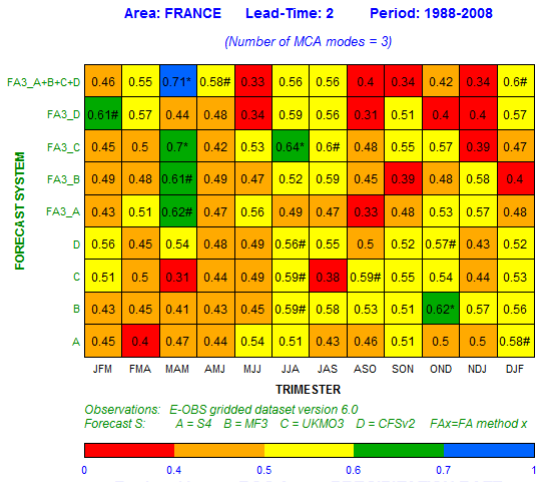
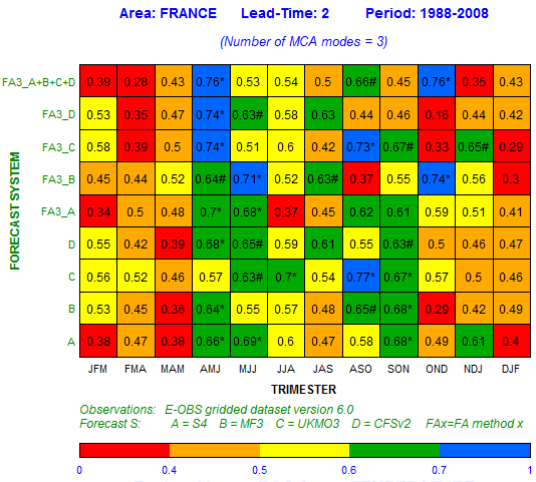


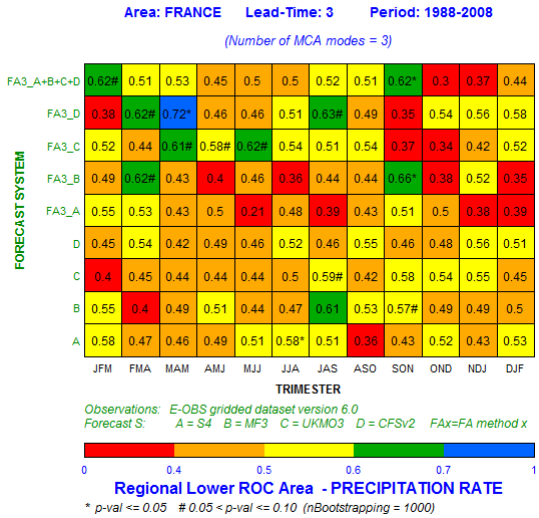
Table 9. The same as Table 1, but for the lower tercile ROC Area.











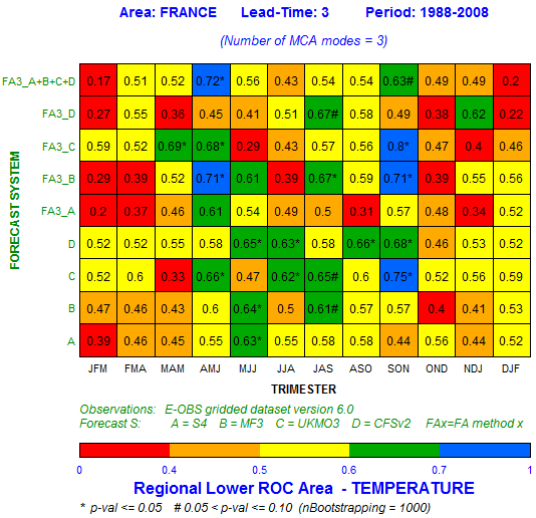
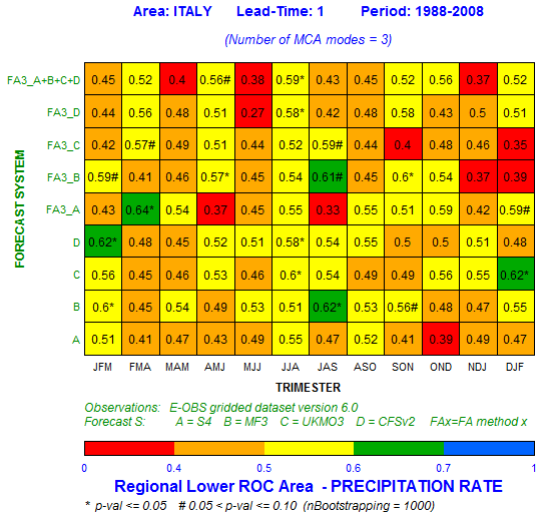
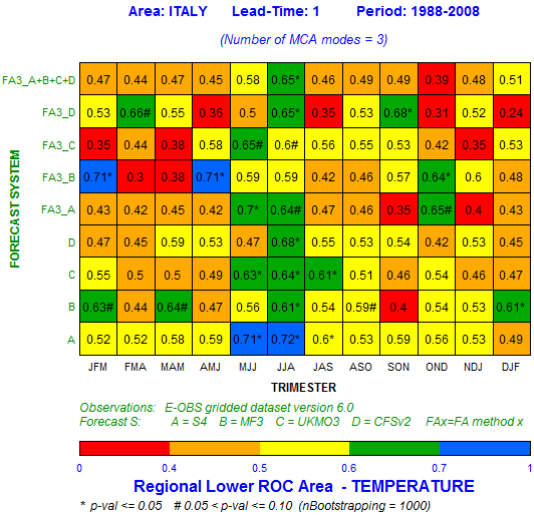
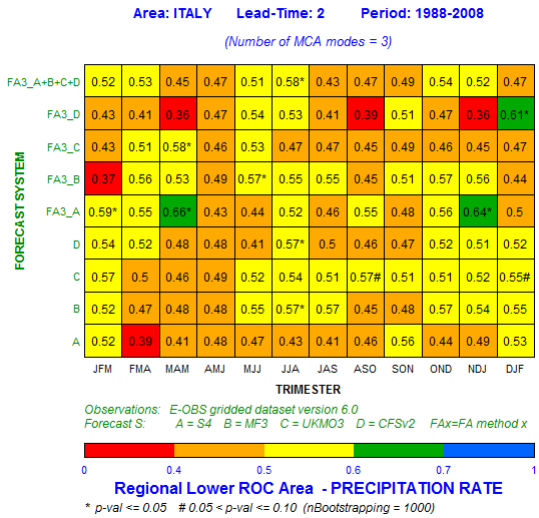
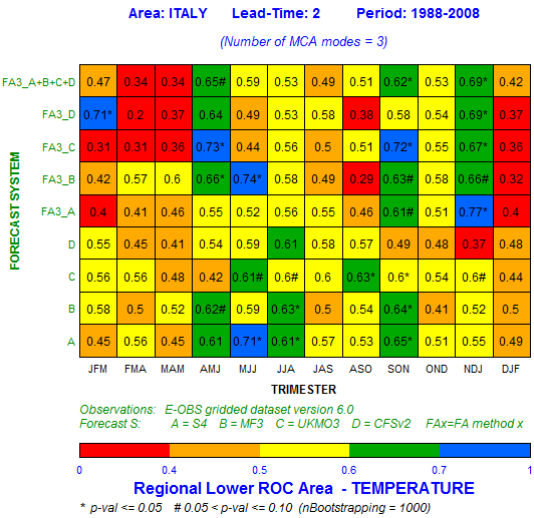


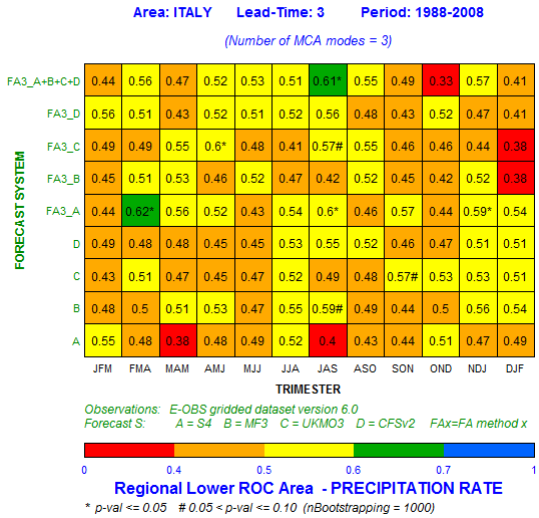
Table 10. The same as Table 1, but for the lower tercile ROC Area over FRANCE domain.











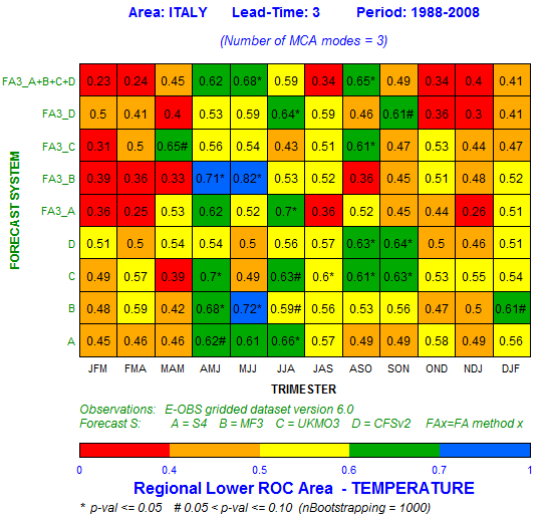
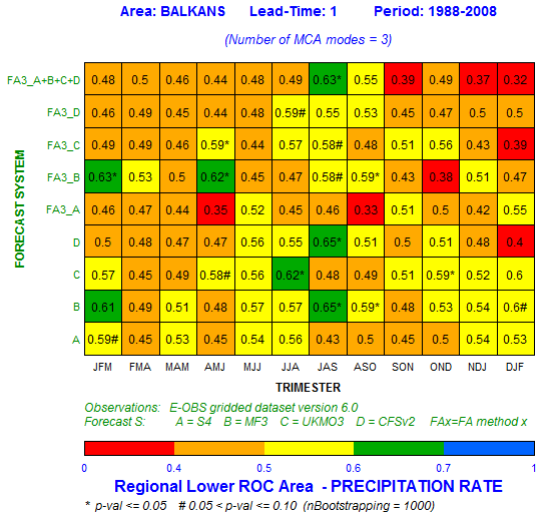
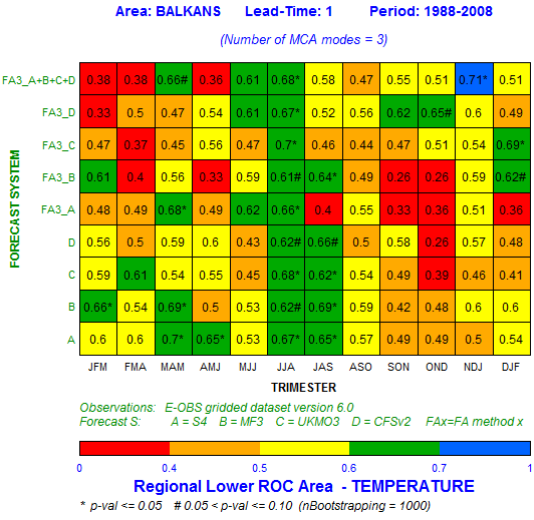
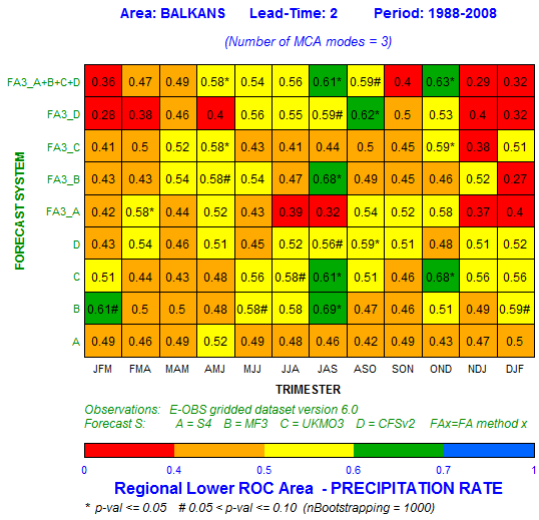
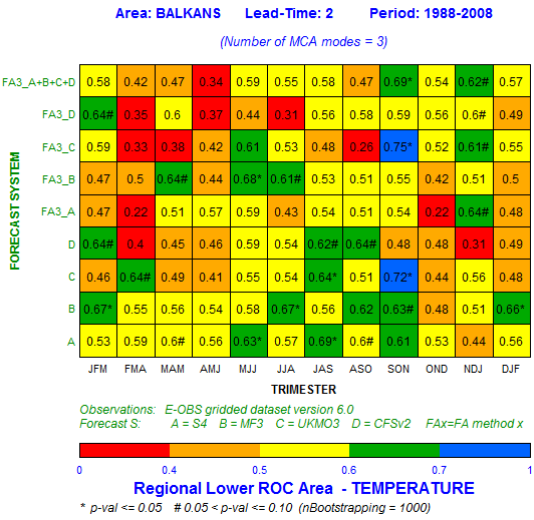


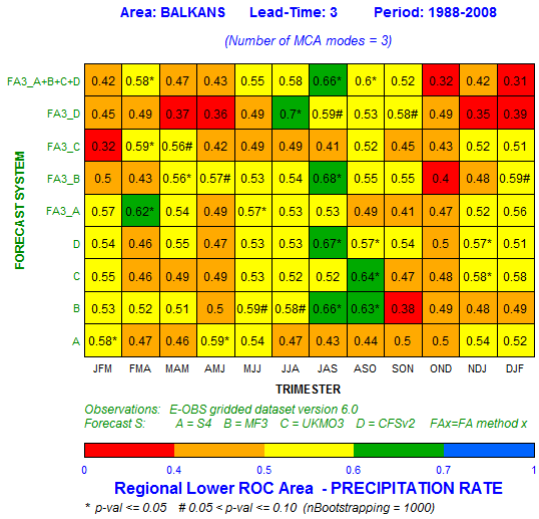
Table 11. The same as Table 1, but for the lower tercile ROC Area over ITALY domain.











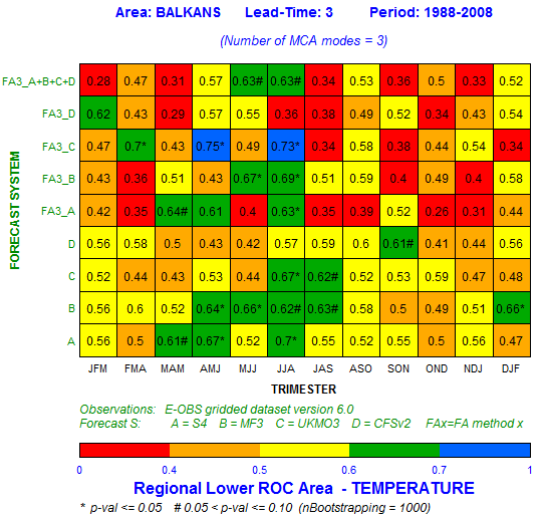
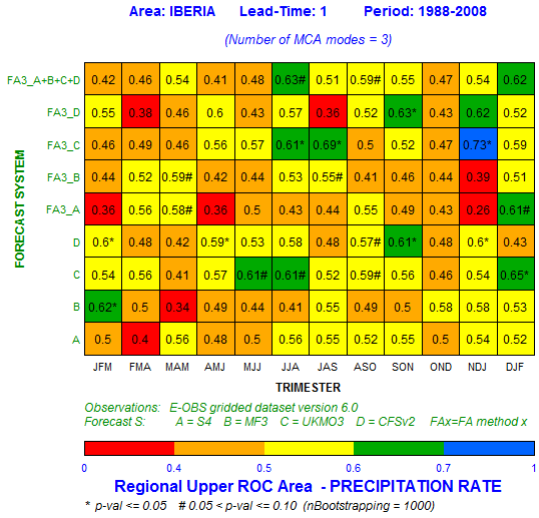
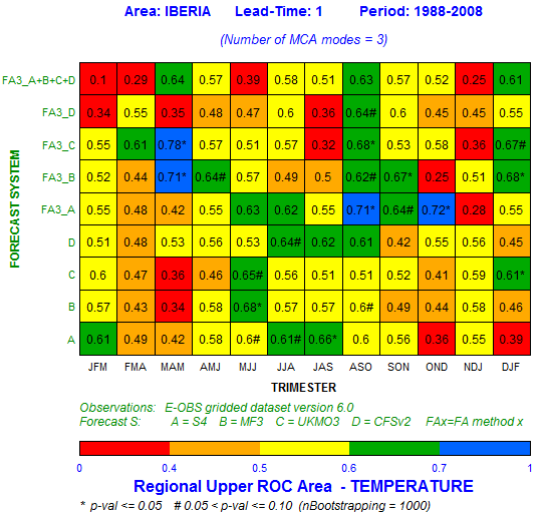
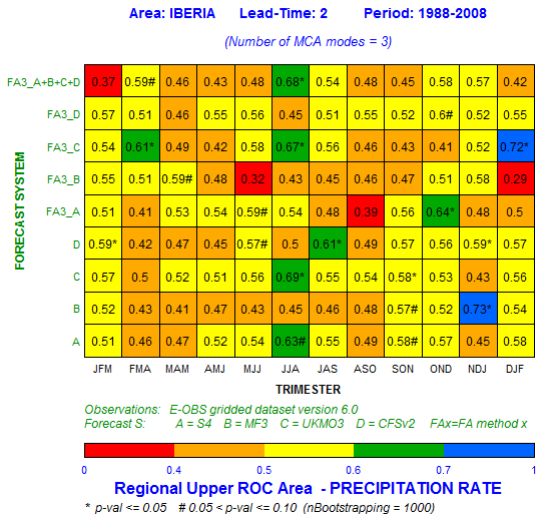
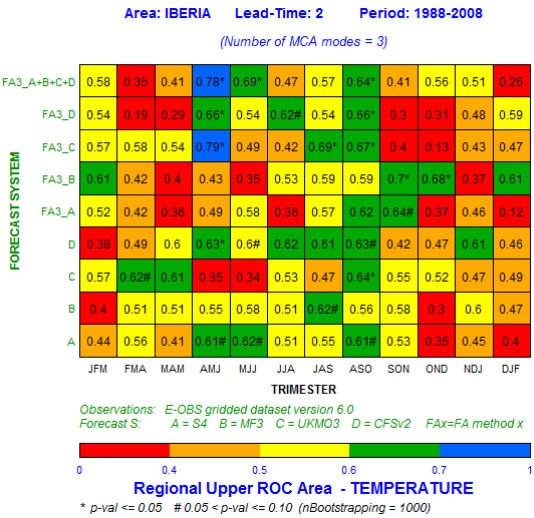


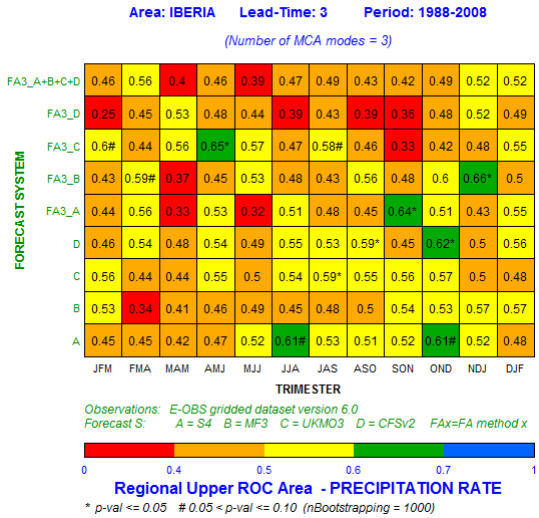
Table 12. The same as Table 1, but for the lower tercile ROC Area over BALKANS domain.











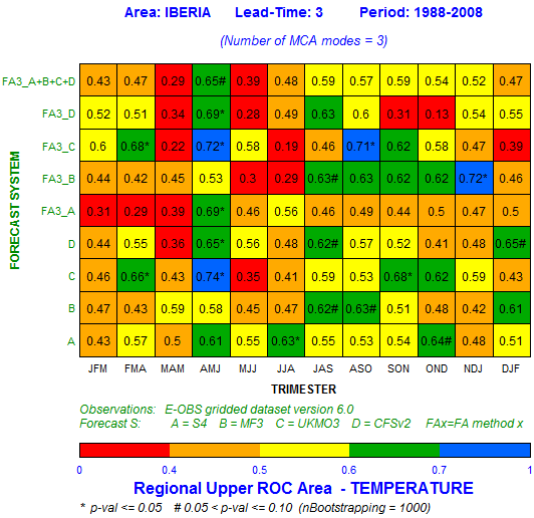
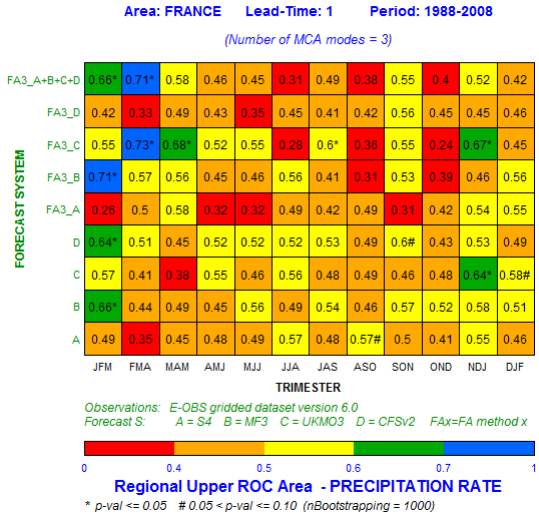
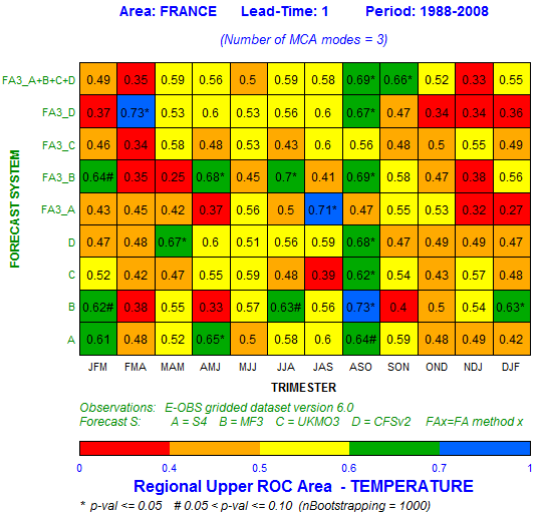
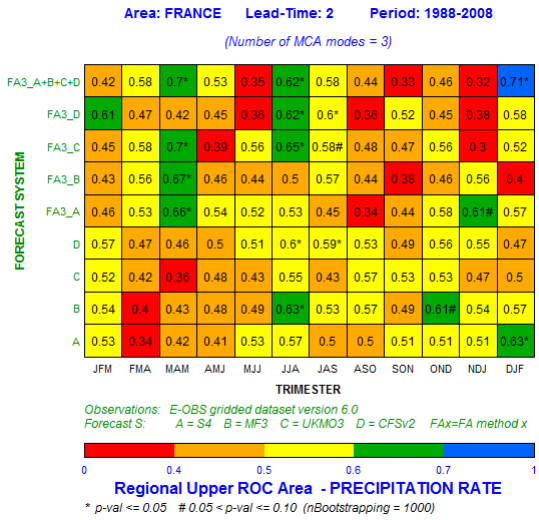
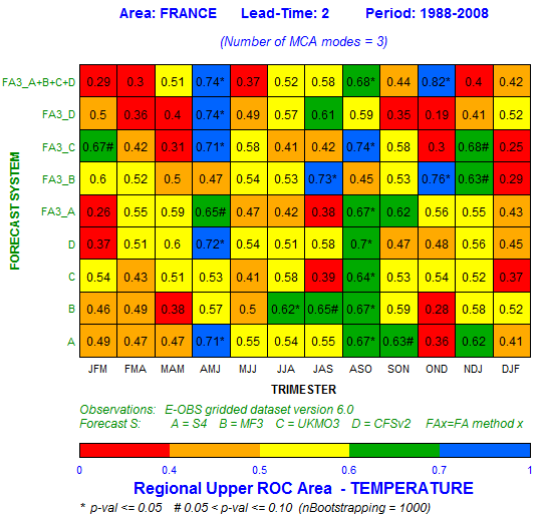


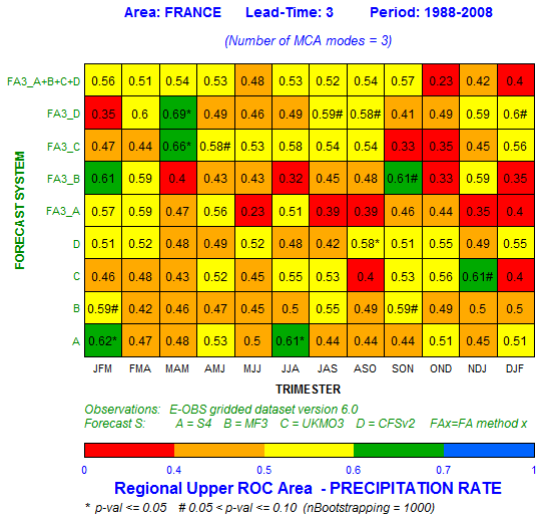
Table 13. The same as Table 1, but for the upper tercile ROC Area.











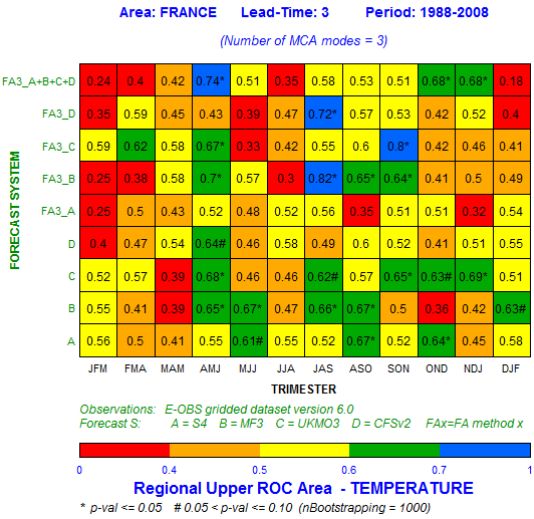
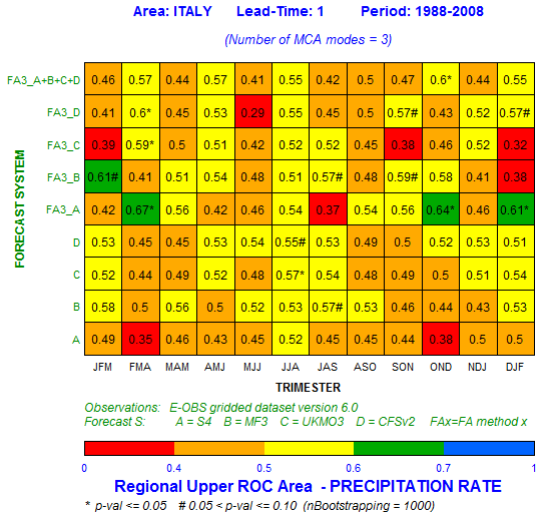
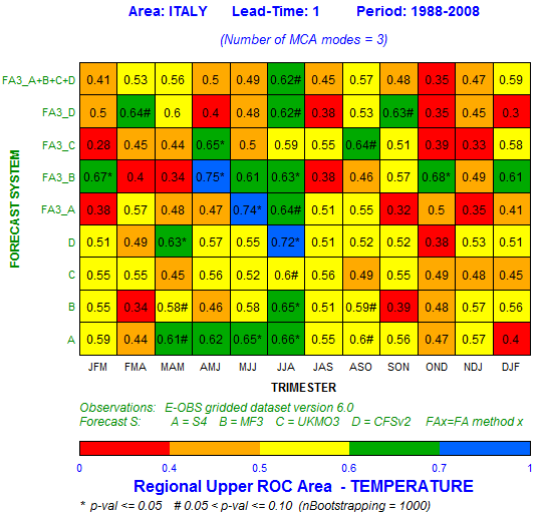
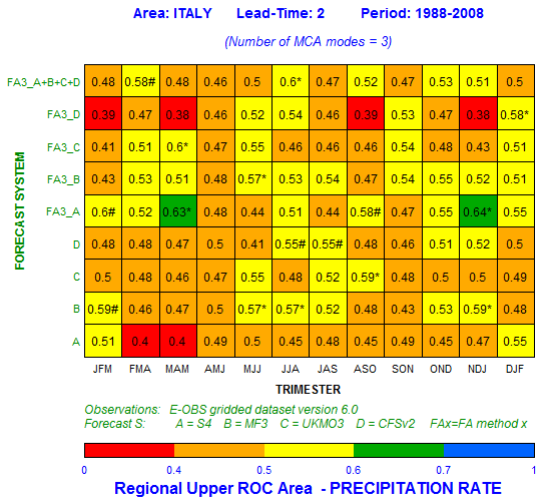
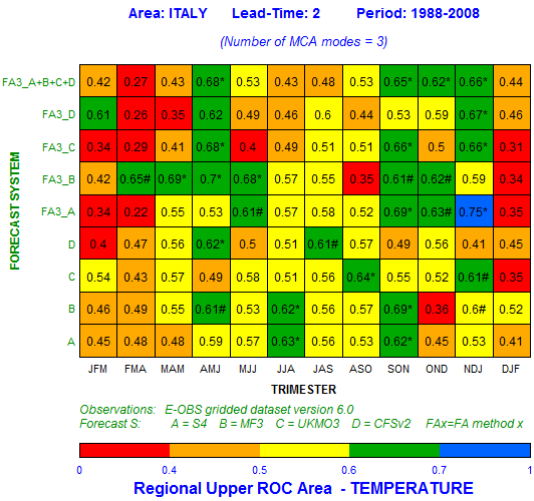


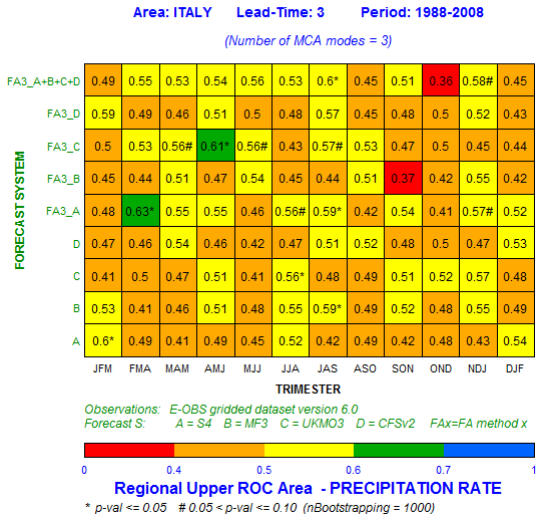
Table 14. The same as Table 1, but for the upper tercile ROC Area over FRANCE domain.











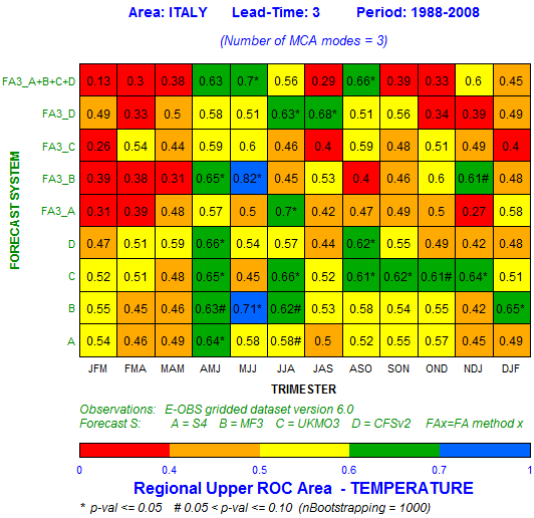
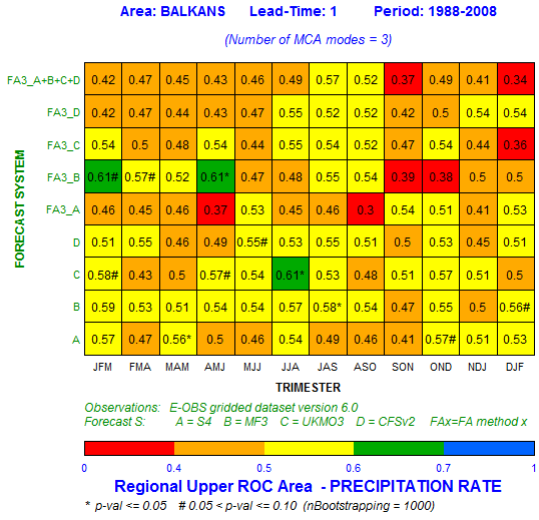
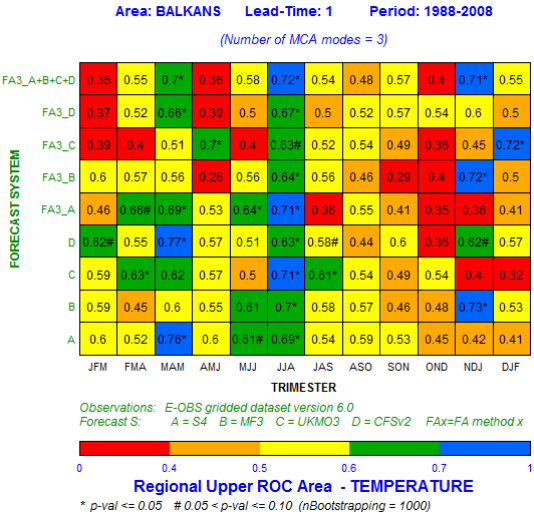
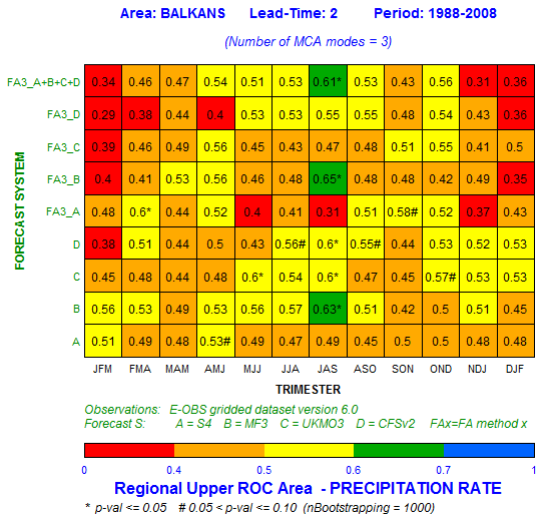
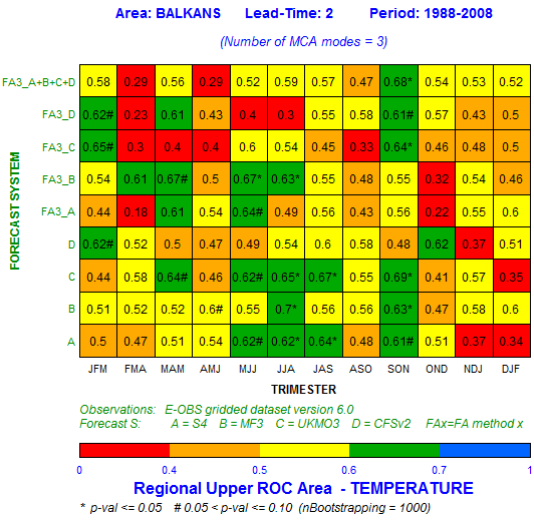


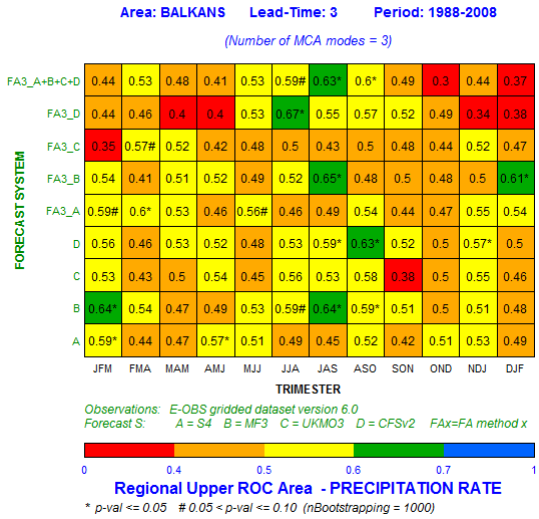
Table 15. The same as Table 1, but for the upper tercile ROC Area over ITALY domain.











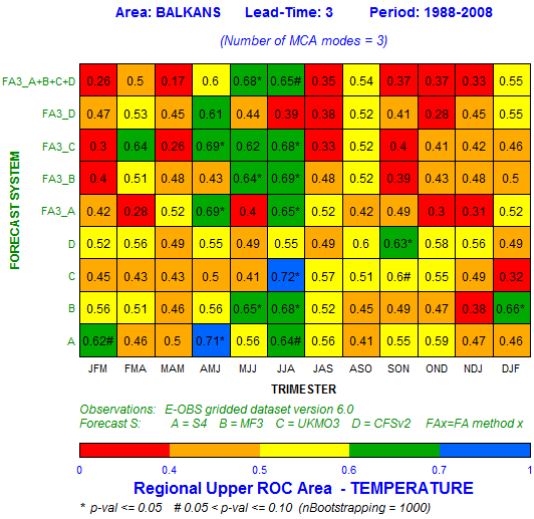
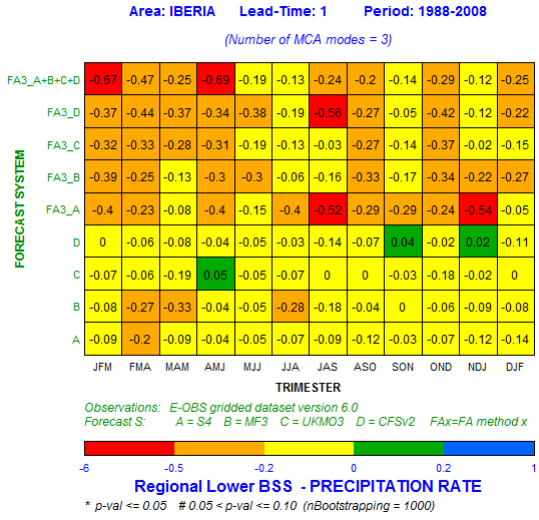
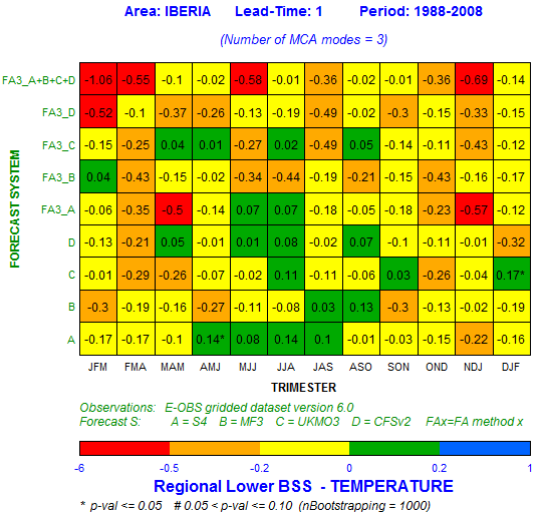
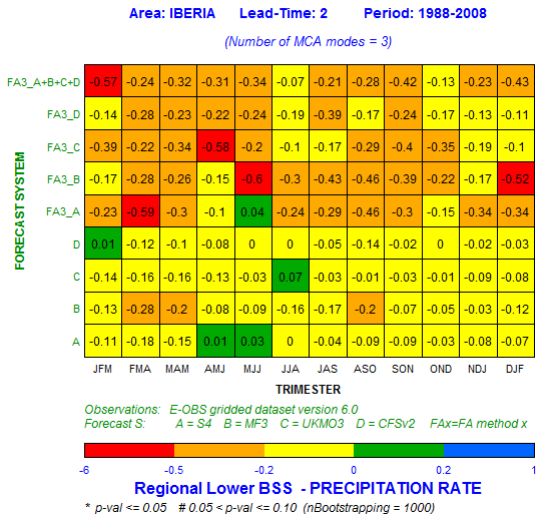
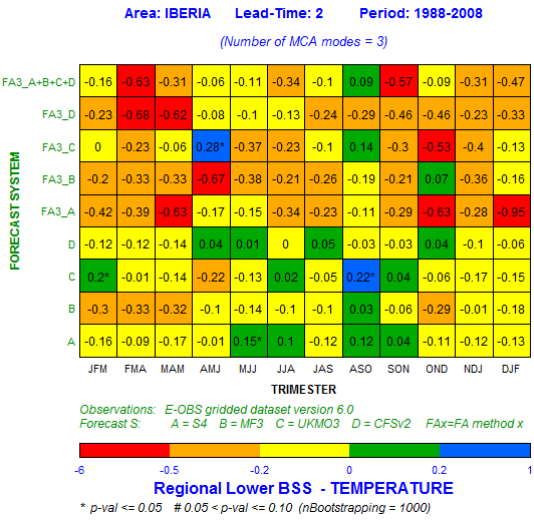


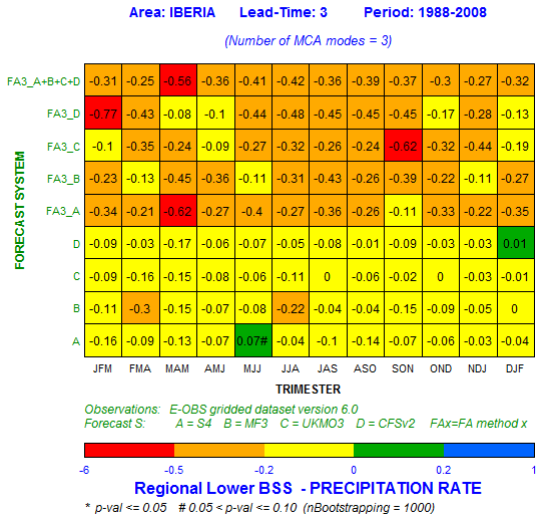
Table 16. The same as Table 1, but for the upper tercile ROC Area over BALKANS domain.











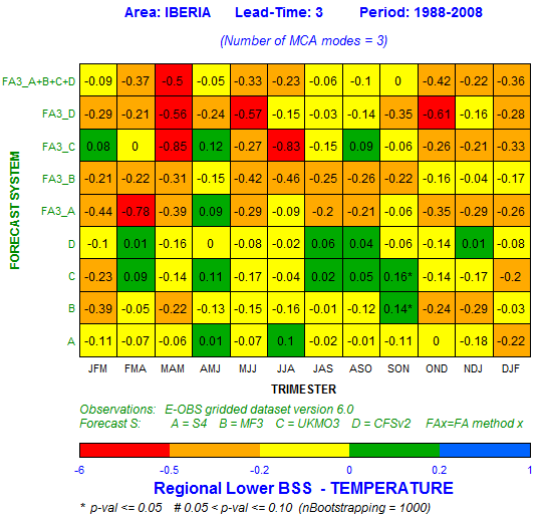
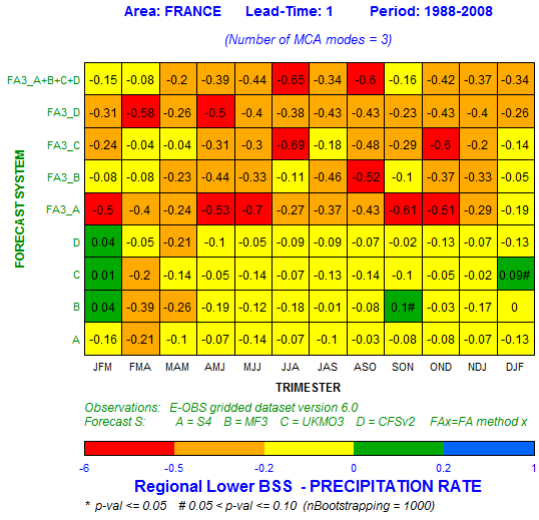
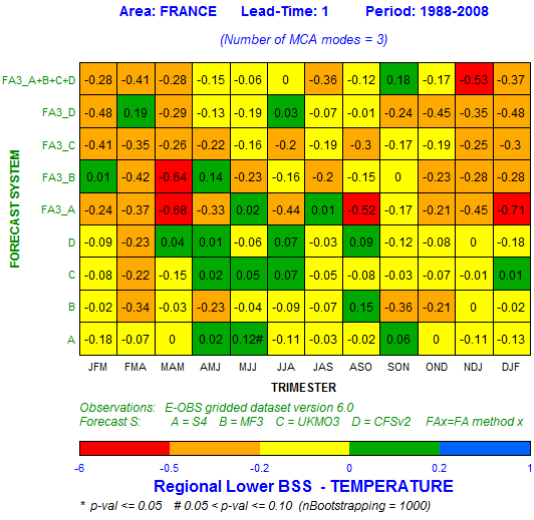
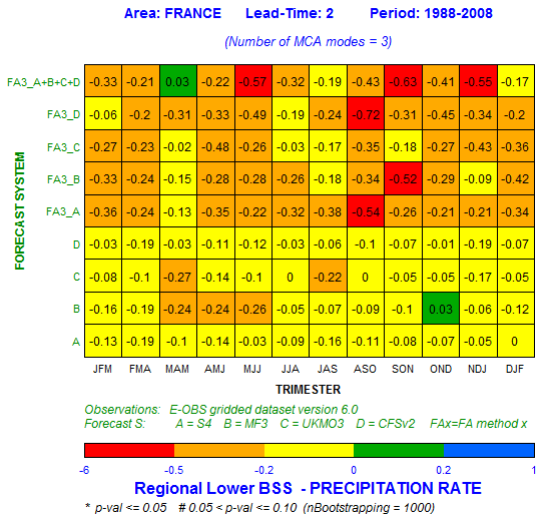
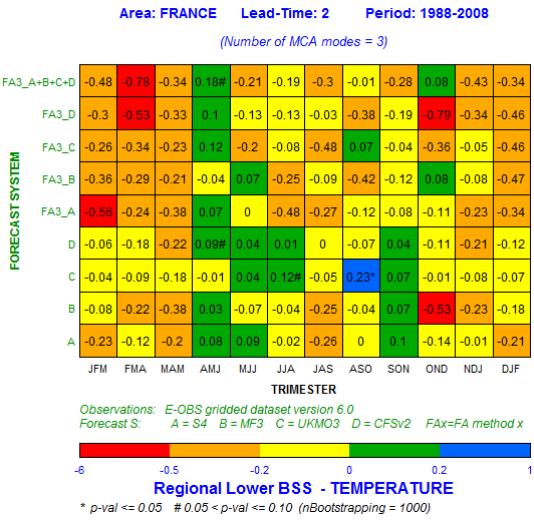


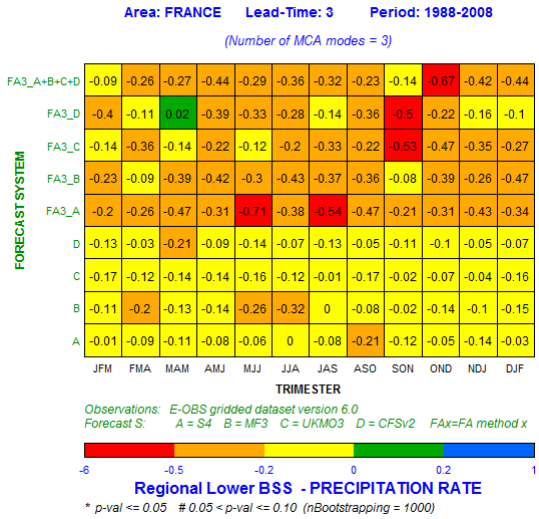
Table 17. The same as Table 1, but for the lower tercile Brier Skill Score.











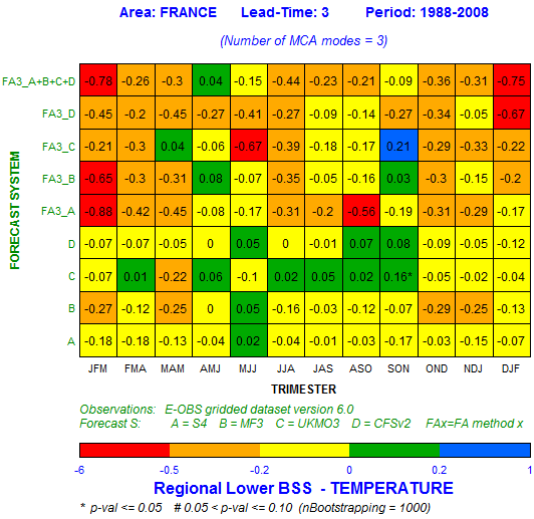
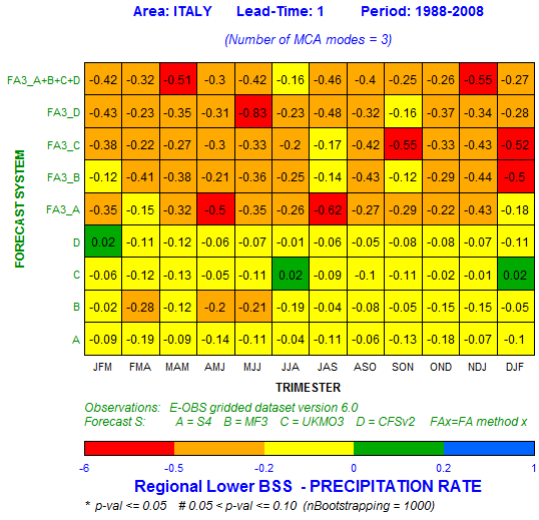
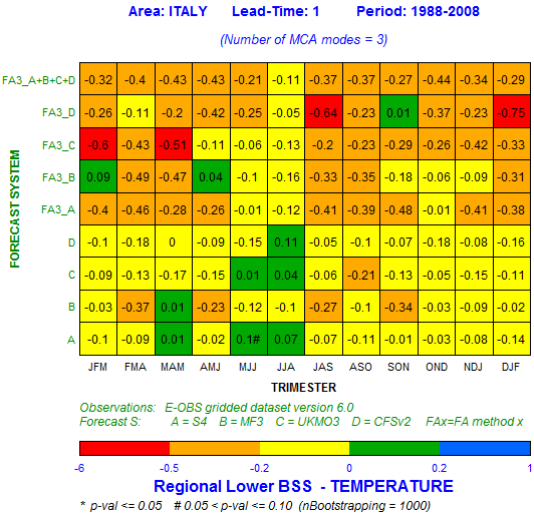
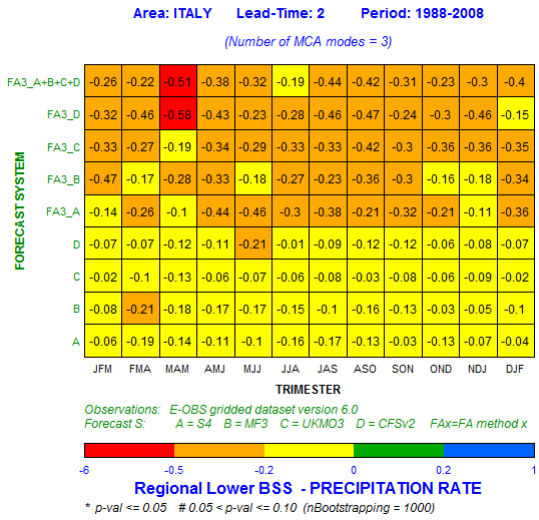
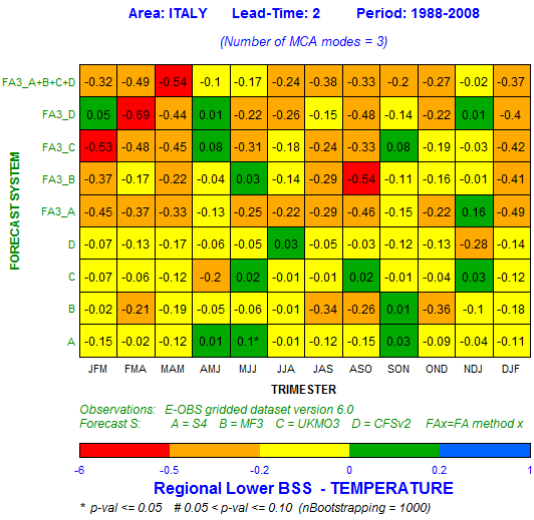


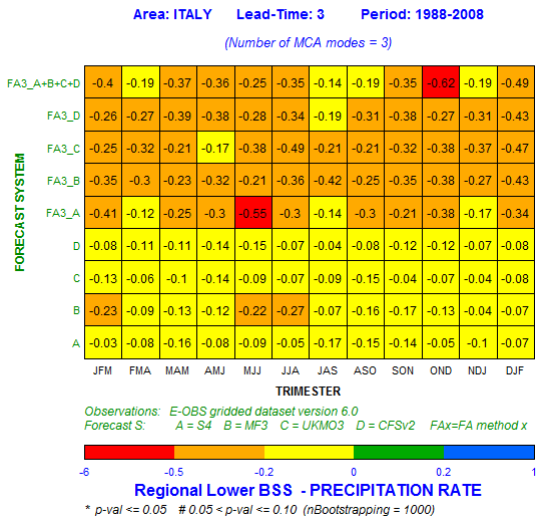
Table 18. The same as Table 1, but for the lower tercile Brier Skill Score over FRANCE domain.











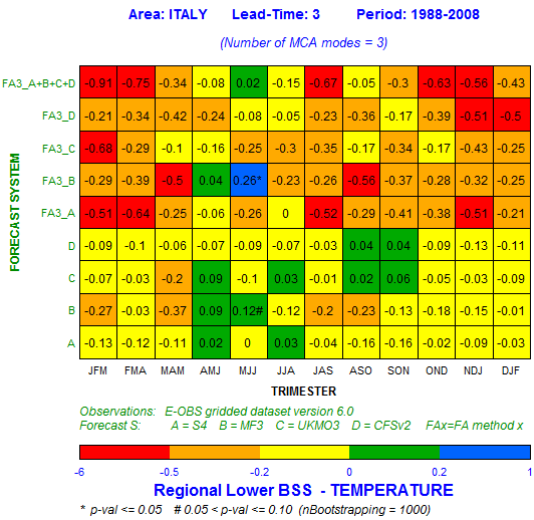
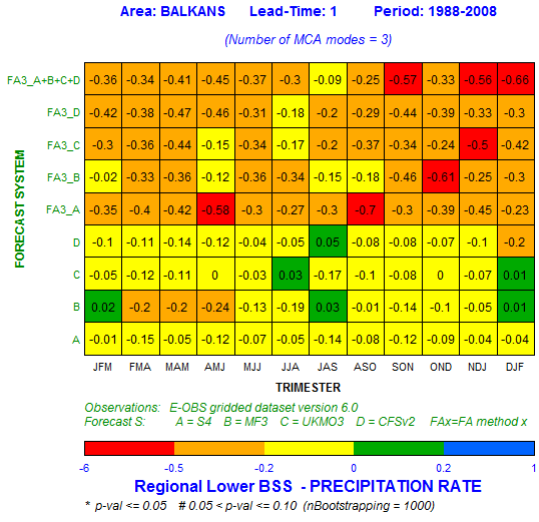
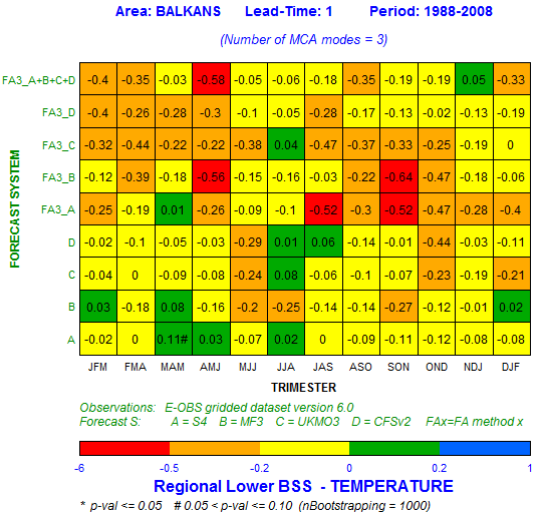
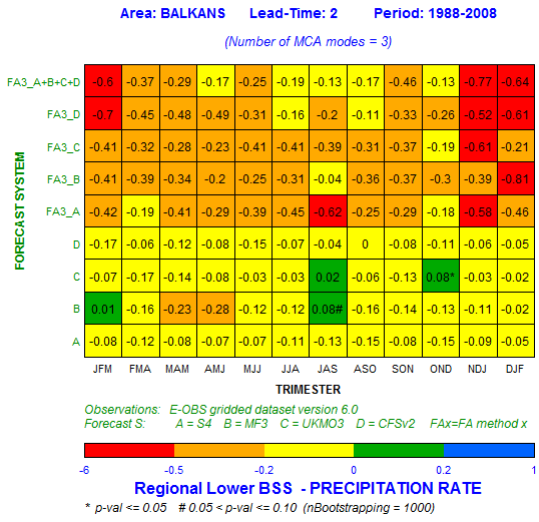
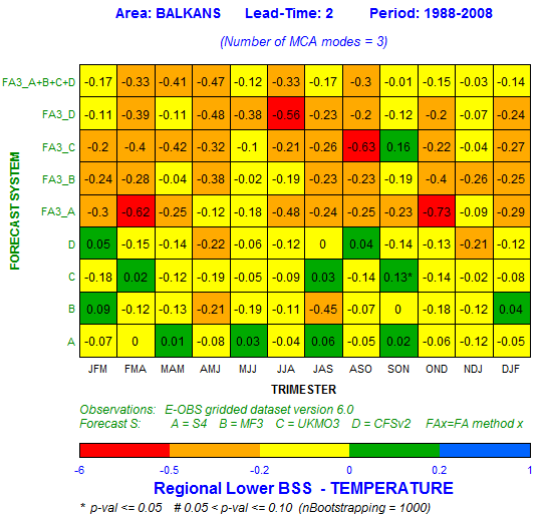


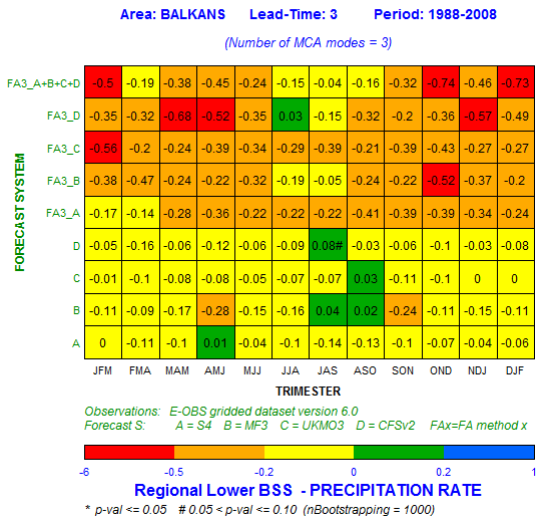
Table 19. The same as Table 1, but for the lower tercile Brier Skill Score over ITALY domain.











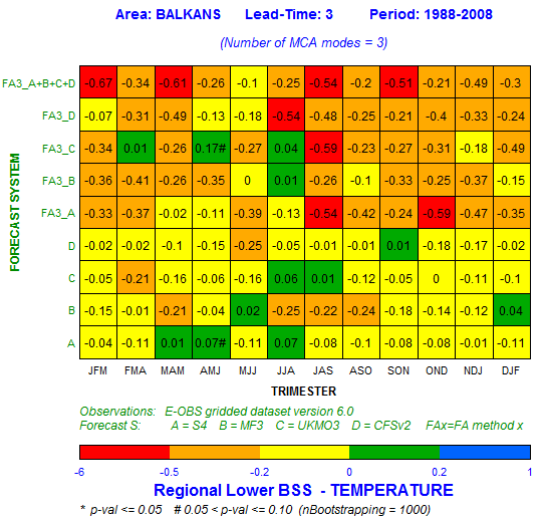
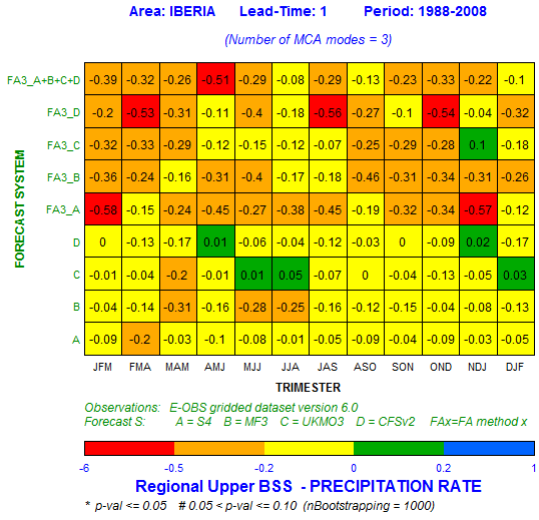
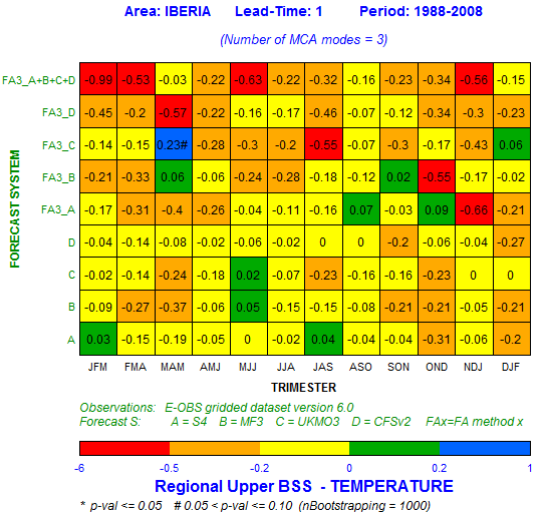
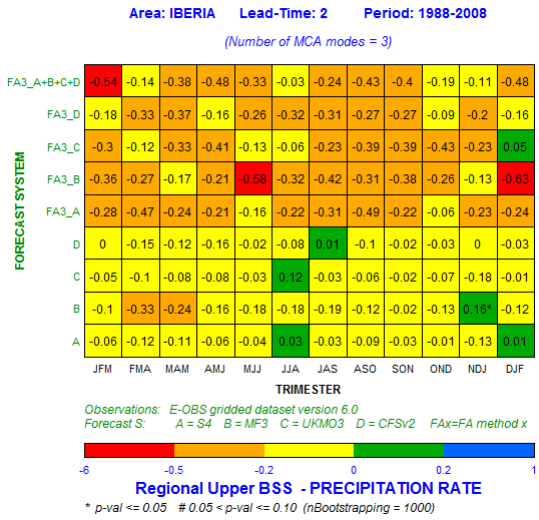
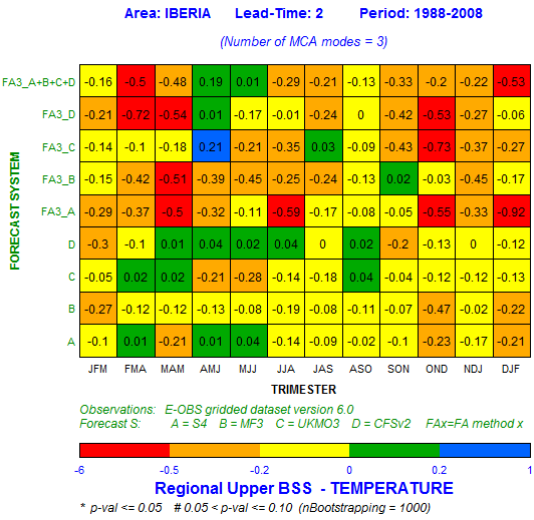


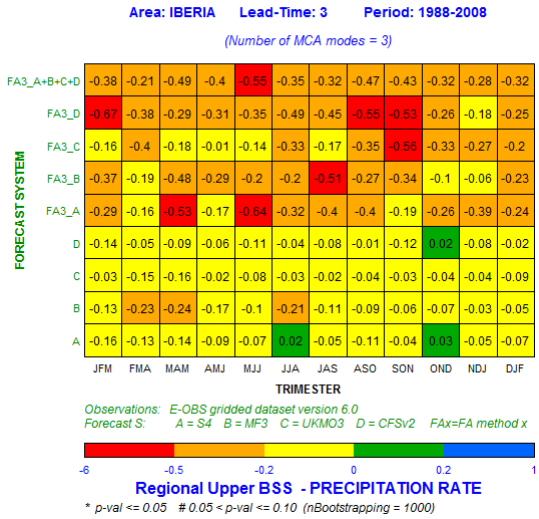
Table 20. The same as Table 1, but for the lower tercile Brier Skill Score over BALKANS domain.











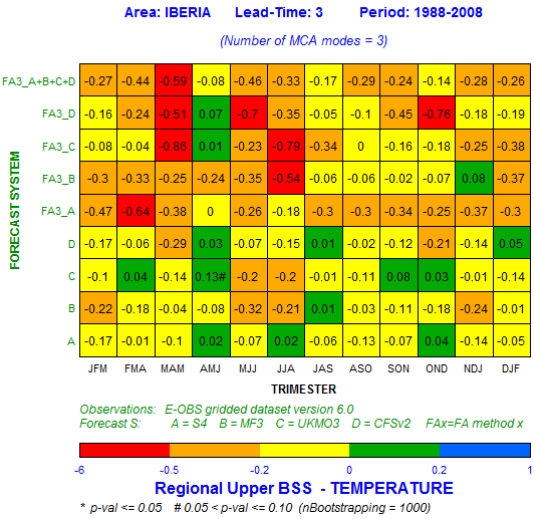
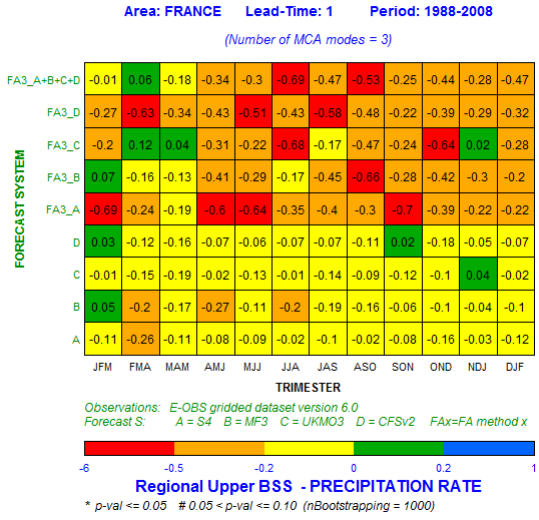
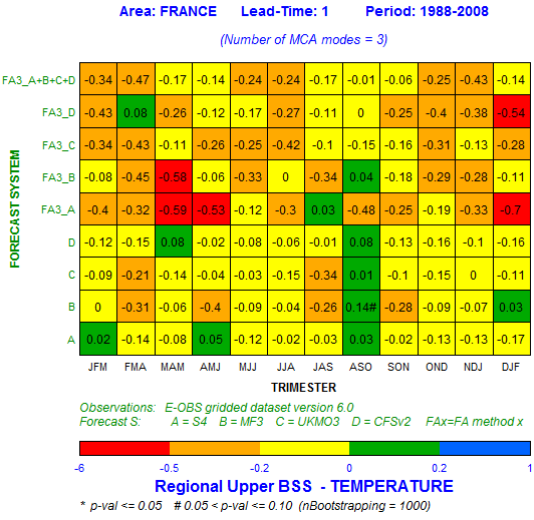
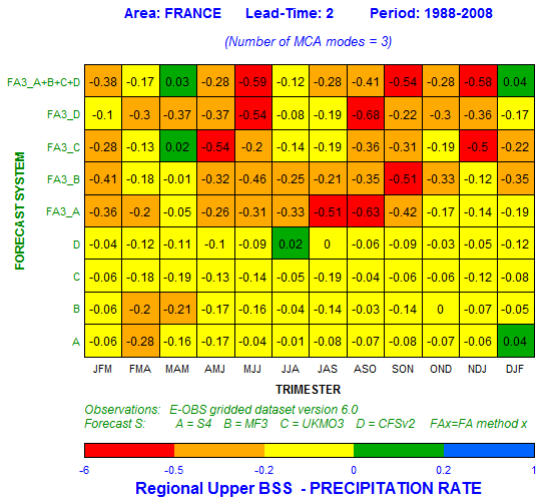
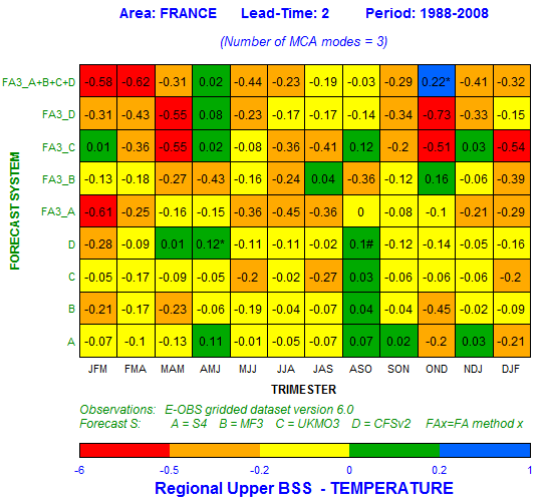


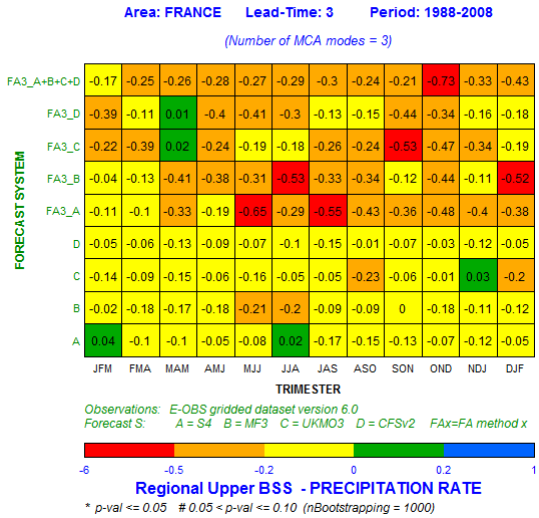
Table 21. The same as Table 1, but for the upper tercile Brier Skill Score.











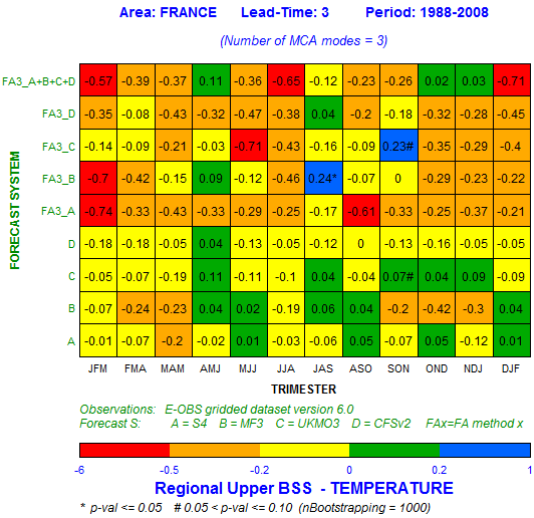
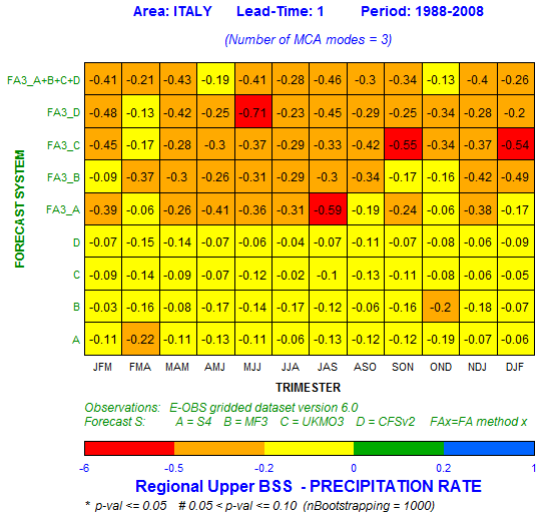
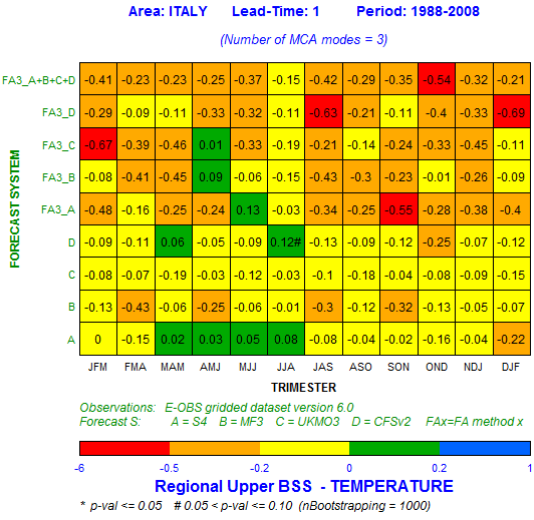
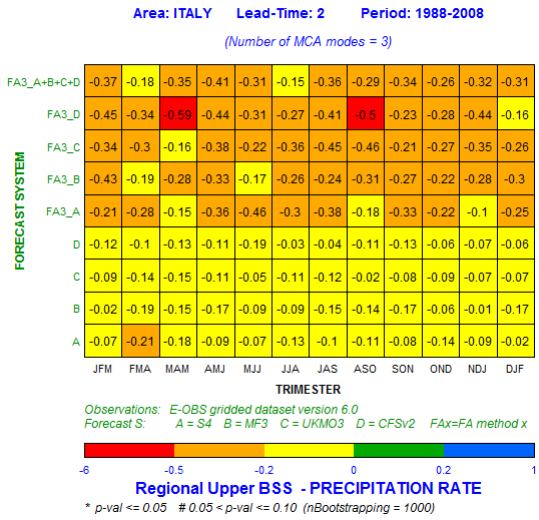
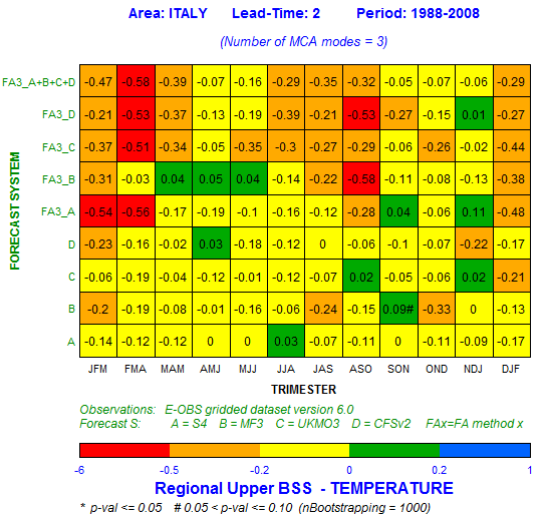


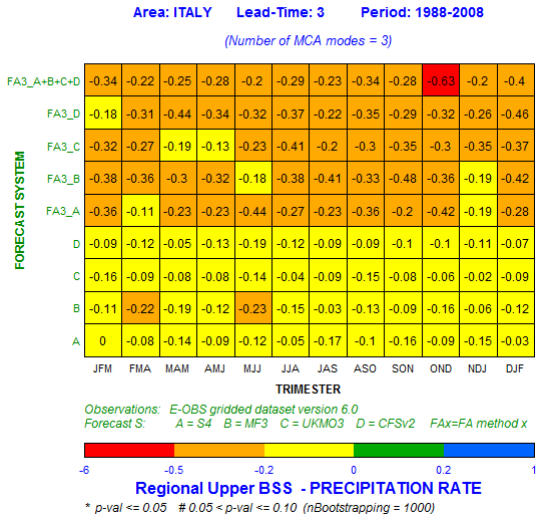
Table 22. The same as Table 1, but for the upper tercile Brier Skill Score over FRANCE domain.











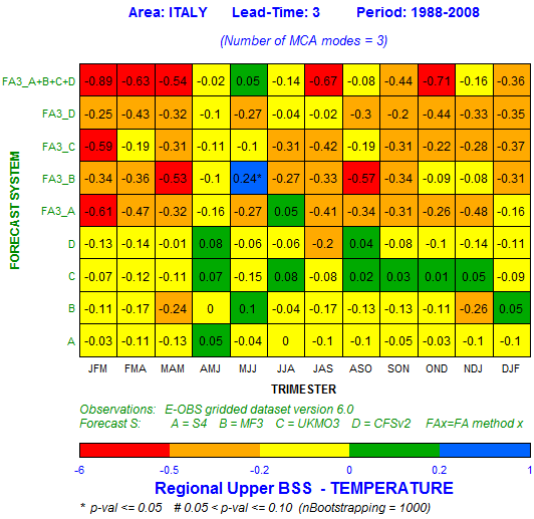
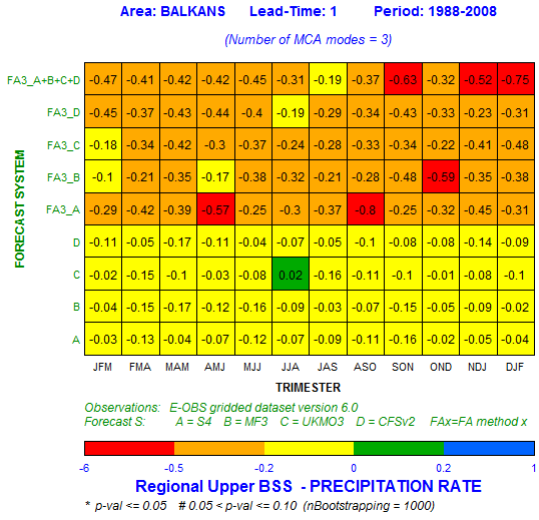
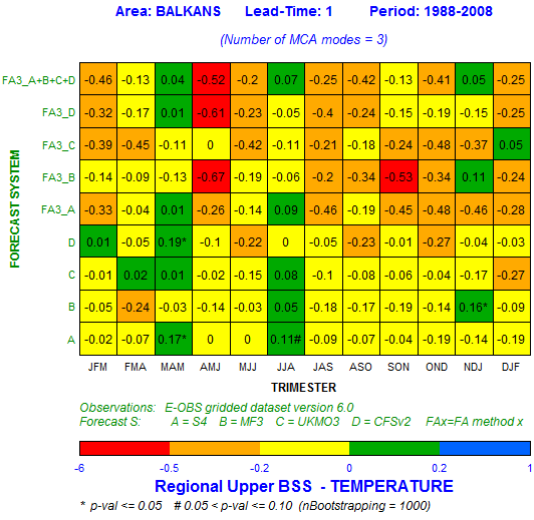
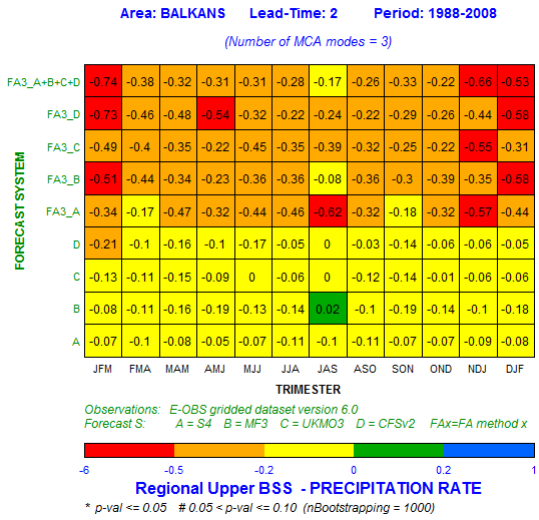
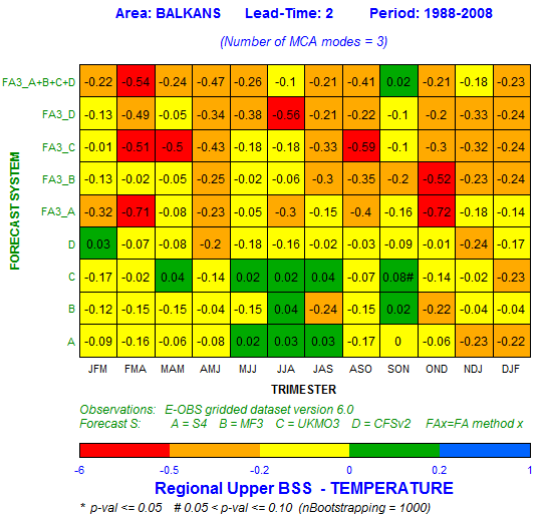


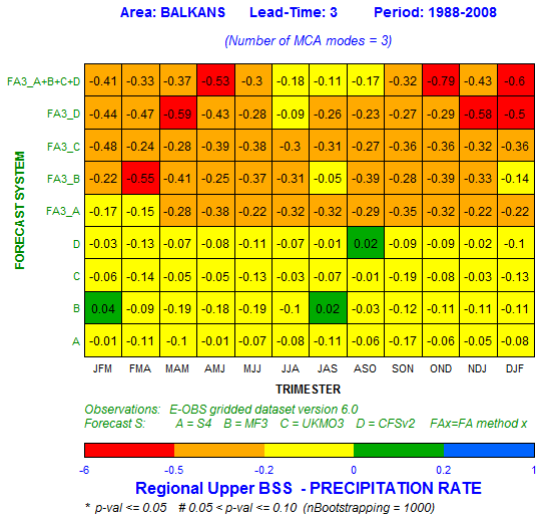
Table 23. The same as Table 1, but for the upper tercile Brier Skill Score over ITALY domain.











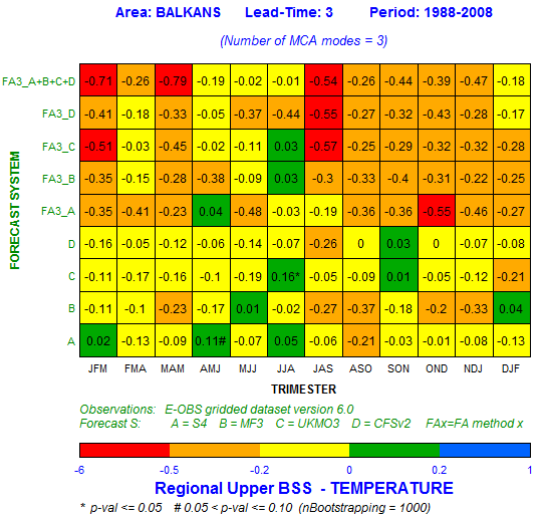


Table 24. The same as Table 1, but for the upper tercile Brier Skill Score over BALKANS domain.

6. Conclusions

In general, verification results show, as expected, low skill at seasonal timescale consequence of the low predictability over mid-latitude regions. However, we can still draw the following conclusions:

- Better skill in general for temperature than for precipitation, as expected over mid-latitude regions.
- Relative consistency among models which allows the identification of some windows of opportunity for seasonal forecasts associated to certain seasons, variables and in some cases limited to certain models. Summer appears as a window of opportunity for temperature, possibly linked to the general trend associated to the warming of the climate system.
- The window of opportunity for temperature is centered around summer and early autumn over the Iberian Peninsula and France domains, whereas it shifts towards spring (MAM and JJA) over the most easterly domains (Italy and Balkans)
- Some scores show more skill than others, e.g., the ROC area skill score -providing an indication of a forecasts system discrimination capacity- tends to have more skill (relative to climatology) than other scores which explore other aspects of forecasts.
- Certain features related with the different quality of models for different seasons are also detected by the verification scores. For example, UKMO3 and MF3 models tend to show more skill in winter time than the rest of the models, whereas S4 shows the highest skill when averaged over all seasons.
- Existence of some barriers or peaks of skill over certain domains, suggesting initial conditions with different predictability. These barriers (peaks) move to the right in the verification tables as lead time increases clearly indicating initial conditions with less (more) predictability. Consequently, higher lead times do not automatically are associated with degradation in terms of skill.
- Benefits from the application of the calibration and combination FA algorithm depend highly on each specific model. Some models seem to have different potential of improvement when the FA algorithm is applied. This may lead in the future to consider different strategies for calibrating and combining each model and to reconsider the application of the same setup (FA3 was selected for this study) for all models.

7. References

- 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.pd)

- 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 Interanual 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

ANNEX I

Description of different setups to apply the FA method

Different setups to apply the Bayesian method FA have been explored in order to improve the seasonal forecast quality:

1. FA1 Configuration:

   - Anomalies of averaged three month values are used for applying the maximum covariance analysis (MCA).
   - The prior function is computing over the common period the hindcasts are available: 1988-2008.

2. FA2 Configuration:

   - Averaged three month standardized values are used for applying the maximum covariance analysis (MCA).
   - The prior function is computing over the common period the hindcasts are available: 1988-2008.

3. FA3 Configuration:

   - Averaged three month standardized values are used for applying the maximum covariance analysis (MCA).
   - The prior function is computing over the extended period: 1960-2010.

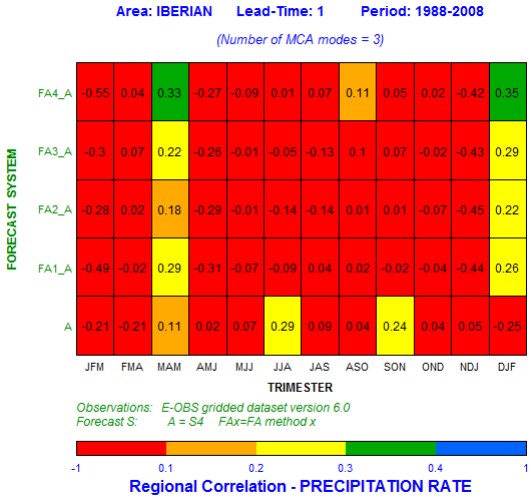
4. FA4 Configuration:

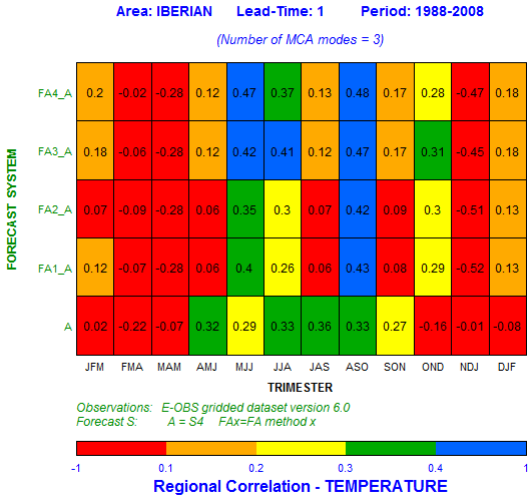
   - Anomalies of averaged three month values are used for applying the maximum covariance analysis (MCA).
   - The prior function is computing over the extended period: 1960-2010.

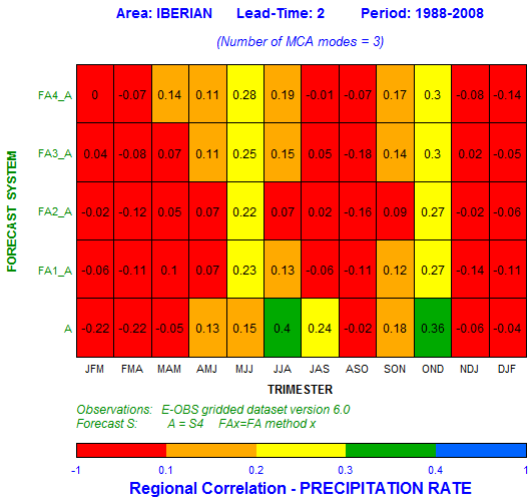
The anomalies and standardized values of the different prediction systems, computed as the difference between the forecasted and climatological values for each system, are obtained by cross-validated forecasts on data not used in the estimation, i.e., the year to be forecast is removed from the data set. Cross-validation method is also applied to compute the anomalies of observation data and for the four calibration/combination FA methods described above.

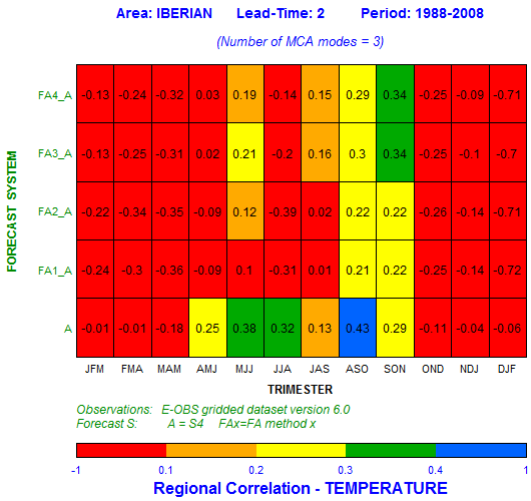
Assessment of the four considered configurations to apply the FA method.

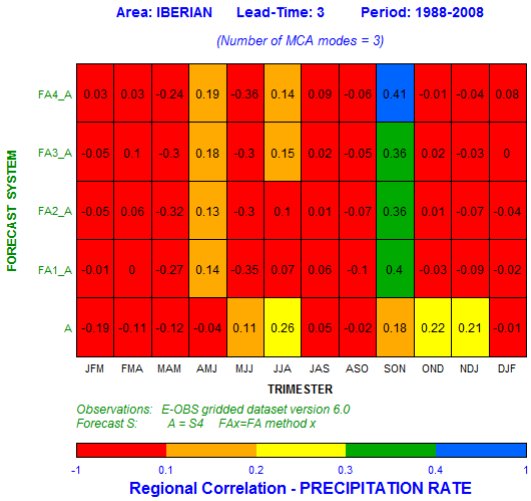
The correlation between the predicted and the observed mean value of anomalies over the Iberian domain for FA1 to FA4 configurations, for temperature and precipitation, for the 12 different three-month periods, for lead times 1, 2 and 3 has been computed in order to select which method is the best to improve the quality of the forecast systems. Analysing the results (see Table 25 to 28) the FA3 setup has been selected to applied to the full study, although the FA4 configuration shows very similar values.











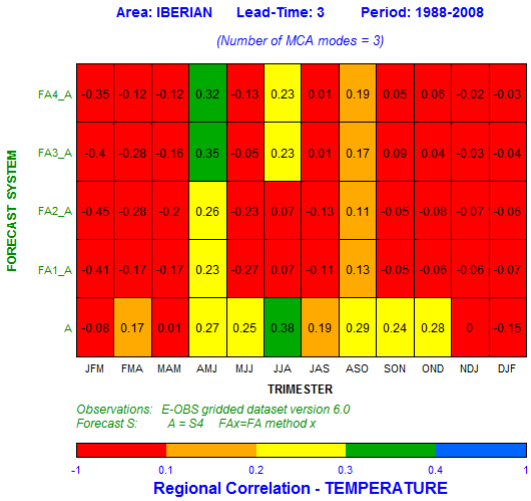
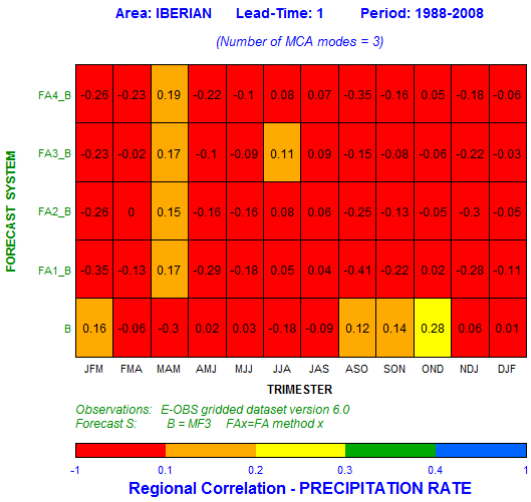
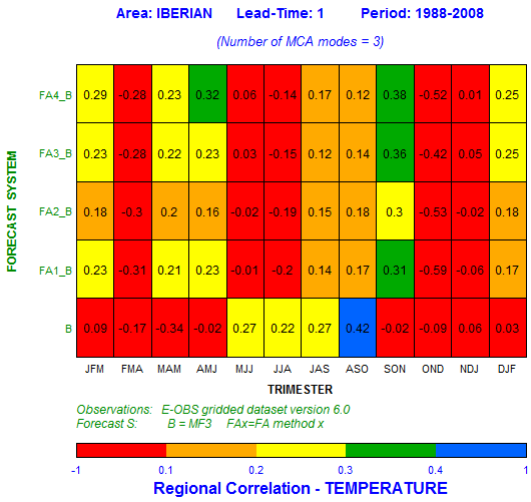
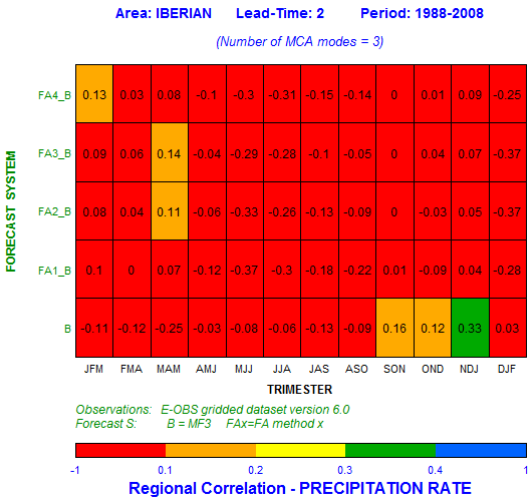
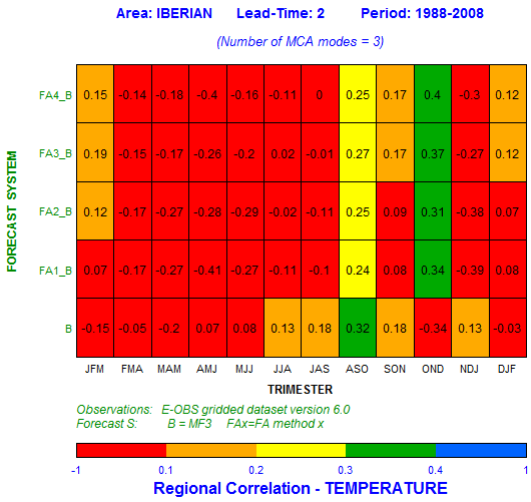


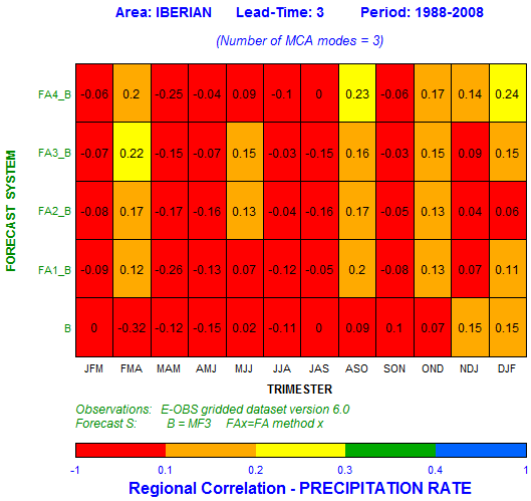
Table 25. Regional correlation coefficients between observations and forecasts (four different FA setups (FA1, FA2, FA3 and FA4 applied only to the S4 system) computed for temperature and precipitation anomalies for 12 different three-month periods and for lead-times 1, 2 and 3 over the Iberian domain.











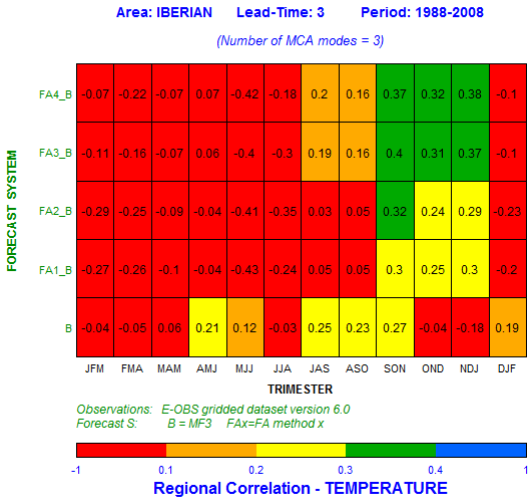
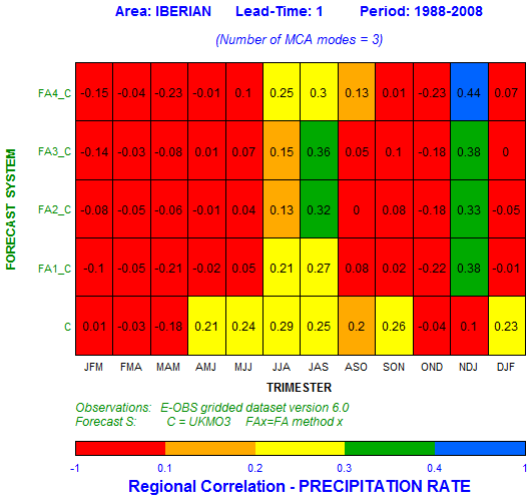
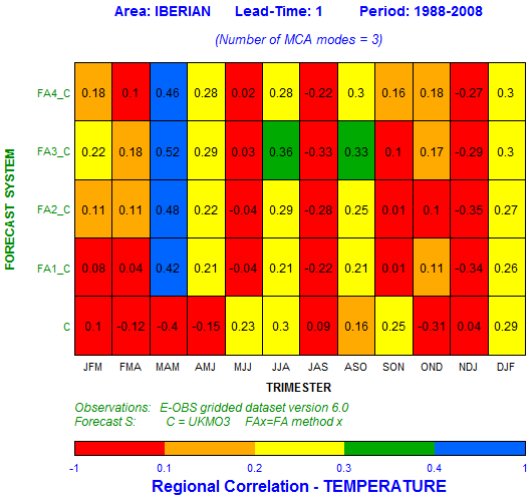
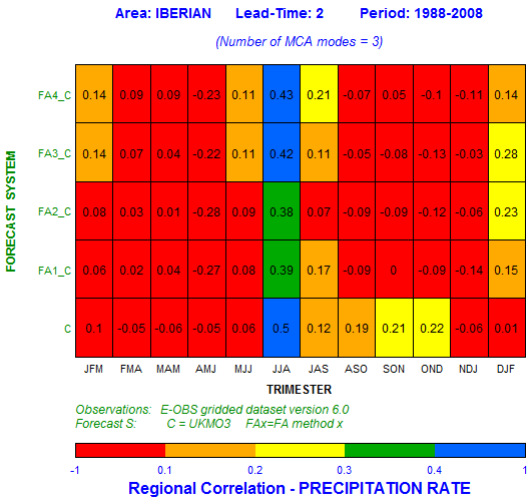
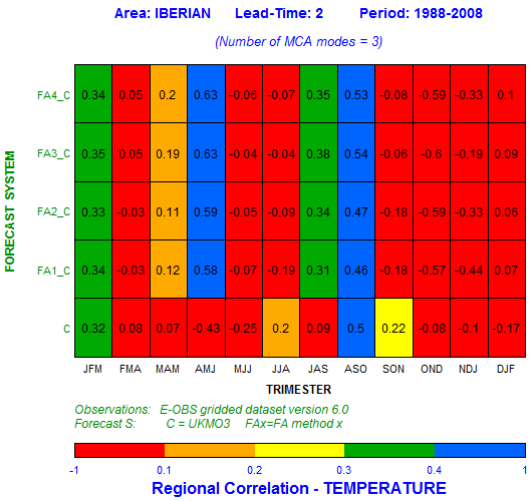


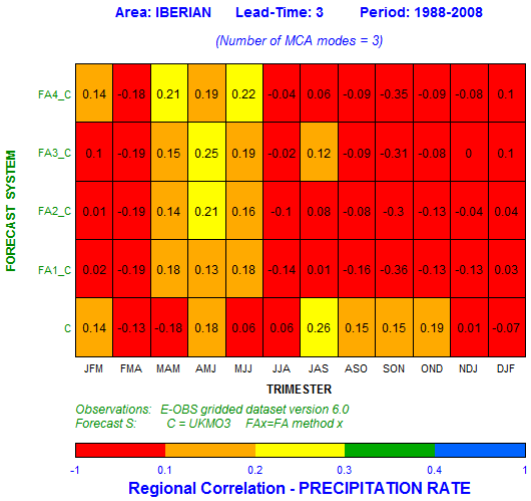
Tabla 26. The same as Table 25, but applying the FA method to the MF3 forecast system.











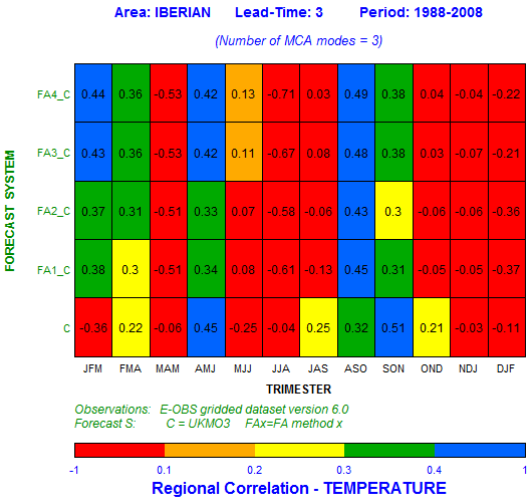
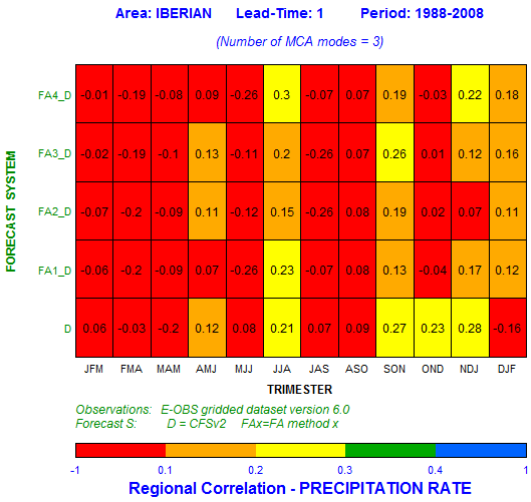
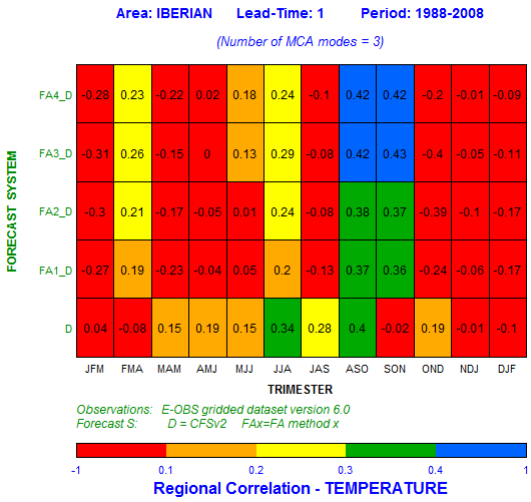
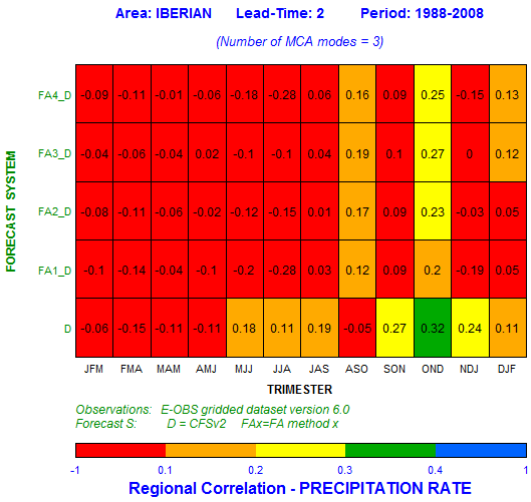
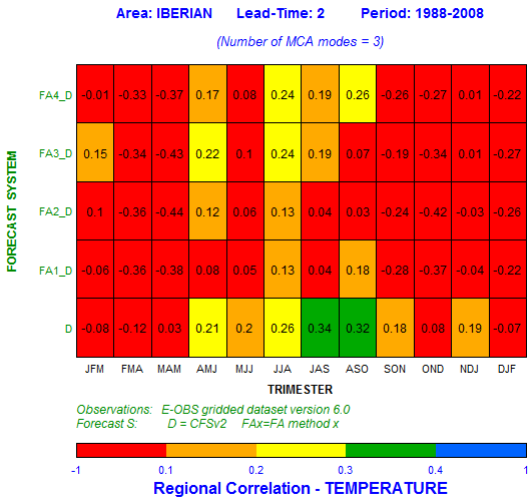


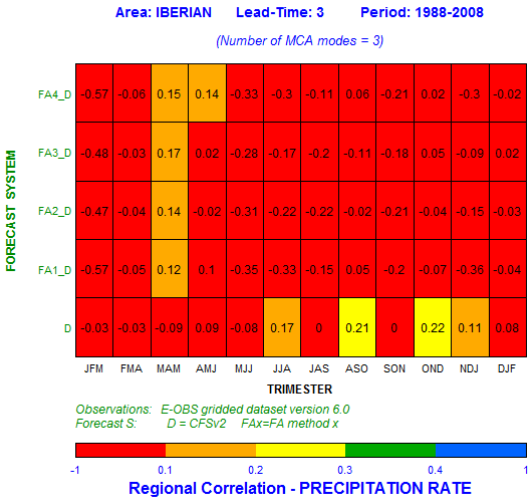
Tabla 27. The same as Table 25, but applying the FA method to the UKMO3 forecast system.











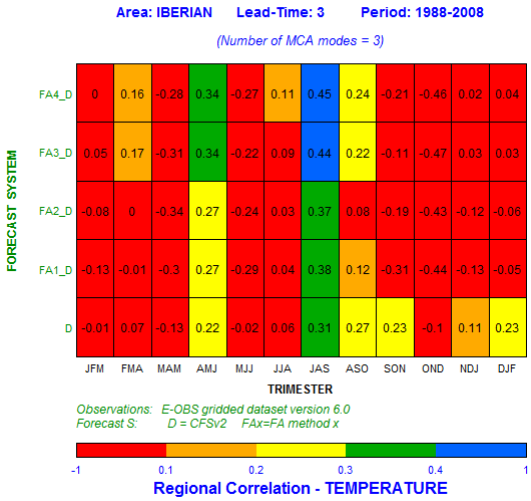


Tabla 28. The same as Table 25, but applying the FA method to the CFSv2 forecast system.

ANNEX II

Number of modes by the MCA.

The explained squared covariance fractions (%) has been used to choose the optimal number of modes retained for MCA when the FA method is applied. Fig. 7 shows squared covariance fractions for three monthly averaged standardized values of temperature from the S4 system. Similar results have been obtained for precipitation and for the other forecast systems. Visual inspection allows to choose 3 modes for the MCA

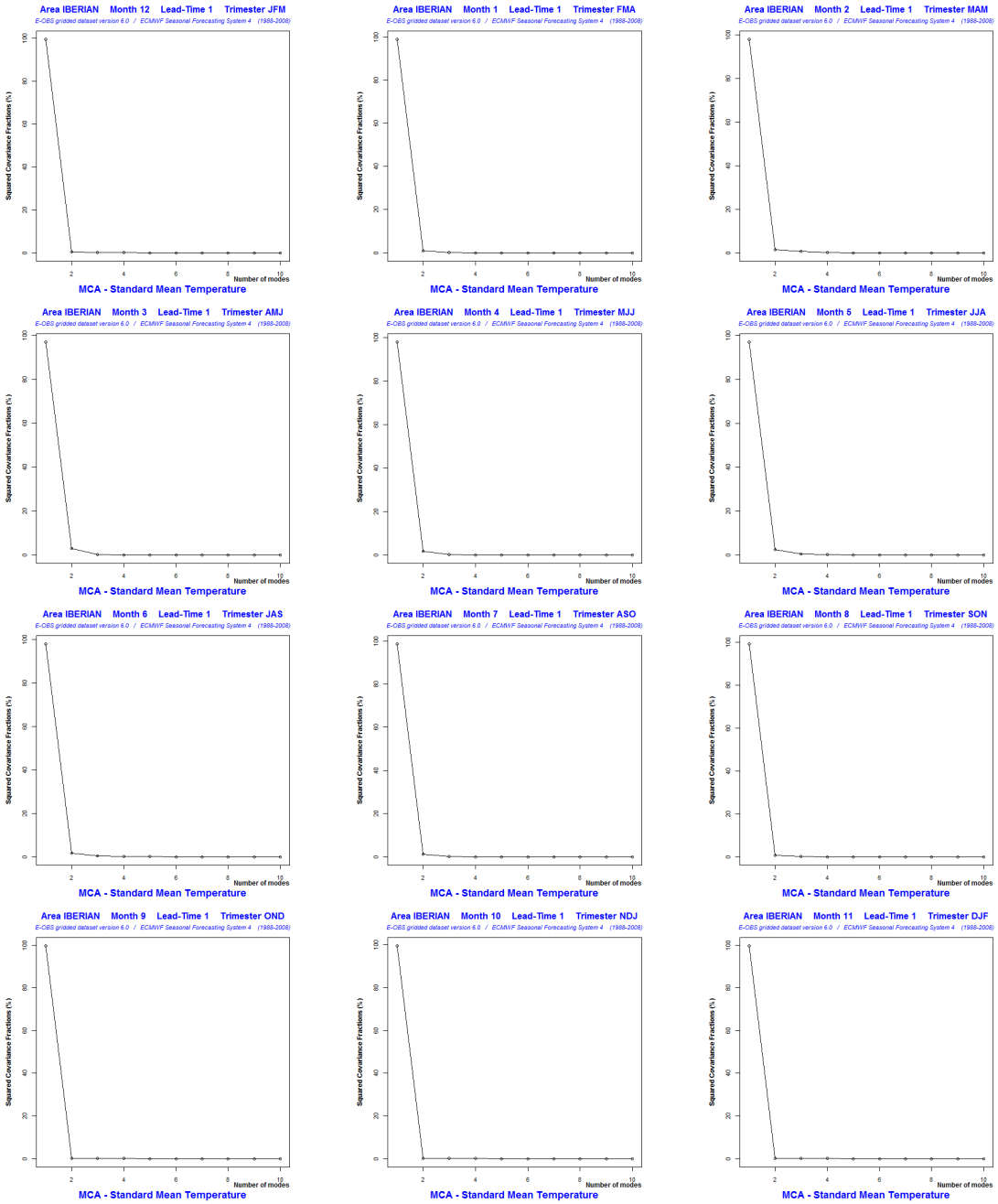


Figure 7. Squared covariance fractions (%) as a function of the retained number of modes for MCA computed for three monthly averaged standardized values of temperature from the S4. Different graphics are referred to 12 successive three monthly periods for the Iberian Peninsula domain and for lead-time 1.