FORECAST QUALITY ASSESSMENT OF SEVERAL SEASONAL CLIMATE PREDICTION SYSTEMS

BSC-ESS-2016

Doo Young Lee, Nube Gonzalez-Reviriego, Verónica Torralba, Nicola Cortesi, Raül Marcos, Albert Soret and Francisco Javier Doblas-Reyes

Earth Sciences Department

Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS)

04 April 2017

TECHNICAL REPORT

Series: Earth Sciences (ES) Technical Report

A full list of ES Publications can be found on our website under:

https://earth.bsc.es/wiki/doku.php?id=library:external:technical_memoranda

Summary

In this report we have assessed the forecast quality of the 10m wind speed and 2m temperature seasonal predictions produced by three different seasonal prediction systems (ECMWF System 4, Météo-France System 3 and Météo-France System 4). To perform the assessment two different reanalysis datasets (ERA-Interim and JRA-55) have been used as a reference. The systematic bias and uncertainty of the seasonal predictions has been minimised with two statistical bias-adjustments methods: simple bias correction and calibration. We have also assembled a multi-model configuration with the three seasonal prediction systems to evaluate the added value of this approach relative to the individual predictions from a wind energy user perspective. In all the cases the deterministic and probabilistic verification measures have been applied in cross-validation, both for the raw and post-processed model outputs. Generally, the ECMWF System 4 shows better skill than other individual prediction systems and the MME prediction indicates consistently higher TCC and FRPSS than the individual models. The cross-validated simple bias corrected predictions show an overall similar performance to the forecasts of raw data for both winter and summer seasons. In winter, the correlation skill of both the raw and bias-adjusted MME predictions show the positively significant performance over the central United States, northern South America, central Europe, eastern Africa and northeastern portion of China. In summer, the significant spatial patterns of the TCC for the MME predictions are mostly concentrated over the tropical region. The FRPSS has a similar distribution to the TCC.

Contents

Index of figures

Index of tables

1. Introduction

Modern society is looking forward to the growth and widespread diffusion of renewable energies such as wind and solar power, hopefully contributing to the major part of the world energy supply. In recent years, their growth shows an increasing and significant trend (Frankfurt School-UNEP/BNEF 2016). Wind power will especially play an increasingly important role in providing a substantial share of renewable energy supply over the coming years (Troccoli et al. 2010). The ability to anticipate and respond to changes in wind energy supply and demand is essential to stabilize and secure the entire electricity network. For this reason, accurate and reliable weather and climate forecasts is required, accordingly to the gradual growth of the need of more and better forecast information in the development and use of wind energy (Troccoli et al. 2010, Vladislavleva et al. 2013).

Previous works have dealt with the sensitivity of the energy system to the variability at either short or long time scales, such as weather forecasts (Amin 2013, Foley et al. 2012, Troccoli et al. 2013, Vladislavleva et al. 2013) or climate change projections (Ebinger & Vergara 2011, IPCC 2012, Koletsis et al. 2016), while there are few research studies on the use of climate predictions for societal applications (e.g. agriculture, energy, health, insurance, etc.) at seasonal time scales due to the improper general perception on their low quality (Doblas-Reyes et al. 2013). Even though the performance of the seasonal climate prediction has been significantly improved, their systematic errors still remain (Feddersen et al. 1999, Wang et al. 2008, Kug et al. 2008, Alessandri et al. 2010). Both climate sciences and climate services communities have tried to deal with this problem for producing better seasonal climate forecast information relevant to user applications (Buontempo et al. 2014, Coelho & Costa 2010, Morse et al. 2005, Palmer et al. 2005, Torralba et al. 2015).

The main aim of the present study is to evaluate the forecast quality of several seasonal climate forecast systems in predicting global wind speed, where simultaneous predicted and observed wind speed values are systematically compared. This is a fundamental step in climate prediction because it allows users to assess whether or not the prediction systems lead to improved forecast with respect to a standard, which is usually the climatology or a simple persistence forecast (Doblas-Reyes et al. 2013). To this end, several deterministic and probabilistic verification measures were applied for individual forecast systems and multi-model ensemble (MME) against two different reanalysis datasets. To reduce the uncertainty and improve the reliability of the seasonal predictions, we apply two statistical bias adjustment techniques: simple bias correction (Pan & den Dool 1998, Leung et al. 1999, Acharya et al. 2013) and calibration (Von Storch 1999, Doblas-Reyes et al. 2005). Then, a multi-model ensemble (MME), based on the combination (with equal weights) of three different independent forecast systems, has been carried out for the enhancement of seasonal predictability and satisfying the high quality needs of the wind-energy community.

The rest of this report is organized as follows. Section 2 presents a brief description of the observational data, seasonal predictions and verification measures used in this study. The results from the forecast quality assessment of the prediction systems and the MME are described in Section 3, together with an analysis of the sensitivity to the observational references. The summary and main conclusions are given in the final section.

2. Data and methodology

In this study we have focused on the quality assessment of the 10m wind speed and 2m temperature seasonal predictions with 1-month lead-time initialised on 1st May (June through August, JJA) and 1st November (December through February, DJF), key variables to wind power supply and energy demand, proceeding from three coupled global seasonal prediction systems: the European Centre for Medium-Range Weather Forecasts seasonal forecast system 4 (ECMWF-S4; Molteni et al. 2011), Météo-France's System 3 (METFR-S3; Alessandri et al. 2011, (Lage Rodrigues 2015)) and Météo-France's System 4 (METFR-S4; Voldoire et al. 2013). Information on the quality assessment of these two variables for the start dates of 1st February and 1st August with 1-month lead time can be found on the Earth System Services (ESS) web catalogue[a]. The criterion for selecting these prediction systems has been the availability of 6-hourly seasonal forecasts for the two variables over the period 1991-2012 and for the four start-dates. See **Error! Reference source not found.** for a brief description of the systems.

Table 1 Description of the coupled atmosphere-ocean general circulation models employed

Model Name	AGCM[b]	Resolution	OGCM[c]	Resolution	Ensemble size
ECMWF-S4	IFS CY36R4	TL255L91	NEMO3.0	1°lat x1°lon L42	51
METFR-S4	ARPEGE5.2	TL127L31	NEMO3.2	1°lat x 1°lon L42	15
METFR-S3	ARPEGE4	~300km x ~300km, L91	OPA8.2	2°lat x 2°lon L31	11

2.1. Forecast systems

The ECMWF-S4 consists of the ECMWF Integrated Forecast Model (IFS) and Nucleus for European Modelling of the Ocean (NEMO) as atmospheric and oceanic components, respectively. Its hindcast (historical forecast) has 51 ensemble members and the standard forecast has seven months' lead time, initialized on the 1st day of every month from 1981 until recently. Details for the ECMWF-S4 can be found in Molteni et al. (2011).

The METFR-S3 (Alessandri et al. 2011, (Lage Rodrigues 2015)) utilizes the version 4 of the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) as atmospheric component. The

[a] www.bsc.es/ESS/catalogue
[b] AGCM: Atmospheric Global Circulation Model
[c] OGCM: Oceanic Global Circulation Model

ocean component is the global version of the Océan PArallélisé (OPA) model version 8.2. Its hindcast has 11 ensemble members, all starting in burst mode on the 1st day of every month at 0 UTC. Simulations are seven-months long and cover the period 1981-2012.

The Météo-France forecast system 4 (METFR-S4, Voldoire et al. 2013) has been running operationally since September 2012. It consists in a 15 ensemble members, 7-month forecast lead, hindcast starting once per month over 1991-2012, based on ARPEGE-Climat version 5.2 coupled with NEMO3.2.

2.2. Observational datasets

For the forecast verification, we have used two reanalysis datasets: the ERA-Interim reanalysis (Dee et al. 2011) and the Japanese 55-year reanalysis (JRA-55; Kobayashi et al. 2015, Harada et al. 2016). ERA-Interim is ECMWF's most recent atmospheric reanalysis, covering the modern satellite era from January 1979 to the present at a spatial resolution of about 80 km (T255). It is based on a 2006 version of the ECMWF Integrated Forecast Model (IFS) and utilizes a four-dimensional variational analysis (4D-Var) for data assimilation. JRA-55 is the high-quality reanalysis and homogeneous climate dataset produced by the Japan Meteorological Agency (JMA) which employs a more sophisticated Data Assimilation (DA) system than JRA-25 (previous version of JRA-55). This is an appropriate dataset for climate research spanning from 1958 to the present at a spatial resolution of 60 km (TL319).

2.3. Methodology

A land mask was introduced to represent information over land only. In order to include points where offshore wind farms might be installed, sea points with a depth equal or less than 50m are not masked.

2.3.1. Post-processing methods

The simple bias correction method adjusts the systematic errors of the model using the standardized anomaly of the ensemble mean. By default, standardized anomalies of the ensemble mean are computed by subtracting the climatology of the ensemble mean and normalizing with the standard deviation of ensemble mean. To estimate the bias adjusted forecast, the standardized anomaly of the ensemble mean is reconstructed by multiplying the observed standard deviation and adding the observed climatology (Pan & den Dool 1998, Leung et al. 1999, Acharya et al. 2013, Torralba et al. 2017).

In the calibration method, we use the variance inflation technique that has been proposed in several studies (Doblas-Reyes et al. 2005, Johnson & Bowler 2009, Charles et al. 2011, Torralba et al. 2017). It assumes that the bias adjusted ensemble forecasts by calibration method should have the same interannual variance as observations. To improve the reliability of the ensemble predictions, the inflation of both the ensemble mean and the ensemble spread is applied to the seasonal predictions.

To derive a more accurate estimate of model prediction performance and avoid overfitting, the standard ‘‘leave-one-out’’ cross-validation method (Michaelsen 1987, Jolliffe & Stephenson 2003) are applied to post-processing procedures. This cross-validation method removes the target year from the climatological means. Both statistical post-processing methods have been applied as in Torralba et al. 2017.

2.3.2. Multi-model ensemble

Many studies have reported that the multi-model ensemble (MME) technique, assigning equal weights to the ensemble mean predictions of each of models, is a competitive method for producing accurate and reliable forecasts (Kharin & Zwiers 2002, Peng et al. 2002, Hagedorn et al. 2005, Min et al. 2009, Weigel et al. 2010). The MME techniques are useful methods for reducing the inherent errors contained in individual models and providing better performance than the constituent individual models (Krishnamurti et al. 2000, Palmer 2000, Pavan & Doblas-Reyes 2000, Peng et al. 2002, Hagedorn et al. 2005, Doblas-Reyes et al. 2005, Yun et al. 2005, Weigel et al. 2008, Min et al. 2009, Lee et al. 2011, 2013a, 2013b). In the present study, only three forecast systems (ECMWF-S4, METFR-S3 and METF-S4) are employed to create the MME because they are the only ones with 6-hourly data available. In order to enhance the deterministic seasonal predictability for wind energy sector, we apply the multi-model combinations as a simple arithmetic mean to the direct ensemble mean of raw and bias-adjusted data sets of the three prediction systems. For the probabilistic MME predictions, forecast probabilities for each tercile (above, near, and below normal) category are estimated separately for each individual model and then such probabilities for each category are combined by applying the simple average with equal weights.

2.3.3. Forecast quality measures

To investigate the ability of the forecast systems to reproduce adequately the observed 10m wind speed and 2m temperature variability, a set of verification measures (deterministic and probabilistic) are applied for the simultaneous comparison of predicted and observed values. In this report, the temporal correlation coefficient (TCC) and fair ranked probability skill score (FRPSS) are estimated over the retrospective forecast periods (Jolliffe & Stephenson 2003, Wilks 2006). The uncertainty in the skill difference between two forecast systems is also quantified. More details on the sensitivity to the observation reference are in Section 3.1.

The TCC is one of the deterministic measures more widely used. To measure the TCC, the ensemble mean forecast and observed values are first converted to cross-validated anomalies by subtracting the climatological average value of the forecast and observed field. The TCC is designed to analyze the spatial distribution of forecast skill and to measure the strength of linear relationship between both departures from the climatological values for forecasts and their corresponding observations, respectively. Positive TCC values indicate that the prediction system can provide additional information relative to the past climatology. The statistical significance of the TCCs is computed using a Student’s two tailed t-test, being the regions with significant correlation at the 90% confidence level hatched.

One of the more commonly used probabilistic measures to evaluate forecasts of multiple categories is the ranked probability skill score (RPSS; Epstein 1969, Murphy 1971, Daan 1985). The RPSS measures the cumulative squared error between the categorical forecast probabilities and the observed categorical probabilities relative to a reference (or standard baseline, Wilks 2006, Weigel et al. 2007, Acharya et al. 2014). When the value of RPSS equals to 1, it implies that the observed category is always predicted with 100% confidence. RPSS = 0 implies that the prediction skill is same as reference prediction (climatology, in our case) and a score <0 means that the forecast system performs worse than climatology.

The RPSS is an unfair skill measure for the evaluation of several prediction systems with different number of ensemble members. In this regards, Ferro (2007) and Ferro et al. (2008) show that the RPS can be adjusted to provide a fair way to evaluate ensembles. In this study, we have applied the fair version of RPSS to seasonal forecasts of the both individual prediction systems and multi-model ensemble. In this way, it is possible to compare forecasts with a different ensemble size. The statistical significance of the FRPSS is calculated using a one-tailed Z-test at the 95% confidence level.

Figure 1 shows a schematic diagram with all the steps followed in the forecast quality assessment of seasonal prediction systems. Even through various results through verification measures are obtained for five different types of data sets (raw data, simple bias corrected data with and without cross-validation, calibrated data with and without cross-validation) in terms of the three individual seasonal forecast systems and MME, in this report we have only focused on two types of data sets: the raw and simple bias corrected data with cross-validation for 10m wind speed and two seasons DJF and JJA. Verification results for other types of data sets, two other seasons (MAM and SON) and 2m temperature variable, refer to the ESS web catalogue[d].

[d] www.bsc.es/ESS/catalogue

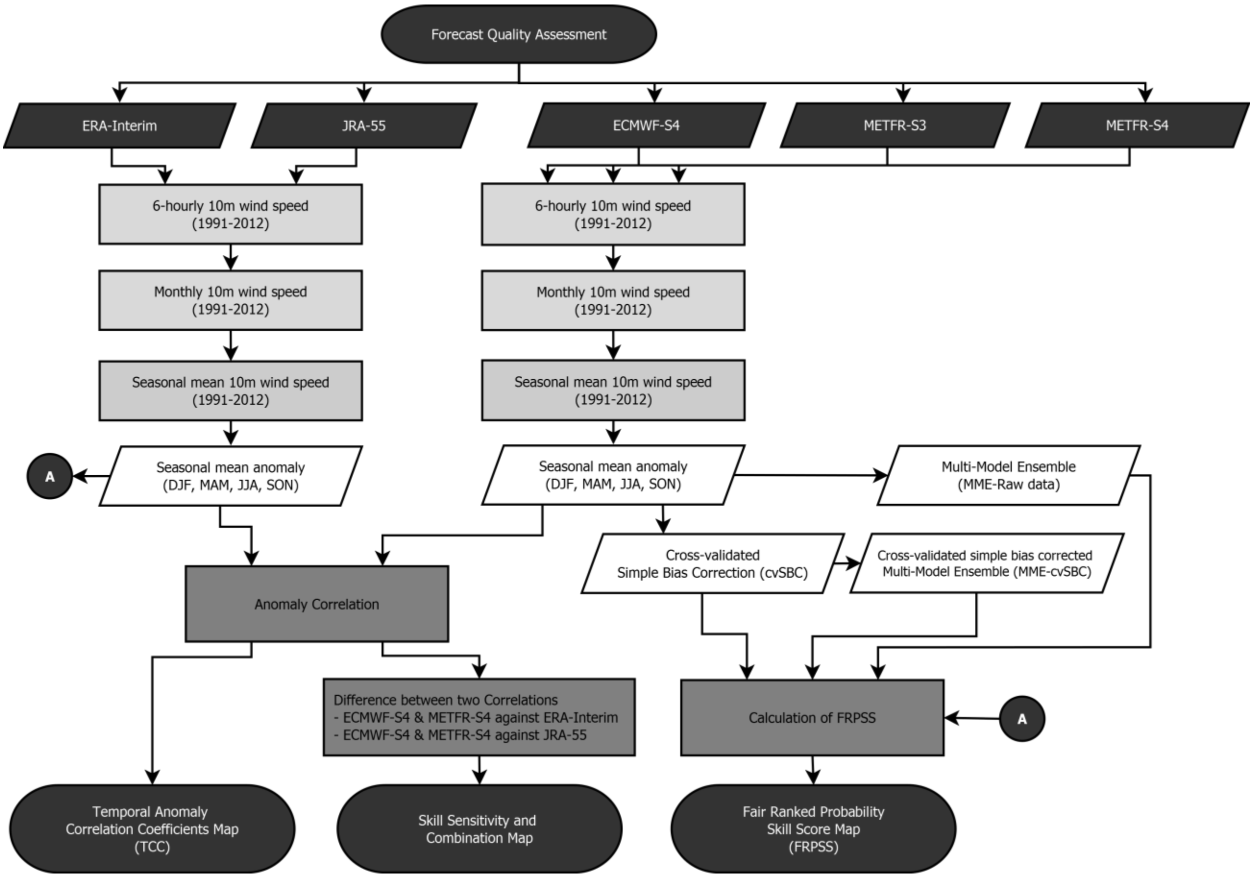

***Figure 1**. Flow chart of the forecast quality assessment of seasonal forecast systems*

Flow chart on the sequence of the steps in a process for the forecast quality assessment of seasonal climate prediction systems in terms of 10m wind speed for the raw and cross-validated simple bias corrected datasets. The same procedure is also adopted for the forecast quality assessment for 2m temperature.

3. Results

3.1. Sensitivity to the observational reference

To investigate the sensitivity of the forecast verification to the observational choices, we have obtained the maps of differences between the correlation coefficients which are obtained from two different prediction systems, ECMWF-S4 (F_E) and METFR-S4 (F_M), against two reanalysis datasets, ERA-Interim (O_E) and JRA-55 (O_J). The correlation coefficients between the forecasts generated by the two prediction systems and two observational datasets are defined by:

$$R_{EE} = \mathrm{Corr}\ (F_E, O_E),\ R_{ME} = \mathrm{Corr}\ (F_M, O_E),\ R_{EJ} = \mathrm{Corr}\ (F_E, O_J),\ R_{MJ} = \mathrm{Corr}\ (F_M, O_J)$$

The maps of differences between two correlation coefficients, R_{EE}-R_{ME} and R_{EJ}-R_{MJ} for 10m wind speed at each grid point for winter (DJF) and summer (JJA) are shown in Figure 2a and b and Figure 3a and b, respectively. Furthermore, we applied the methods suggested by Steiger (1980), Zou (2007) and Siegert et al. (2016) to calculation of the differences between two dependent correlation coefficients. Hatched lines highlight areas where differences are statistically significant at the 90% confidence level from a two-sided Student's t-test.

To evaluate and summarize the information from skill differences against two different reanalysis datasets more easily, we also created a single map in which their information is combined (Figure 3c and Figure 3c), taking into account the sign and significance of differences. The red (blue) colors mean positive (negative) values of differences in correlation for both cases (against ERA-Interim and JRA-55), indicating that the sign of difference between correlations is always independent to the reference dataset selected. Yellow and green colors show the other possible combinations of differences and significance, highlighting those regions where the sign of the correlation differences is influenced by the choice of the reference dataset we use for the forecast quality assessment.

In the maps of differences between correlation coefficients in the ECMWF-S4 and METFR-S4 prediction systems in terms of 10m wind speed for winter, some parts of the North America, northern Europe, northern Africa, and northern portion of China show statistically significant positive values against both reanalysis datasets (Figure 2a and b). That means that the ECMWF-S4 produces more predictable wind speeds in those regions than the METFR-S4. In the combination map of skill differences (Figure 2c), the regions which the skill differences show the opposite signs (e.g., yellow or green areas) associated with the larger discrepancies according to the different reanalysis datasets, are found mainly in some parts of South America and Africa. So we should be cautious in the interpretation of the TCC in those regions where the use of one reanalysis or another can provide different results in correlation.

The regions which have the significant skill difference between two prediction systems against the ERA-Interim (Figure 3a) and JRA-55 (Figure 3b) reanalysis datasets for 10m wind speed during JJA are much smaller than those during DJF. This indicates that for JJA the the two prediction systems produce more similar results than in DJF independently of which reference dataset has been selected. The combination map of the correlation differences shows a wide distribution of the uncertainty due to the reference dataset, being the yellow or green areas mainly over the western Brazil, central equatorial Africa, some parts of India and southwestern Australia (Figure 3c).

Similar results illustrating the sensitivity of the seasonal forecast verification with respect to different observational datasets for other prediction systems and other variables, such as 2m temperature, can be found in the ESS web catalogue[e]. Generally, the uncertainty regions for 2m temperature decreases, compared to those for 10m wind speed, indicating that for this variable the selection of the most appropriate reanalysis for verification is not as crucial as for wind speed. These results show that we need to be vigilant on the evaluation and comparison of performance of the prediction systems in the particular regions indicating opposite sign of skill difference according to the selection of the reanalysis.

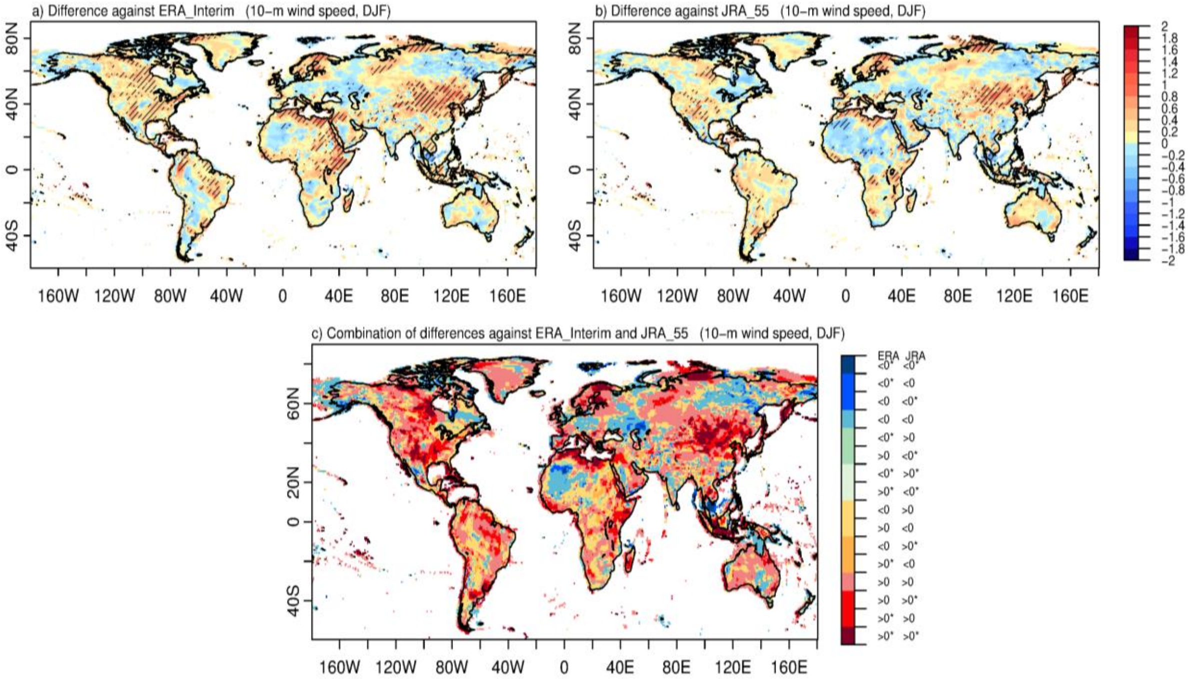

***Figure 2**. Difference between TCCs for 10m wind speed during winter (DJF).*

a) Map of difference between TCCs of two prediction systems (ECMWF-S4 and METFR-S4) against ERA-Interim for 10m wind speed over the globe in winter (DJF) during period 1991-2012 with 1-month lead time. b) Same as a), but for JRA-55 reanalysis as reference. Hatching lines indicate differences significant at the 90% confidence level from a two-tailed Student's t-test. c) Combination map of correlation differences classified by sign and significant differences in TCC for the two prediction systems against each reanalysis. The asterisk in the color bar indicates significant difference at the

[e] www.bsc.es/ESS/catalogue

90% confidence level from a two-tailed Student's t-test.

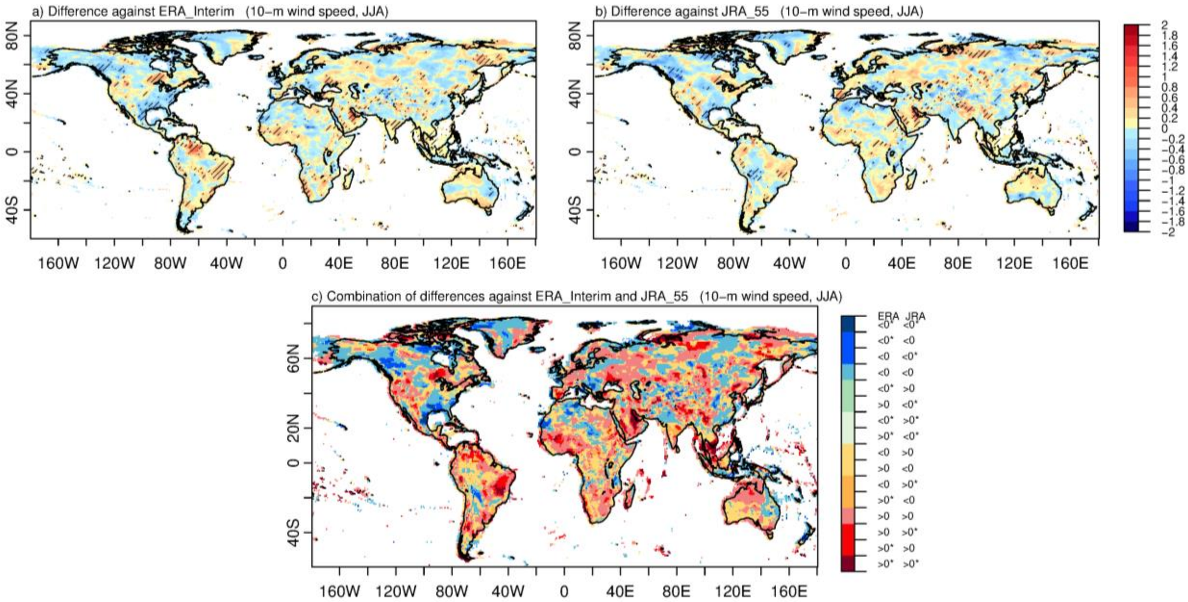

***Figure 3**. Same as Fig. 2, but for summer (JJA).*

3.2. Forecast verification of individual models

Verification measures used in this report to assess the forecast quality of the individual seasonal prediction systems include both deterministic (TCC) and probabilistic (FRPSS) measures. Both verification measures have been computed in cross-validation for raw and simple bias corrected data and the ERA-Interim reanalysis has been used as a reference. Results from this analysis are shown for 10m wind speed and two seasons: winter (DJF) and summer (JJA), focusing on 1-month lead time. Similar analysis was done for 2m temperature in the ESS web catalogue[f].

3.2.1. Raw model output

Figure 4a, c, and e illustrate the TCC global spatial distribution 10m wind speed raw data provided by individual forecasts systems in winter (DJF) over the period 1991–2012. Overall, common areas showing high correlation values from the three prediction systems are confined to the Maritime Continent, central U.S. and northern South America (Figure 4a, c, and e). Though the prediction skill of the ECMWF-S4 in Figure 4a is superior to that of METFR-S4 (Figure 4c) and METFR-S3 (Figure 4e), the correlation values for ECMWF-S4 are not significant in some regions such as the northern part of Africa, southern Europe and eastern Russia.

The correlation coefficient of the ECMWF-S4 seasonal forecast shows statistically significant positive scores over the North America, northern South America, most of maritime continent, eastern Africa and northern portion of China (Figure 4a). The METFR-S4 (Figure 4c) does not show the significant positive skill over China, eastern Africa, and most of North America and there are more areas with negative correlation. In Figure 4e, the METFR-S3 shows similar features as METFR-S4 except for the positive skill in central Europe and eastern Africa.

Figure 5a, c, and e illustrate the spatial distribution of the TCC for raw model output in summer. Compared to winter, remarkably, the distribution of the positive significant skills of the prediction systems tends to concentrate along the tropical areas such as maritime continents, Indian subcontinent, and northern South America.

Correlation skill of the ECMWF-S4 (Figure 5a) indicate the positively significant performance over the eastern Brazil and near Venezuela in South America, eastern Africa, Indian subcontinent and most of maritime continent. The positively high correlations of the METFR-S4 forecast have similar patterns as those of the ECMWF-S4 (Figure 5c), excluding the northeastern part of South America. This prediction system also shows statistically significant negative correlation skill in the proximity of the Gulf of Guinea. The METFR-S3 simulation shows significantly positive (negative) performance over the eastern Canada and central Brazil (central Russia), as shown in Figure 5e.

[f] www.bsc.es/ESS/catalogue

Left column of Figure 6 displays the global FRPSS distribution of the raw datasets from the different prediction systems using the FRPS of the ERA-Interim reanalysis as the reference forecast in terms of 10m wind speed for winter season. The forecasts of prediction systems are valuable only when RPSS > 0, i.e., when the forecasts perform better than the reference. Hatch lines show area where FRPSS is significant at the 95% confidence level from a one-tailed Z-test.

The ECMWF-S4 forecasts have significantly positive skill over the central United States, northeastern South America, northern China, and near the Indonesia. The negative skill scores for the METFR-S4 are found over large areas, though they are never significant. The number of significant regions with positive skill decreases compared to the ECMWF-S4. The positively significant skills can be seen in the western Russia and southern Indonesia (Figure 6c). The METFR-S3 has few significant positive regions, even less than in METFR-S4. The METFR-S3 forecast indicates similar negative patterns to METFR-S4. There is positive skill in the some parts of central Europe, North America, and maritime continent (Figure 6e).

The global patterns of the FRPSS for raw model datasets in summer are shown in the left column of Figure 7. The tropical distributions of the significant positive FRPSS of probabilistic forecasts from the prediction systems in summer are almost similar to those of the TCCs. Figure 7a shows the significantly positive skills for the ECMWF-S4 over the maritime continent and Indian subcontinent. The skill performance of the METFR-S4 is comparable to that of the ECMWF-S4, except for decreasing of significant positive skill areas in the maritime continent and non-significant skill in the Indian subcontinent (Figure 7c). In Figure 7e most areas have the negative skill, except for the significantly positive skill in the western Brazil.

3.2.2. Bias adjusted predictions

To examine deterministic seasonal prediction skill of the cross-validated simple bias corrected data sets, the correlation coefficients between ERA-Interim reanalysis and ensemble mean anomalies for 10m wind speed from each prediction system are calculated for winter (Figure 4b, d, and f) and summer seasons (Figure 5b, d, and f) over 22 years.

The post-processed prediction of the ECMWF-S4 generally shows a similar performance to the raw prediction in terms of significant correlation skill areas, even though the leave-one-out cross-validation was applied to the bias adjusted predictions (Figure 4b) and this could produce a decrease of the skill (Torralba et al. 2017). In Figure 4d, the bias adjusted prediction of the METFR-S4 still has a wide distribution of negative correlation, similarly to the prediction of raw data in Figure 4c. For the METFR-S3 prediction, the skill performance of the simple bias correction is indistinguishable from that of the raw data (Figure 4f).

The significant prediction skill of the three post-processed seasonal climate predictions in summer shows a similar spatial distribution to that of the raw prediction data sets (Figure 5b, d, and f). Areas with significant skill of the bias adjusted ECMWF-S4 prediction are similar to those of the raw prediction skill, but they are distributed around the tropical areas, as opposed to winter (Figure 5b). Bias corrected skill of the METFR-S4 forecast in Figure 5d has almost the same distribution patterns as the raw prediction skill of Figure 5c. The significant skill of the bias corrected prediction is lower in northern South America, Indian subcontinent and maritime continent, compared to the ECMWF-S4 (Figure 5d). In Figure 5f, there are no significant differences in skill between raw and bias corrected prediction. The METFR-S3 has a similar distribution to METFR-S4 except for the significant skill in central Brazil, eastern Canada and central Siberia.

The right columns (b, d and f) of Figure 6 and Figure 7 depict the global spatial distribution of the FRPSS for the cross-validated bias-adjusted data sets. In Figure 6b, the significant skill of simple bias corrected ECMWF-S4 prediction has similar patterns as the raw prediction skill of Figure 6a. The significant high skill of the bias-adjusted prediction for the METFR-S4 is almost indistinguishable from that of the raw prediction (Figure 6d). It is difficult to find the skill improvement in the bias-adjusted prediction of METFR_S3 (Figure 6f) relative to the raw prediction (Figure 6e).

Compared to the raw prediction skill for the ECMWF-S4 in Figure 7a, there are no significant differences, except for slight decrease of skills over the southern part of China (Figure 7b). The difference in skill between the raw (Figure 7c) and bias adjusted (Figure 7d) prediction for the METFR-S4 is distinctly shown in the central Africa and northern China. In Figure 7f, the METFR_S3 prediction obtained from the bias-correction shows the similar pattern to the raw prediction of Figure 7e. Even though the bias correction minimizes the forecast errors (see RMSSS in the ESS web catalogue), the degeneracy in the skill is found. But it can be negligible.

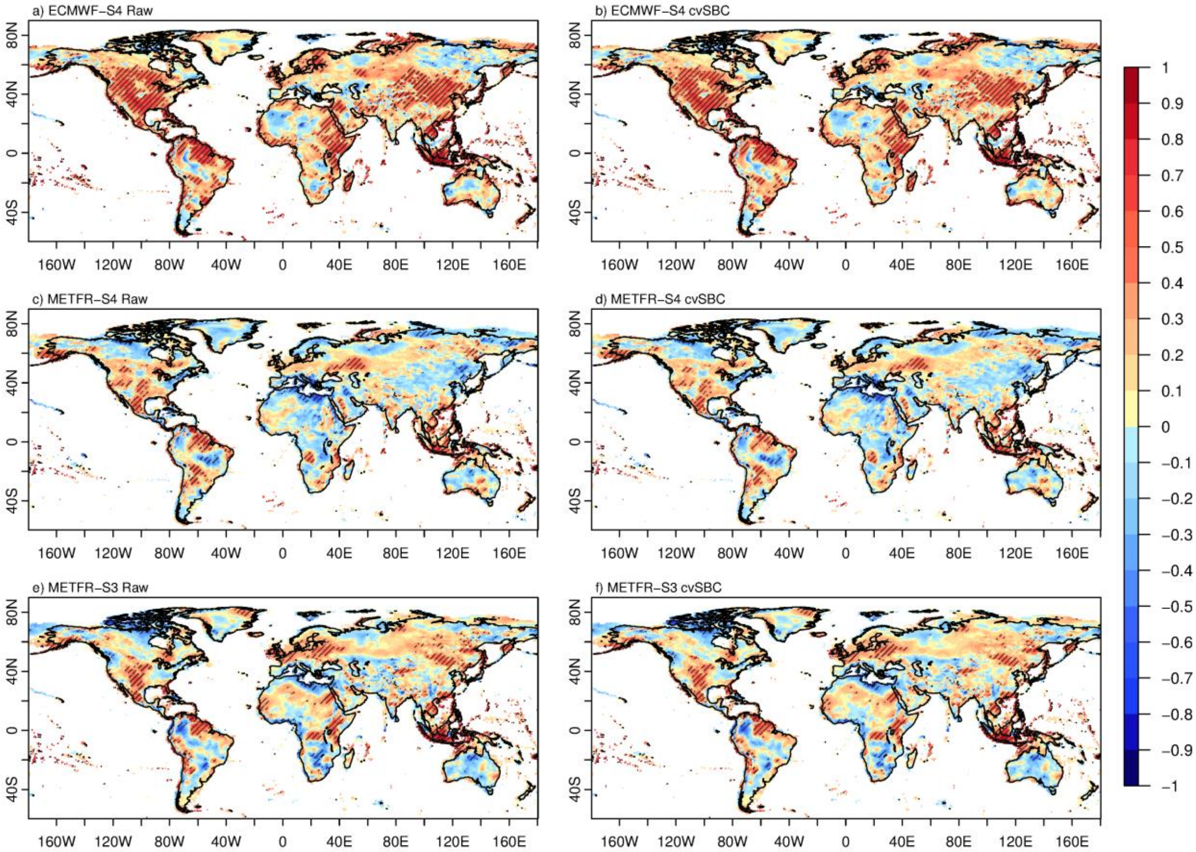

***Figure 4**. Temporal correlation coefficient for 10m wind speed during winter (DJF)*

Temporal correlation coefficient (TCC) between the observed and predicted 10m wind speed during winter (DJF) for period 1991-2012. Left and right columns show the TCC for raw (a, c, and e) and cross-validated simple bias corrected (b, d, and f) data against ERA-Interim reanalysis. Hatched areas highlight regions where TCC is significant at the 90% confidence level from a two-tailed Student's t-test.

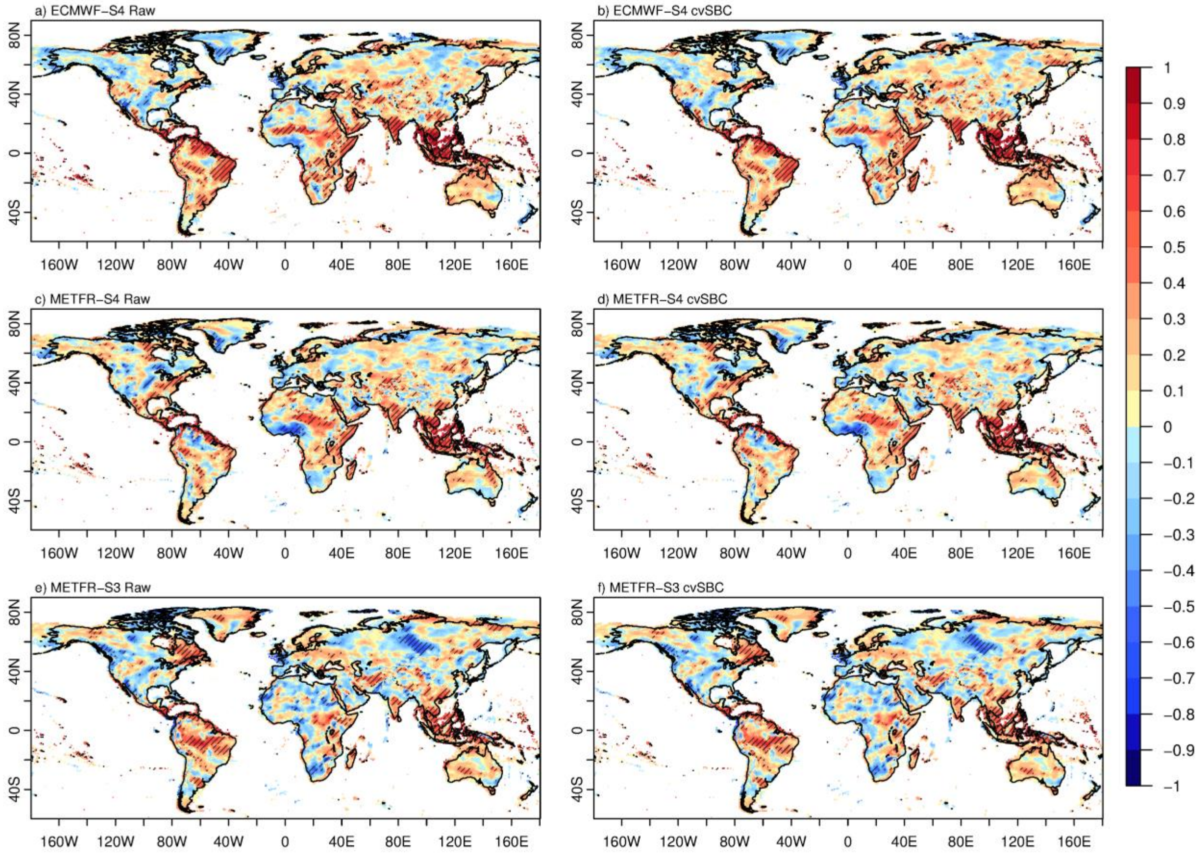

Figure 5*. Same as Fig. 4, but for summer (JJA).*

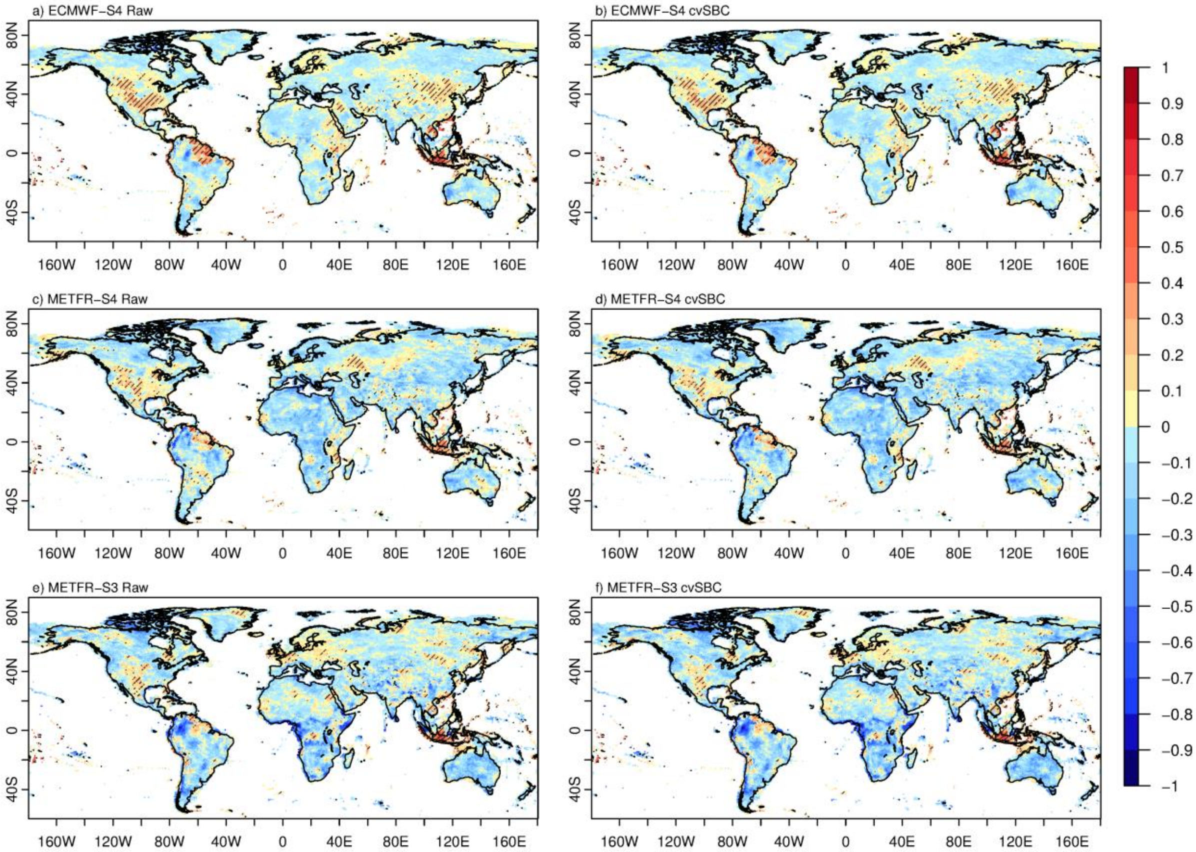

***Figure 6**. FRPSS for 10m wind speed during winter (DJF)*

FRPSS of each individual model with respect to the reference observed climatology for 10m wind speed during winter (DJF) for period 1991-2012. Left and right columns show FRPSS for raw (a, c, and e) and cross-validated simple bias corrected (b, d, and f) data against ERA-Interim reanalysis. Hatched areas highlight regions where the FRPSS is significant at the 95% confidence level from a one-tailed Z-test.

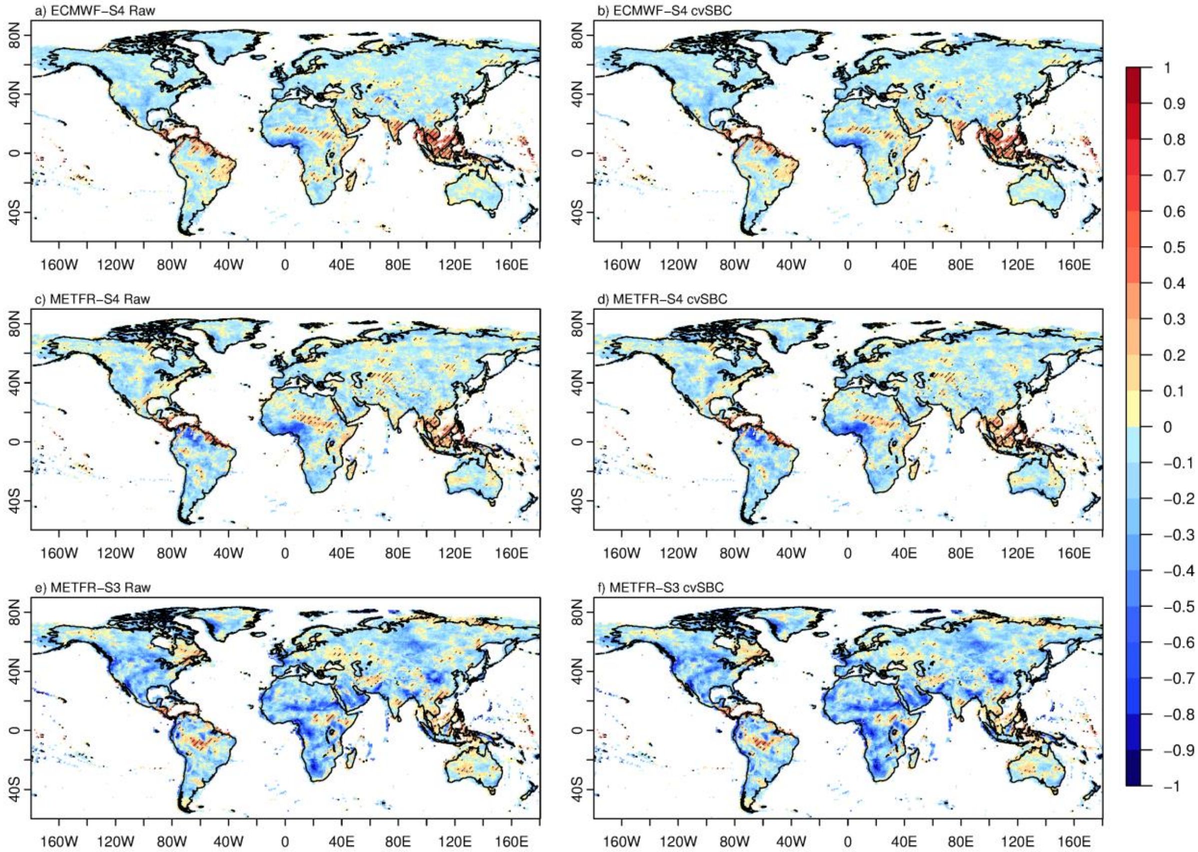

***Figure 7**. Same as Fig. 6, but for summer (JJA).*

3.3. Forecast verification of Multi-Model Ensemble predictions

In this section we describe and compare the performance of the 10m wind speed multi-model ensemble predictions of the original raw data sets and cross-validated bias adjusted data sets in terms of TCC and FRPSS.

Spatial maps of TCC for predictions of two data sets (raw and simple bias corrected data) based on MME technique with the equal weights are shown in Figure 8 (winter) and Figure 9 (summer). Statistical significance at the 90% confidence level from a two tailed t-test of the TCCs is also provided. In winter, the performance of the cross-validated bias adjusted multi-model ensemble prediction indicates almost the same distribution as that of the MME prediction of the original raw data. The correlation skill of the both MME predictions show the positively significant performance over the central United States, northern South America, most of maritime continent, central Europe, eastern Africa and eastern portion of China (Figure 8).

The raw MME prediction in summer shows significantly positive high performance over the northern South America, eastern Africa, Indian subcontinent, and most of maritime continent. Spatial pattern of the bias corrected MME prediction skill is similar to that of the raw prediction. In both the MME predictions, it is difficult to find positively significant skill over the North America and Eurasia continents, compared to winter season (Figure 9).

Comparing to the TCC spatial patterns of individual climate prediction systems for winter, it is clear that the significant temporal anomaly correlations for the MME predictions in Figure 8 are obviously better than the METFR-S3 and METFR-S4 in Figure 4. However, they are worse than the ECMWF-S4 of Figure 4 in most of the regions, except over central Europe. For summer, the significant TCCs of the MME predictions in Figure 9 are comparable to (or slightly better than) those of the individual prediction systems in Figure 5.

FRPSS values of the MME predictions obtained from the original raw and cross-validated bias-adjusted data sets for winter and summer are presented in Figure 10 and Figure 11, respectively. Generally, the significant FRPSS distribution of the MME prediction from the raw data sets displays results similar to those from the bias corrected data sets for both seasons. The positive values of the FRPSS shown in both results mean that the forecast from the MME technique performs better than observed climatology.

In Figure 10a, the winter MME prediction of the raw data depicts significant positive skill over the central United States, northern South America, northern China, and the Indonesia. Negative skill scores are found over the large areas for the raw MME prediction. Most of the skill patterns of the bias adjusted MME prediction are almost indistinguishable from those of the raw MME prediction (Figure 10).

In summer, the distribution of significantly positive skill for the raw MME prediction is found over the maritime continent, Indian subcontinent, central Africa and some parts of northern South America (Figure 11a). There are no significant differences in skill between the raw and cross-validated bias correction for the MME prediction. As opposed to winter, there is no positively significant skill over the North America in both raw and bias adjusted MME predictions (Figure 11). From this fact it may be deduced that the predictions of individual models over the same region for summer have poor skills (Figure 7).

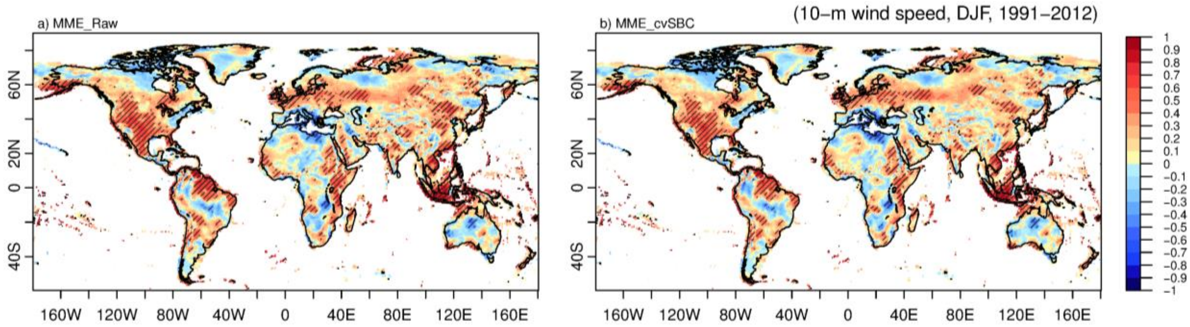

***Figure 8**. TCC of multi-model ensemble for 10m wind speed during winter (DJF)*

TCCs between the observed and predicted 10m wind speed over the globe during winter (DJF) for period 1991-2012. Left and right columns show TCC for (a) raw and (b) cross-validated simple bias corrected multi-model ensemble (MME) data against ERA-Interim reanalysis. Hatched areas highlight regions where correlation is significant at the 90% confidence level from a two-tailed Student's t-test.

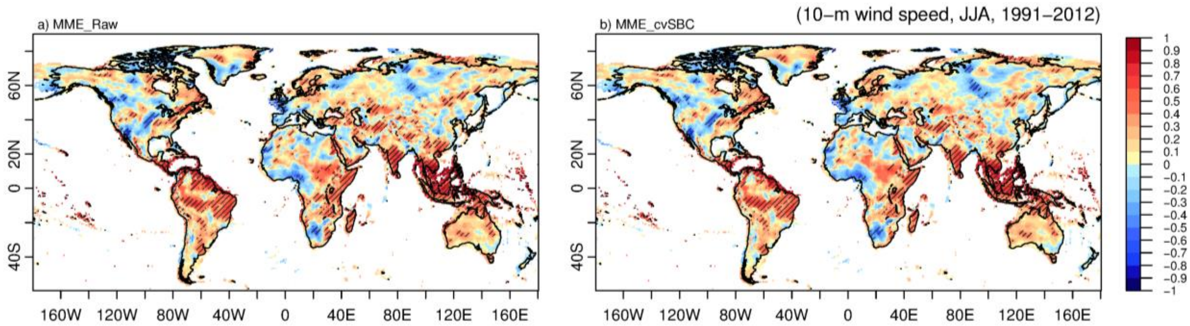

***Figure 9**. Same as Fig. 8, but for summer (JJA)*

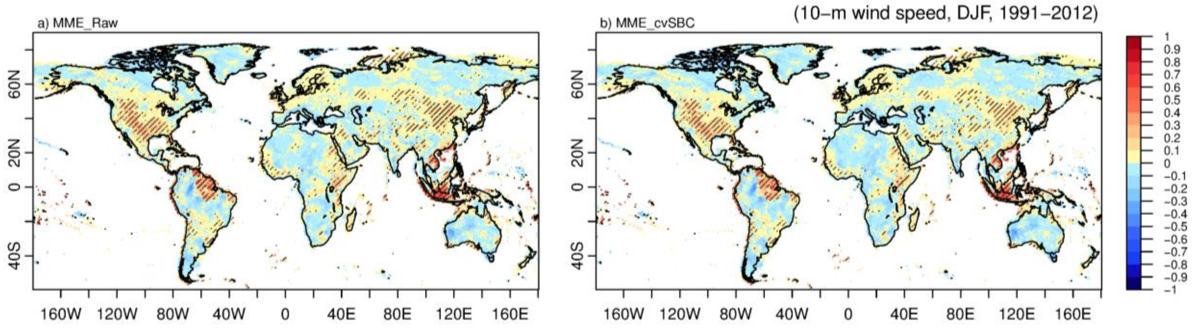

***Figure 10**. FRPSS of multi model ensemble for 10m wind speed during winter*

FRPSS of multi model ensemble (MME) with respect to the reference observed climatology for 10m wind speed over the globe during winter (DJF) for period 1991-2012. Left and right columns show the skill score for (a) raw and (b) cross-validated simple bias corrected MME data against ERA-Interim reanalysis. Hatch areas highlight regions where the FRPSS is significant at the 95% confidence level from a one-tailed Z-test.

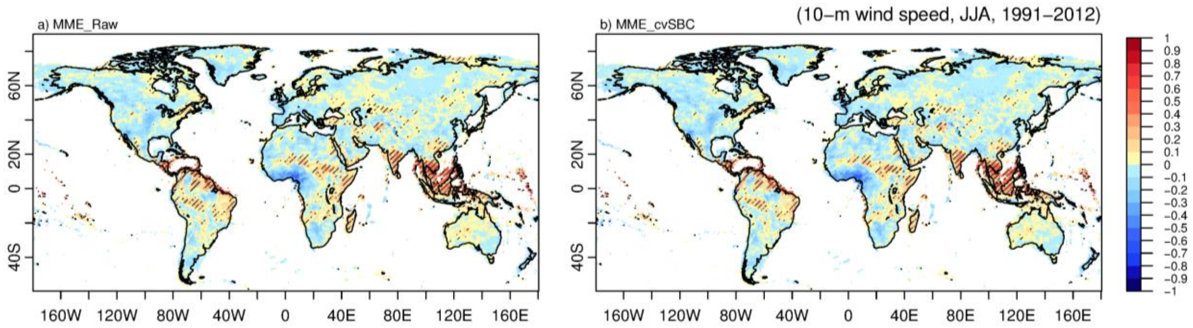

***Figure 11**. Same as Fig. 10, but for summer (JJA).*

4. Conclusions

In this study we have investigated the forecast ability of three global coupled seasonal climate prediction systems (ECMWF-S4, METFR-S3 and METFR-S4) for both 10m wind speed and 2m temperature (see the catalogue web site for temperature). The aim was to provide an overview of the seasonal prediction systems quality to predict climate variables that can be relevant for wind energy users. In this work we have developed more useful and reliable climate information that is tailored to the wind energy industry.

The assessment of the wind speed forecast quality has been carried out by deterministic and probabilistic verification measures (TCC and Fair RPSS) over the 22 years period 1991–2012. Then, two statistical post-processing techniques have been applied to the original raw forecasts to reduce the systematic model bias and improve the reliability and accuracy of forecasts. The sensitivity of the forecast verification to the reference dataset selected for the evaluation of three climate prediction systems forecasts has been explored by two different reanalysis datasets, ERA-Interim and JRA-55, in terms of differences between the correlation coefficients.

The results reveals that ECMWF-S4 shows better skill than the other prediction systems: globally averaged winter TCCs for raw data from ECMWF-S4, METFR-S4 and METFR-S3 are 0.175, 0.023 and 0.066, respectively. The cross-validated simple bias adjusted predictions provided similar performance to the predictions of raw data for both winter and summer seasons, which is an expected result taking into account the simple bias correction only changes the mean and the variance of the predictions to resemble to those observational parameters. Regions with the highest sensitivity, according to the change of reference data, are more widespread in South America and Africa. The geographical distribution of such high uncertainty that comes from the observational reference for temperature (not shown) is generally reduced than that for wind speed. This indicates that the sensitivity to the reference choice is higher for wind speed than for temperature. In case of regions associated to large skill differences depending on the choice of the reanalysis, special care should be taken to avoid an over (or under) estimation of the skill in a particular location.

Using the MME approach (with equal weights for each forecast system), we have also tried to further enhance the predictability of the seasonal forecasts. The significant skill of the bias adjusted MME prediction for the TCC and FRPSS has almost the same distribution as that of the raw prediction. The MME predictions show a slight improvement in skill over some areas, particularly in central Europe, compared to the ECMWF-S4. Therefore, the use of the multi-model can be recommended for those users who are interested in the climate information for particular locations such as central Europe where the MME has better skill than the individual prediction systems.

From the comparison of spatial distribution of skill, we have seen that the performance of the MME prediction is derived from the main features of the each of the integrated prediction systems. For this reason, the enhancement of the skill of each individual prediction system is necessary to improve seasonal predictions, especially using the MME technique. Nonetheless, even in its present state, this forecast quality assessment demonstrates the possibility of providing relevant climate information to improve the current sources of information used in wind energy applications and decision-making.

In this work we have shown, for the first time for wind speed variable, the advantages and drawbacks of using the MME based on the combination of independent different forecast systems. Furthermore, the results of this study suggest that the wind-energy sector can take advantage of the seasonal prediction systems to reduce the uncertainty of seasonal energy estimates. More detailed information for other seasons, variables (temperature) and post-processed data can be found in the catalogue web site (www.bsc.es/ESS/catalogue).

5. Acknowledgements

The authors acknowledge funding support from the RESILIENCE (CGL2013-41055-R) project, funded by the Spanish Ministerio de Economía y Competitividad (MINECO). We are grateful to ECMWF and Météo France as the supporting institutions that provide with climate prediction datasets.

6. References

Acharya, N. et al., 2013. On the bias correction of general circulation model output for Indian summer monsoon. *Meteorological Applications*, 20(3), pp.349–356.

Acharya, N. et al., 2014. Prediction of Indian summer monsoon rainfall: A weighted multi-model ensemble to enhance probabilistic forecast skills. *Meteorological Applications*, 21(3), pp.724–732.

Alessandri, A. et al., 2011. Evaluation of Probabilistic Quality and Value of the ENSEMBLES Multimodel Seasonal Forecasts: Comparison with DEMETER. *Monthly Weather Review*, 139, pp.581–607. Available at: http://dx.doi.org/10.1175/2010MWR3417.1.

Alessandri, A. et al., 2010. The INGV–CMCC Seasonal Prediction System: Improved Ocean Initial Conditions. *Monthly Weather Review*, 138(7), pp.2930–2952.

Amin, M., 2013. Energy: The smart-grid solution. *Nature*, 499(7457), pp.145–147. Available at: http://www.nature.com/nature/journal/v499/n7457/full/499145a.html%5Cnhttp://www.nature.com/nature/journal/v499/n7457/pdf/499145a.pdf.

Buontempo, C. et al., 2014. Climate service development, delivery and use in Europe at monthly to inter-annual timescales. *Climate Risk Management*, 6, pp.1–5. Available at: http://dx.doi.org/10.1016/j.crm.2014.10.002.

Coelho, C.A.S. & Costa, S.M.S., 2010. Challenges for integrating seasonal climate forecasts in user applications. *Current Opinion in Environmental Sustainability*, 2(5–6), pp.317–325. Available at: http://dx.doi.org/10.1016/j.cosust.2010.09.002.

Daan, H., 1985. Sensitivity of Verification Scores to the Classification of the Preditand. *Mon.\ Wea.\ Rev.*, 113(9), pp.1384–1392.

Dee, D.P. et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), pp.553–597.

Doblas-Reyes, F.J. et al., 2013. Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4), pp.245–268.

Doblas-Reyes, F.J., Hagedorn, R. & Palmer, T.N., 2005. The rationale behind the success of multi model ensembles in seasonal forecasting-II. Calibration and combination. *Tellus A*, 57(3), pp.234–252.

Ebinger, J. & Vergara, W., 2011. *Climate Impacts on Energy Systems*, Available at: http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8697.

Epstein, E.S., 1969. A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology*, 8, pp.985–987.

Feddersen, H., Navarra, A. & Ward, M.N., 1999. Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *Journal of Climate*, 12(7), pp.1974–1989.

Ferro, C. a. T., 2007. Comparing Probabilistic Forecasting Systems with the Brier Score. *Weather and Forecasting*, 22(5), pp.1076–1088.

Ferro, C. a. T., Richardson, D.S. & Weigel, A.P., 2008. On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15, pp.19–24.

Foley, A.M. et al., 2012. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1), pp.1–8. Available at: http://dx.doi.org/10.1016/j.renene.2011.05.033.

Frankfurt School-UNEP/BNEF, 2016. *Global Trends in Renewable Energy Investment 2016*,

Hagedorn, R., Doblas-Reyes, F.J. & Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 57(3), pp.219–233.

Harada, Y. et al., 2016. The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan*, 94(3). Available at: http://www.crossref.org/deleted_DOI.html.

IPCC, 2012. *Renewable Energy Sources and Climate Change Mitigation*, Available at: http://www.cro3.org/cgi/doi/10.5860/CHOICE.49-6309.

Jolliffe, I.T. & Stephenson, D.B., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169207005001214.

Kharin, V. V. & Zwiers, F.W., 2002. Climate predictions with multimodel ensembles. *Journal of Climate*, 15(7), pp.793–799.

Kobayashi, S. et al., 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), pp.5–48. Available at: https://www.jstage.jst.go.jp/article/jmsj/93/1/93_2015-001/_article.

Koletsis, I. et al., 2016. Assessment of offshore wind speed and power potential over the Mediterranean and the Black Seas under future climate changes. *Renewable and Sustainable Energy Reviews*, 60, pp.234–245. Available at: http://dx.doi.org/10.1016/j.rser.2016.01.080.

Krishnamurti, T.N. et al., 2000. Multimodel Ensemble Forecasts for Weather and Seasonal Climate. *Journal of Climate*, 13(23), pp.4196–4216.

Kug, J.S., Lee, J.Y. & Kang, I.S., 2008. Systematic Error Correction of Dynamical Seasonal Prediction of Sea Surface Temperature Using a Stepwise Pattern Project Method. *Monthly Weather Review*, pp.3501–3512.

Lage Rodrigues, L.R., 2015. *Calibration and combination of seasonal climate predictions in tropical and extratropical regions*.

Lee, D.Y., Ahn, J.B., et al., 2013. Improvement of grand multi-model ensemble prediction skills for the coupled models of APCC/ENSEMBLES using a climate filter. *Atmospheric Science Letters*, 14(3), pp.139–145.

Lee, D.Y., Ahn, J.B.J.B. & Ashok, K., 2013. Improvement of multimodel ensemble seasonal prediction skills over east asian summer monsoon region using a climate filter concept. *Journal of Applied Meteorology and Climatology*, 52(5), pp.1127–1138.

Lee, D.Y., Ashok, K. & Ahn, J.B., 2011. Toward enhancement of prediction skills of multimodel ensemble seasonal prediction: A climate filter concept. *Journal of Geophysical Research Atmospheres*, 116(6).

Leung, L.R. et al., 1999. Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin. *Bulletin of the American Meteorological Society*, 80(11), pp.2313–2329.

Michaelsen, J., 1987. Cross-Validation in Statistical Climate Forecast Models. *Journal of Climate and Applied Meteorology*, 26(11), pp.1589–1600.

Min, Y.-M., Kryjov, V.N. & Park, C.-K., 2009. A Probabilistic Multimodel Ensemble Approach to Seasonal Prediction. *Weather and Forecasting*, 24(3), pp.812–828. Available at: http://journals.ametsoc.org/doi/abs/10.1175/2008WAF2222140.1.

Molteni, F. et al., 2011. *The new ECMWF seasonal forecast system (System 4)*,

Morse, A.P. et al., 2005. A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 57(3), pp.464–475.

Murphy, A.H., 1971. A Note on the Ranked Probability Score. *Journal of Applied Meteorology*, 10(1), pp.155–156.

Palmer, B.T.N., 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quarterly Journal of the Royal …*, 126, pp.2013–2033.

Palmer, T.N. et al., 2005. Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1463), pp.1991–1998.

Pan, J. & den Dool, H. Van, 1998. Extended-Range Probability Forecasts Based on Dynamical Model Output. *Wea. Forecasting*, 13, pp.983–996.

Pavan, V. & Doblas-Reyes, F.J., 2000. Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climate Dynamics*, 16(8), pp.611–625. Available at: http://link.springer.com/10.1007/s003820000063.

Peng, P. et al., 2002. An analysis of multimodel ensemble predictions for seasonal climate anomalies. *Journal of Geophysical Research Atmospheres*, 107(23), pp.1–12.

Siegert, S. et al., 2016. Detecting improvements in correlation forecast skill : Statistical tests and power analysis. *Monthly Weather Review*.

Steiger, J.H., 1980. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), pp.245–251.

Von Storch, H., 1999. On the use of “inflation” in statistical downscaling. *Journal of Climate*, 12(12), pp.3505–3506.

Torralba, V. et al., Development of a wind energy climate service based on seasonal climate prediction. Available at: https://upcommons.upc.edu/bitstream/handle/2117/90235/43-45 Development of a wind energy climate 3rd BSC International Doctoral Symposium 2016_1.pdf [Accessed March 13, 2017].

Torralba, V. et al., 2017. Seasonal climate prediction: a new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, p.JAMC-D-16-0204.1. Available at: http://journals.ametsoc.org/doi/10.1175/JAMC-D-16-0204.1 [Accessed March 13, 2017].

Troccoli, A. et al., 2013. Promoting new links between energy and meteorology. *Bulletin of the American Meteorological Society*, 94(4).

Troccoli, A. et al., 2010. Weather and climate risk management in the energy sector. *Bulletin of the American Meteorological Society*, 91(6), pp.785–788.

Vladislavleva, E. et al., 2013. Predicting the energy output of wind farms based on weather data: Important variables and their correlation. *Renewable Energy*, 50, pp.236–243. Available at: http://dx.doi.org/10.1016/j.renene.2012.06.036.

Voldoire, A. et al., 2013. The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, 40(9–10), pp.2091–2121.

Wang, B. et al., 2008. How accurately do coupled climate models predict the leading modes of Asian-Australian monsoon interannual variability? *Climate Dynamics*, 30(6), pp.605–619.

Weigel, A.P. et al., 2010. Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23(15), pp.4175–4191.

Weigel, A.P., Liniger, M.A. & Appenzeller, C., 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134(630), pp.241–260. Available at: http://doi.wiley.com/10.1002/qj.210 [Accessed October 28, 2016].

Weigel, A.P., Liniger, M.A. & Appenzeller, C., 2007. The Discrete Brier and Ranked Probability Skill Scores. *Monthly Weather Review*, 135(1), pp.118–124. Available at: http://journals.ametsoc.org/doi/abs/10.1175/MWR3280.1.

Wilks, D.S., 2006. *Statistical methods in the atmospheric sciences*,

Yun, W.T. et al., 2005. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 57(3), pp.280–289.

Zou, G.Y., 2007. Toward using confidence intervals to compare correlations. *Psychological Methods*, 12(4), pp.399–413.