THEME ENV.2012.6.1-1

EUPORIAS

(Grant Agreement 308291)

European Provision Of Regional Impact Assessment on a

Seasonal-to-decadal timescale

Deliverable D*32.1*

Report on assessment and combination of S2D predictions

EUPORIAS

Deliverable Title	*Report on assessment and combination of S2D predictions*	
Brief Description	*Results presented in this report will be based on results from WP21. Work carried out under this deliverable will be coordinated with Task 31.2.*	
WP number	32	
Lead Beneficiary	*UC*	
Contributors	*R. Manzanas, UC. J.M. Gutiérrez, UC* *J. Fernández, UC. M.D Frías, UC* *A.S. Cofiño, UC* *E. Sánchez, AEMET. J. Voces, AEMET* *E. Rodríguez, AEMET*	
Creation Date Version Number Version Date	16/4/2014 v3.0 1/8/2014	
Deliverable Due Date Actual Delivery Date	M18 (postponed to M21)	
Nature of the Deliverable	x	*R - Report*
		P - Prototype
		D - Demonstrator
		O - Other
Dissemination Level/ Audience	x	*PU - Public*
		PP - Restricted to other programme participants, including the Commission services
		RE - Restricted to a group specified by the consortium, including the Commission services
		CO - Confidential, only for members of the consortium, including the Commission services

Version	**Date**	**Modified by**	**Comments**
v1.0	16/4/2014	all authors	Initial draft.
v2.0	20/7/2014	UC	Updated draft, including downscaling.
V2.1	31/07/2014	Carlo Buontempo	Comment added and minor modifications made
V3.0	01/08/2014	UC	final version

Table of Contents

1. Executive Summary

Different EUPORIAS activities require global and regional (downscaled) predictions to build sector-specific indicators (WP22) and to feed different impact models (WP23) in agriculture, hydrology, energy, etc., among others. Since seasonal predictability strongly varies from region to region and from season to season, assessing the skill of state-of-the-art operational seasonal forecasting systems in the different regions of the world is a key task for these activities. This allows identifying the most suitable regions for impact applications on seasonal time-scales. Moreover, since several seasonal forecasting systems are available nowadays, it is also important to study the best approach to combine them into a single consensus prediction taking into account their performance on the target region and variable.

This report describes the validation of global European seasonal forecasting systems: The operational ECMWF System4 and four models of the ENSEMBLES multi-model hindcast, which correspond to former versions of the current operational seasonal forecasting systems, including System3 (the validation of other operational models will be performed following the same procedure when they become available). Moreover, an assessment of the added value of calibrated/downscaled forecasts —obtained by means of different bias-correction and statistical downscaling techniques— is also performed. Global forecasts are assessed worldwide, whereas downscaling techniques are analysed in two target regions representative of the low and high skill patterns found in the global models (Spain and Philippines, respectively). Finally, different forecast combination techniques —based on equal weighting, regression and Bayesian combination— are tested in the target regions, considering both global and regional predictions.

The most important findings are:

1) in agreement with previous studies, the skill of the global models analysed is mainly located in the tropics (particularly for precipitation) and no significant skill is found in Europe;
2) important seasonally varying biases are found in the global models, thus requiring bias-correction/downscaling to be suitable for impact applications;
3) statistical downscaling methods can improve global model outputs in some cases (particularly when the model is more skilful with the circulation variables than with the target one); however, bias-correction methods do not improve model skill;
4) No clear benefit is observed when applying model combination techniques (neither to the global model outputs nor to the downscaled values); thus, simple model combination obtained by equal weighting is recommended in general.

Finally, new operational models will be analysed when they become available through the ECOMS-UDG server, in order to explore the promising new high-resolution hindcasts with potential new sources of predictability (e.g. the NAO in the GloSea5 model, with 0.25º ocean and around 60 km atmosphere; see e.g. Scaiffe et al., 2014).

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOW, Section B1.1):

No.	Objective	Yes	No
1	Develop and deliver reliable and trusted impact prediction systems for a number of carefully selected case studies. These will provide working examples of end to end climate-to-impacts-decision making services operation on S2D timescales.		x
2	Assess and document key knowledge gaps and vulnerabilities of important sectors (e.g., water, energy, health, transport, agriculture, tourism), along with the needs of specific users within these sectors, through close collaboration with project stakeholders.		x
3	Develop a set of standard tools tailored to the needs of stakeholders for calibrating, downscaling, and modelling sector-specific impacts on S2D timescales.	x	
4	Develop techniques to map the meteorological variables from the prediction systems provided by the WMO GPCs (two of which (Met Office and MeteoFrance) are partners in the project) into variables which are directly relevant to the needs of specific stakeholders.		x
5	Develop a knowledge-sharing protocol necessary to promote the use of these technologies. This will include making uncertain information fit into the decision support systems used by stakeholders to take decisions on the S2D horizon. This objective will place Europe at the forefront of the implementation of the GFCS, through the GFCS's ambitions to develop climate services research, a climate services information system and a user interface platform.		x
6	Assess and document the current marketability of climate services in Europe and demonstrate how climate services on S2D time horizons can be made useful to end users.		x

3. Detailed Report

3.1. Assessment of Global Seasonal Forecasts

Seasonal predictability strongly varies from region to region and from season to season (see, e.g., van Oldenborgh, 2004; Doblas-Reyes et al., 2010). Therefore, a key task in EUPORIAS WP32 (uncertainty framework) is to globally assess the forecast skill of the seasonal forecasting models used in the different activities. The ECMWF System 4 seasonal forecasting system has been selected as the initial model to undertake this activity. The available System 4 hindcast is formed by 15 members covering the period 1981-2010, and several variables have already been included in the ECOMS-UDG server [1] — a joint EUPORIAS-SPECS initiative developed to provide homogeneous hindcast information from the different operational European seasonal forecasting systems to the ECOMS partners.— This task will be updated in the coming months to include the validation results for further seasonal forecasting systems to be included in the ECOMS-UDG (e.g. GloSea5 by Met Office). We focus on precipitation and surface temperatures, which are the most demanded variables in impact applications.

In order to compare results among different forecasting systems, we have also considered the state-of-the-art operational models from the EUROSIP initiative. However, due to data availability constraints only a limited number of studies have been performed by AEMET with this dataset. Therefore, we have also analysed four of the models included in the ENSEMBLES multi-model hindcast — the only homogeneous multi-model seasonal hindcast publicly available to date (Weisheimer et al., 2009).— These models (see Table 1) correspond to former versions of the current operational seasonal forecasting systems and, interestingly, the model from the ECMWF is System3, the predecessor of System4. The global validation of this multi-model hindcast has been performed in collaboration with SPECS, and the results have been already published in Manzanas et al. (2014b). In particular the global skill pattern found for precipitation (the only variable validated in Manzanas et al. (2014b)) is very similar to the one found when applying the same methodology to System4. Therefore, in this section we only describe the results for the latter.

Table 1: Main components of the four atmosphere-ocean coupled models considered in this work. All of them contributed to the ENSEMBLES multi-model seasonal hindcast.

Centre	Amospheric model / Resolution	Ocean model / Resolution
ECMWF	IFS CY31R1 (T159/L62)	HOPE (0.3º − 1.4º/L29)
IFM-GEOMAR	ECHAM5 (T63/L31)	MPI-OM1 (1.5º/L40)
CMCC-INGV	ECHAM5 (T63/L19)	OPA8.2 (2.0º/L31)

[1] The ECOMS User Data Gateway (ECOMS-UDG) is based on a THREDDS server which allows friendly data access, both directly and through an R-package for homogenous data access. More information in http://meteo.unican.es/ecoms-udg

Meteo France	ARPEGE4.6 (T63)	OPA8.2 (2.0º/L31)

Model assessment is performed gridbox by gridbox, calculating the bias, correlation and the Relative Operating Characteristic Skill Score (ROCSS) —a tercile-based probabilistic score which is not sensitive to model's bias and is recommended by the WMO Lead Centre for the SVS-LRF[2] for the verification of probabilistic seasonal forecasts.— In particular, one-month lead forecasts of precipitation and maximum temperature from System4 were validated for the standard seasons DJF, MAM, JJA and SON over the whole hindcast period (1981-2010). The 15-member ensemble mean was considered for bias and correlation, whereas the tercile probabilities given by the 15 members were considered for ROCSS. In the case of precipitation (maximum temperature), the model was validated against GPCC[3] v5 at 1º (CRU[4] ts 3.10 at 0.5º). In both cases, System4 output was bi-linearly interpolated from its native 0.75º to the resolution of the observations.

Figure 1 shows the bias (mean error) for precipitation (top) and maximum temperature (bottom). Wet/dry biases alternate in the different regions of the world for precipitation, whereas cold biases are predominant for maximum temperature (except in North America and East Asia).

Figure 2 shows the Pearson correlation (computed upon the inter-annual time-series). In agreement with previous studies, the highest correlations for precipitation are concentrated over tropical regions (see, e.g. van Oldenborgh et al., 2005). In the case of temperatures, trends can lead to artificial skill and, therefore, an alternative analysis for the de-trended series is given in Figure 3. In this case, skill spans over large regions of the globe.

Figure 4 shows the ROCSS for the dry (left) and wet (right) terciles of precipitation. In agreement with the results found in Manzanas et al. (2014b), the most skilful regions are Northern South America (all seasons), the Malay archipelago and Oceania (during JJA and SON), South Africa (during DJF) and Middle East (during SON). There are also some spots of skill in Africa, in the Gulf of Guinea (in DJF) and Sahel-East Africa (JJA and SON), but the spatial significance of these latter regions should be further analysed. Finally, Figure 5 shows the ROCSS for the cold (left) and warm (right) terciles of maximum temperature (de-trended series).

For the sake of simplicity, in the following sections we restrict the analysis to the case of precipitation.

---

2 http://www.bom.gov.au/wmo/lrfvs/users.shtml

3 ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_download.html

4 http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_ fe67d66a-5b02-11e0-88c9-00e081470265

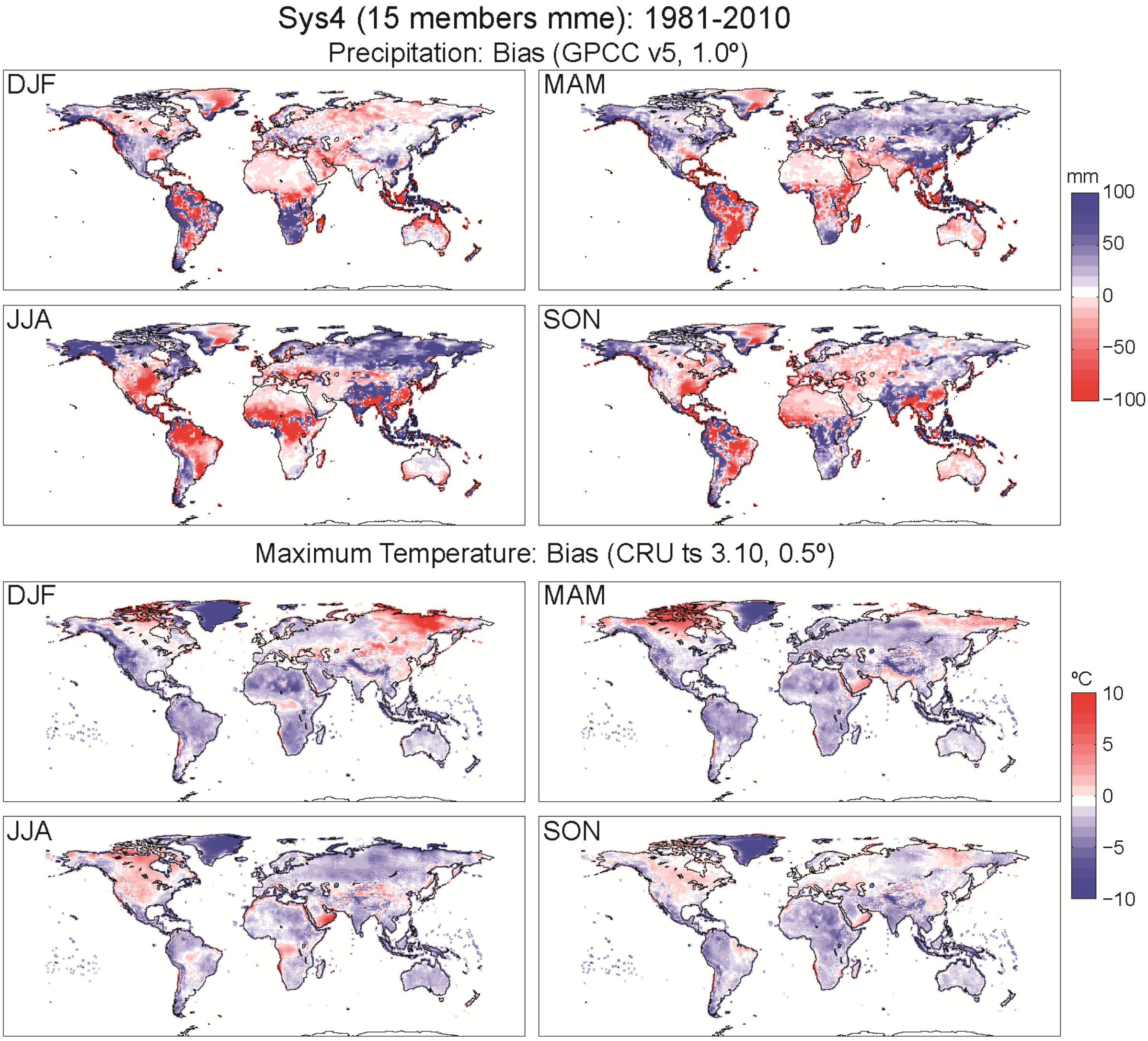

Figure 1: Biases for the 15-member ensemble mean for the whole hindcast period (1981-2000) for (top) precipitation and (bottom) maximum temperature

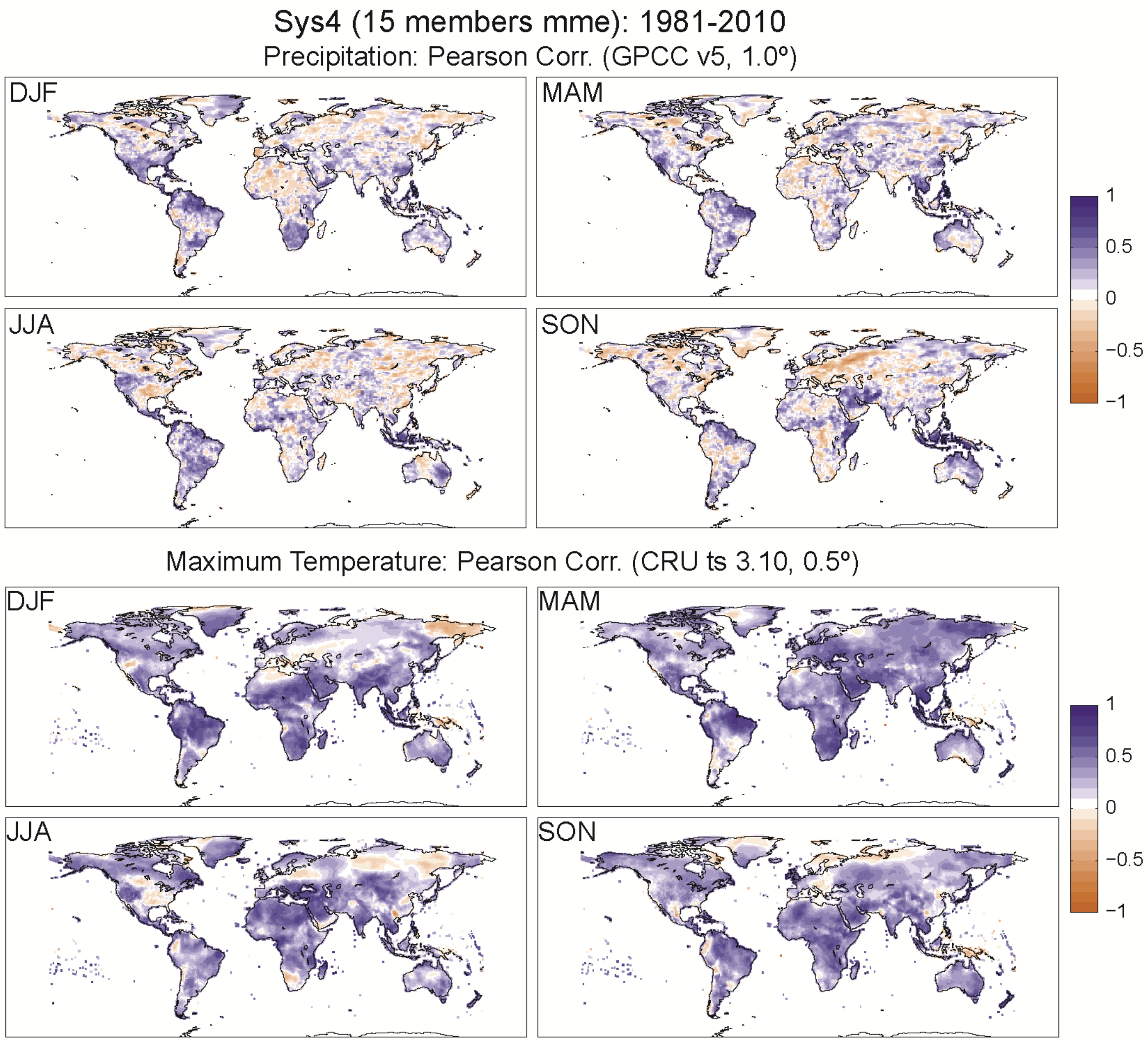

Figure 2: As Figure 1, but for Pearson correlation

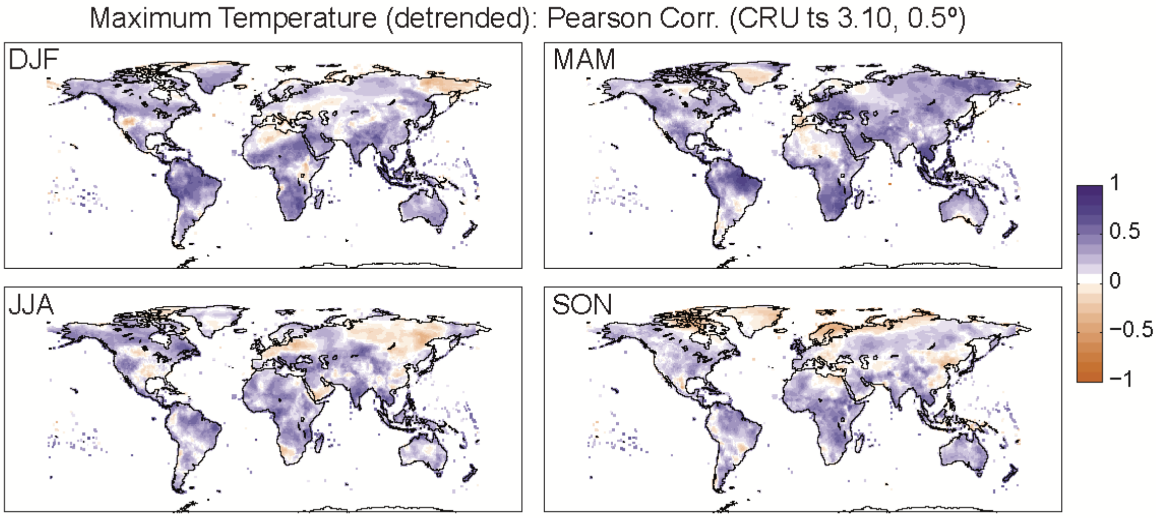

Figure 3: As Figure 2 (bottom), but for the de-trended series. Significant trends after a Mann-Kendall test (at a 5% level) are removed both from System4 and CRU TS 3.10 for the computation of correlations

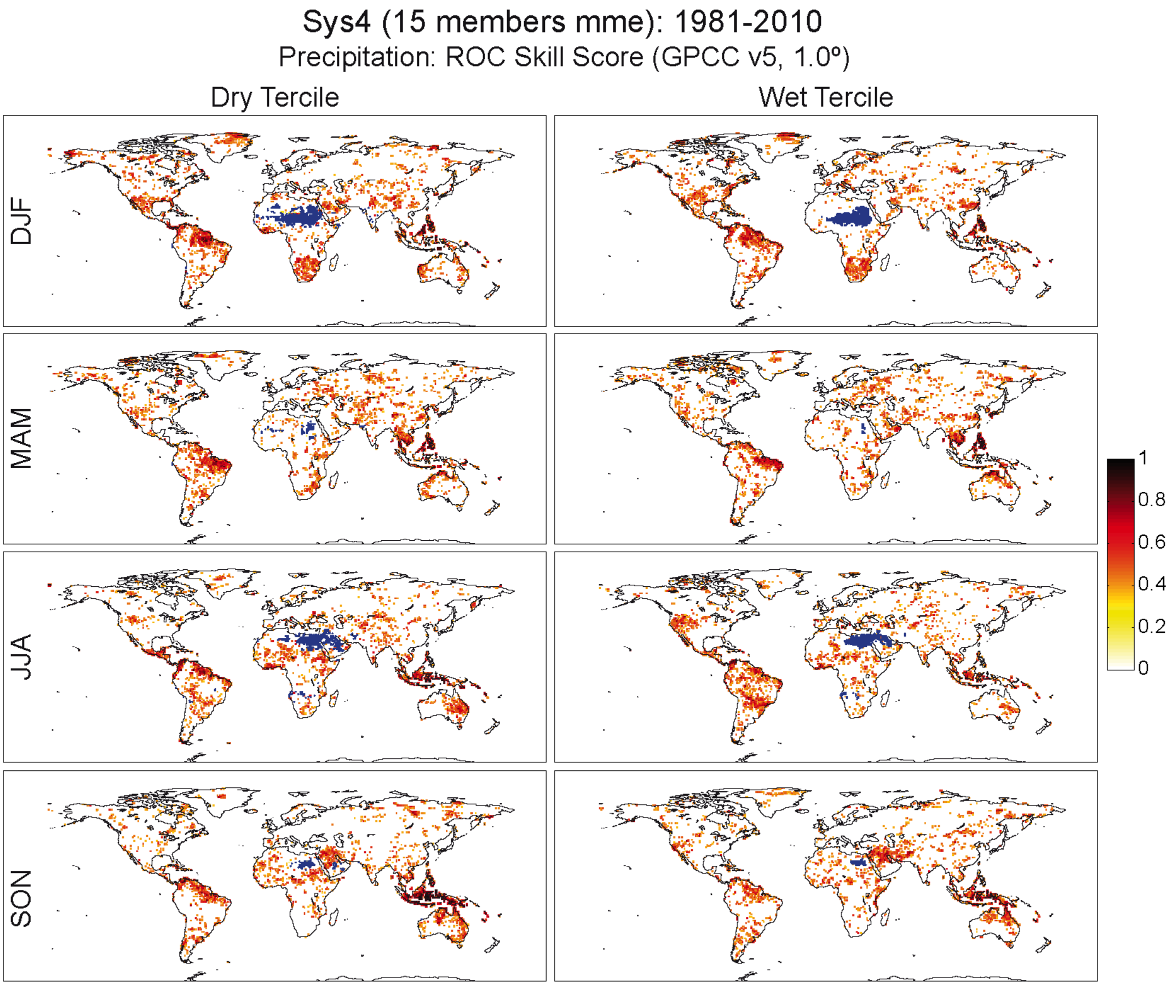

Figure 4: ROCSS for the dry and wet terciles of precipitation for the whole hindcast period 1981-2000. Only statistically significant (at a 5% level) ROCSS are shown. Blue points indicate grid-boxes where the corresponding tercile category has never been observed

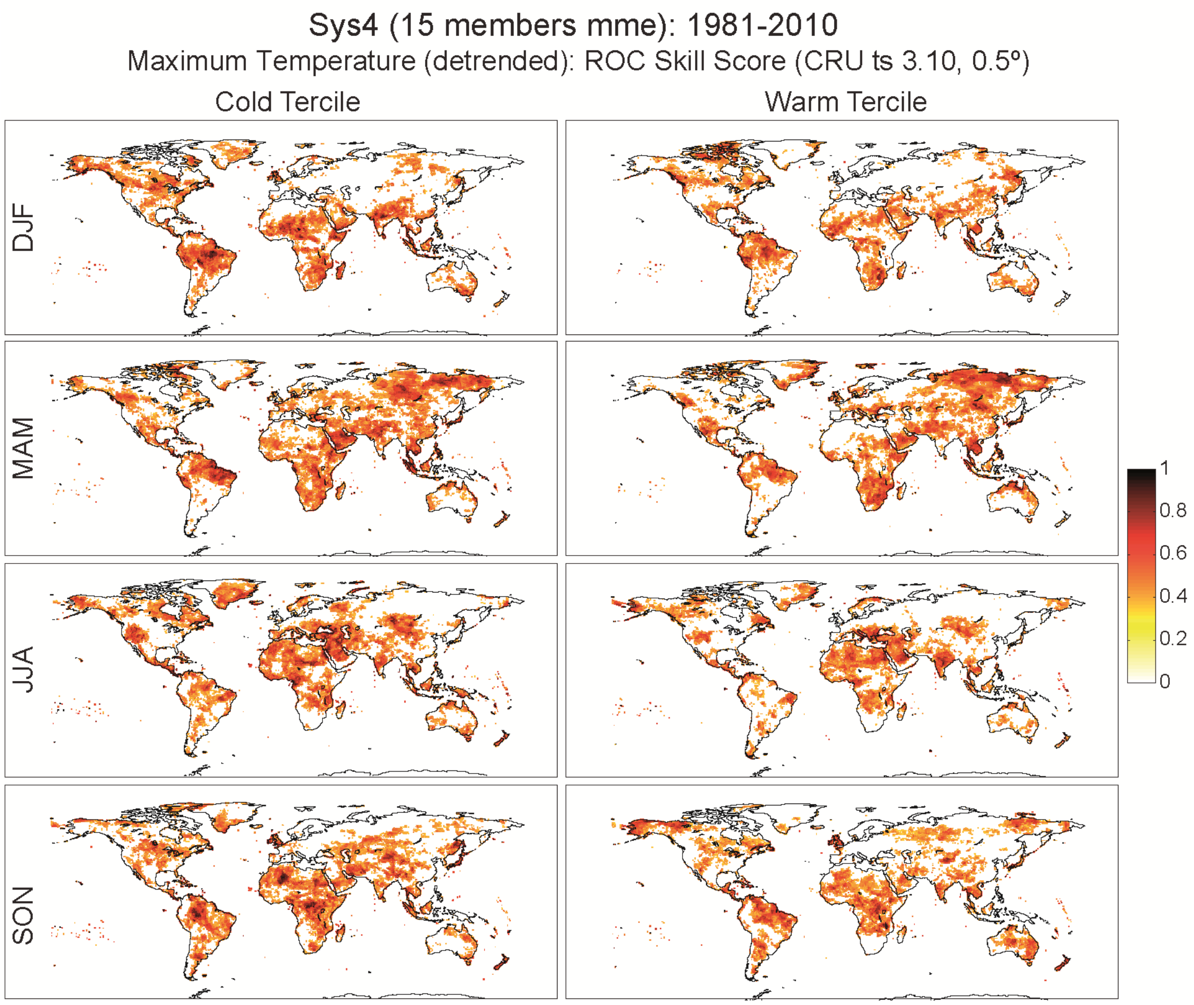

Figure 5: As Figure 4 but for detrended maximum temperature. Significant trends after a Mann-Kendall test (at a 5% level) are removed both from System4 and CRU ts 3.10 for the computation of the ROCSS

Finally, we want to remark that although no clear signals of skill have been found in the European region when analysing the System4 and ENSEMBLES models, promising results have been recently reported with more recent versions of seasonal forecasting (see e.g. Scaife et al., 2014). Therefore, work is in progress to collect the necessary information from some of these models (in particular GloSea5 from Met Office) and apply the same validation methodology used in this work in order to test the state-of-the-art skill in the European region.

3.2. Assessment of calibrated/downscaled seasonal forecasts

In this section we focus on precipitation and consider the multi-model ENSEMBLES hindcast (see Table 1). Each of the ENSEMBLES models in Table 1 was run four times a year —the first of February, May, August and November— for seven months,

considering an ensemble of nine initial conditions (nine members). Therefore, we analyse the standard seasons DJF, MAM, JJA and SON over the whole available hindcast period (1960-2000). We consider one-month lead predictions.

Seasonal forecasts of precipitation are calibrated/downscaled on a member basis and are subsequently compared with the Direct Model Output (DMO) in terms of their performance (interannual correlation and bias). Given the low/high skill found for this variable over European/tropical areas (see Figure 4), we considered two target areas in these regions (Spain and The Philippines, respectively) where high-quality observations were available. In both regions we considered 17 quality-controlled gauges (see Figure 6) with daily information during most of the ENSEMBLES multi-model hindcast period.

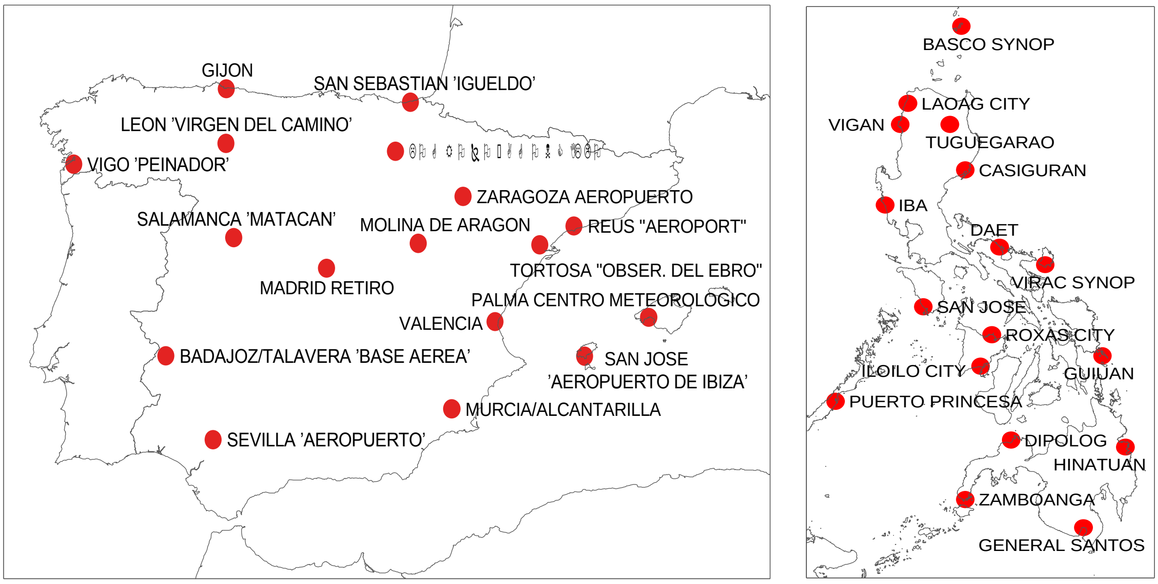

Figure 6: Location of the 17 gauges considered in each region (red points)

A screening of predictors and geographical domains has been already conducted in Spain (Manzanas et al., 2014c) and the Philippines (Manzanas et al., 2014a) in Perfect Prognosis conditions —i.e., using ERA-Interim (Dee et al., 2011) reanalysis data for calibration— obtaining the following optimum configurations: SLP, T850, Q850 (for Spain), and U850, U200, T850 and Q850 (for the Philippines), defined in the geographical areas shown in Figure 6 — SLP=sea level pressure, T=temperature, Q=specific humidity, U=meridional wind; numbers refer to vertical levels in hPa.— These predictor configurations are used in this work.

In order to explore the performance of different Statistical Downscaling Methods (SDMs), different techniques and configurations were applied and tested in the two target areas in collaboration with SPECS WP62 — the reader is referred to

deliverable D52.1[5] of SPECS for a review of the different types of techniques.— Finally, two different SDMs were considered in this deliverable.

On the one hand, SDM1 is based on Generalized Linear Models (GLM), an extension of linear regression which allows for non-normal predictand distributions (see Nelder and Wedderburn, 1972, for an introduction; Brandsma and Buishand, 1997; Fealy and Sweeney, 2007; Hertig et al., 2013). SDM1 follow the typical two-stage implementation used to model precipitation, consisting of a GLM with Bernoulli distribution and logit link for occurrence —equivalent to a logistic regression— and a GLM with gamma distribution and log link for the amount (see., e.g., Coe and Stern, 1982; Chandler and Wheater, 2002; Abaurrea and Asin, 2005). As predictors the 15 leading principal components (PCs; Preisendorfer, 1988) are used.

On the other hand, SDM2 is based on the analogue method (Lorenz, 1969; Zorita and von Storch, 1999), a popular non-parametric technique based on the assumption that similar local occurrences are expected for similar — as measured by the Euclidean distance here— atmospheric configurations (Wetterhall et al., 2005; Brands et al., 2011; Cubasch et al., 1996; Timbal et al., 2003; Moron et al., 2008; Timbal and Jones, 2008). In this work, we use the standard configuration using the closest analogue. The 30 leading PCs are considered as predictor data.

Similarly, different bias-correction techniques were applied. However, the results are very similar for all of them and, therefore, only results for the parametric method introduced by Piani et al. (2010) are shown in the following.

In the work, the different methods are calibrated using predictors from ERA-Interim (i.e., under perfect prognosis conditions) during the period 1981-2005. Afterwards, local predictions are obtained by applying the trained methods to the predictors given by the different seasonal forecasting models, member by member during the period 1960-2000.

Figures 7 and 8 show the results obtained for Spain and the Philippines, respectively. For each gauge, the performance of the direct model output (DMO) — the nearest gridbox in the model is considered—, the calibrated (Piani bias-corrected) model output (BCMO) and the downscaled predictions obtained from the two SDMs considered (SDM1 and SDM2) were computed. For each region, boxplots show the spread of the results along the 17 gauges. Each colour (excluding yellow, which corresponds to the multi-model results) corresponds to a different forecasting system from those in Table 1 (see the legend). Performance is measured in terms of Pearson correlation and bias (left and right columns, respectively), both computed upon the inter-annual time-series (note that the statistical downscaling is performed on a daily basis). It is important to notice that both the bias correction and the

[5] http://www.specs-fp7.eu/SPECS/Deliverable_files/SPECS_D52. 1.pdf

statistical downscaling are performed for each individual member. Then, for each model, the nine different calibrated/downscaled predictions (nine members) are averaged into a single prediction obtaining an inter-annual series which is then validated.

As can be seen from these figures, the mean error (bias) of the DMO is generally large for all models (especially in the Philippines, which might be due to a misrepresentation of the ocean-land interphase). In this case, results show that the BCMO yields nearly zero biases, whereas statistical downscaling techniques lead to small biases typically larger for SDM1 with opposite signs in some cases.

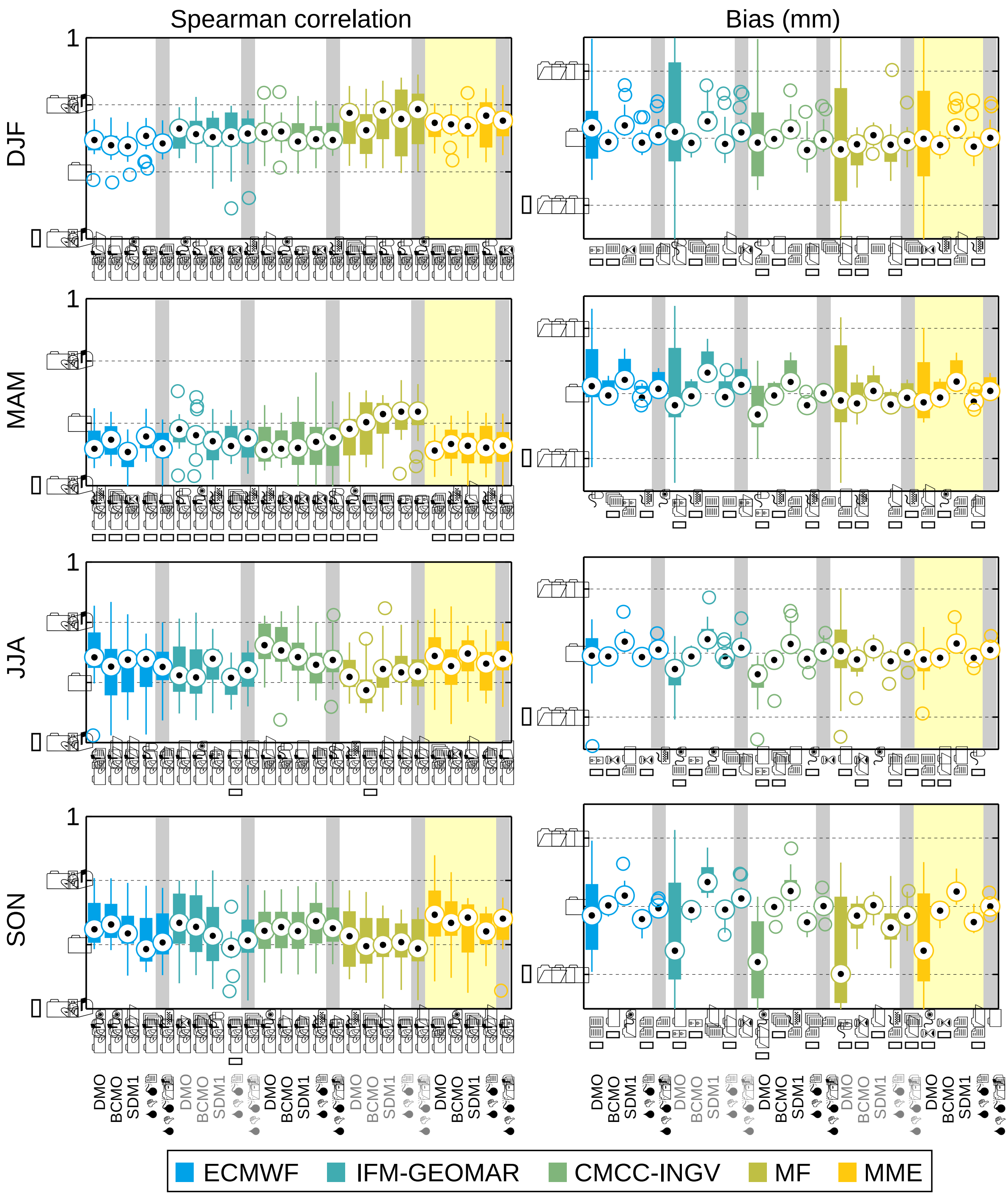

Figure 7: Performance of the raw/calibrated/downscaled seasonal forecasts for different models (see colors in legend) at the 17 gauges considered in Spain (see the text for details)

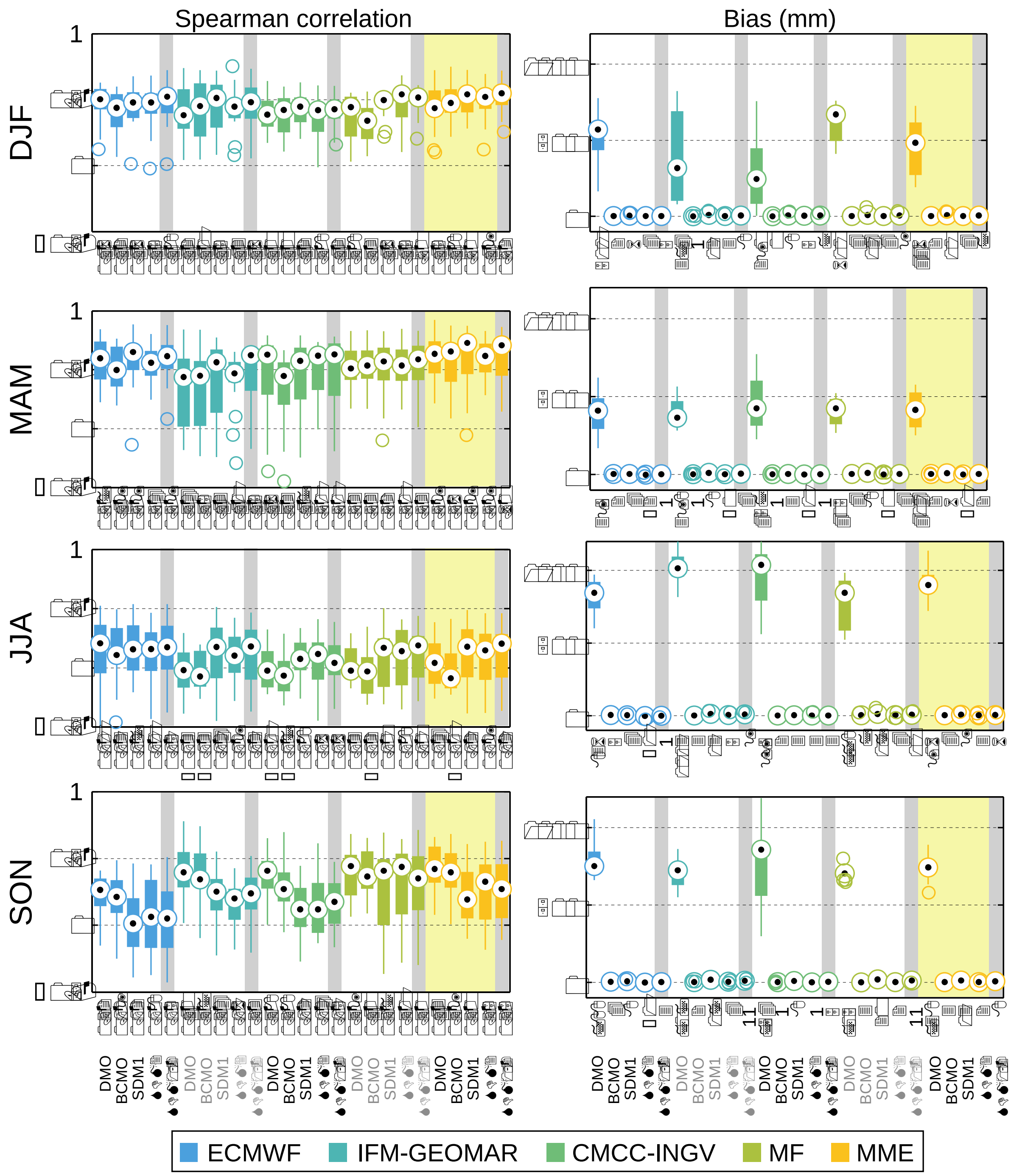

Figure 8: As Figure 7 but for the Philippines

The correlation results show that the BCMO does not improve the correlation of the model in any case, being smaller in some cases. For the case of the SDMs results are highly dependent on the location, season, model and method. Broadly speaking, Figs. 7 and 8 depict the three different situations than can occur:

1. The model exhibits no skill for precipitation and no skill for the large-scale circulation variables (SLP, T, U) —i.e. the model is not able to properly predict

the synoptic circulation — In this situation, SDMs will not provide any added value over the direct model output;

2. The model exhibits no skill for precipitation but it is skilful for the large-scale circulation. In this situation, if there is a physical link between the large- and the local-scales, a good SD method might capture this link and provide some added value to the model output. However, it might also occur that this link does not exist (convective precipitation). In the latter case, SD should not be expected to yield a significant added value;
3. The model exhibits skill for precipitation but not for the large-scale circulation. In this situation, SD is not appropriate.

Our results suggest that condition 1 is what predominantly occurs in Spain (and in Europe in general)[6]. However, a wider range of situations takes place in the Philippines, which is why this region is an ideal test-bed for this kind of study. In particular, condition 2 might be the case in DJF and MAM, when SD gives some added value with respect to the DMO. Note that SDMs also yield better results in JJA; however, the correlations reached in this season are still very low. Condition3 is particularly interesting and might be the case of SON in the Philippines, where SD techniques provide worse results than the DMO. It can be seen that bias-correction is more recommendable in this case.

The previous results indicate that the usefulness of calibrated/downscaled seasonal forecasts is limited by a number of restricting factors and, therefore, it should be assessed for each particular location/season/forecasting system/SDM.

3.3. Combination of multi-model seasonal forecasts

In general, when working with a multi-model ensemble of seasonal forecasting systems, one of the models will be preferable over the others for certain seasons, regions and target variables (see, e.g. Figures 7 and 8). Therefore, an important question in this context is finding the best way to automatically combine the model predictions in order to maximize the skill of the resulting consensus forecast. It has been demonstrated in several applications that the simple multi-model (denoted as MME and obtained by assigning equal weights to all models) improves the average performance of the individual models (see e.g. Doblas-Reyes et al. 2005, Wang et al. 2009, Weigel et al. 2010).

However, more sophisticated model combination approaches have been proposed to take into account some measure of individual skill, as calculated from the past performance in a hindcast dataset. Some of the most popular model combination

[6] Note: the most recent seasonal forecasting models are shown to be better than older models in describing larger scale patterns over Europe and, therefore, the added value of statistical downscaling methods may be relevant for those models.

techniques include forecast assimilation (Stephenson et al. 2005), based on a Bayesian combination of the model predictions, and different forms of linear regression techniques (Doblas-Reyes et al. 2005), including super-ensemble methods (Yu et al. 2005). Although the relative merits of these multi-model combinations over the individual models have been shown in particular applications (specially considering sea surface temperature), most of the inter-comparison studies undertaken so far concluded that it is difficult to improve the MME forecasts (DelSole et al. 2012; Rodrigues et al. 2014). In particular, Rodrigues et al. (2014) conclude that further research is needed to analyse how combination methods will behave if applied to less predictable regions such as the extratropics or to other variables such as air temperature or precipitation.

In this section we consider both the direct model and downscaled precipitation outputs in Spain and the Philippines computed in previous sections and analyse the performance of the most popular model combination approaches.

3.3.1 Simple multi-model and multi-SDM approaches

Yellow (shadowed) boxplots in Figures 7 and 8 correspond to the simple multi-model (MME) constructed by applying equal weights to all contributing models and members, i.e., by averaging the 4 (models) × 9 (members) = 36 predictions into a single prediction which was then validated. In addition, for each model (and for the MME), we also assessed the potential of an ensemble of SDMs (grey-shadowed boxplots). To this, predictions from SDM1 and SDM2 were averaged into a single prediction (SDM1-2), which was then validated.

From these figures it can be shown that, in agreement with previous results, the MME yields generally slightly better results than any of the individual models. Moreover, the MME obtained by combining the two SDMs is also generally better than the individual SDMs, particularly in skilful seasons and regions (e.g. DJF and MAM in the Philippines). Therefore, the simple multi-model provides a successful benchmark.

3.3.2 Model combination techniques: Regression and Bayesian combination

In this section, we analyse the added value (with regards to the simple model combination, MME) of two representative model combination approaches: forecast assimilation (Stephenson et al. 2005; using the R package provided by the authors) and model weighting by least-square multiple linear regression of the observations on the anomaly values of the four forecast systems (see Doblas-Reyes et al. 2005). In all cases, a Leave-One-Out (LOO) cross-validation approach is performed, so the final combined series is obtained by joining the test results for each individual year

(the weight model is build in each case using the interannual series, excluding the test year).

For the sake of simplicity we focus on the DJF season, both in Spain and the Philippines, where both global models and downscaling methods exhibit varying performances. In general, both model combination techniques failed to improve the performance (interannual correlation) of the simple multi-model both for the raw model outputs and for the bias-corrected and downscaled results. Each of the panels in Figure 9 shows in the first four boxplots the results over the 17 stations of the DMO, for the four individual models. These figures also show the results of two multi-model combinations: simple multi-model (equal weights) and forecast assimilation (FA). These techniques are applied to the direct model outputs (DMO), the bias corrected outputs (BCMO), the two statistical downscaling methods (SDM1, SDM2), the ensemble formed by the two SDMs (SDM1-2) and the full ensemble (AlLL), including all the previous information. Each of these blocks is bounded by grey shade (the first boxplot corresponds to the simple model combination and the second to the FA method). This figure shows that, in general the simple multimodel approach is a convenient choice and no further improvement is possible in general with the alternative more complex techniques. Moreover, the performance of the model combination techniques seems to be higher over skilful regions (e.g. the example of Philippines) than over non-skilful ones (e.g. in Spain).

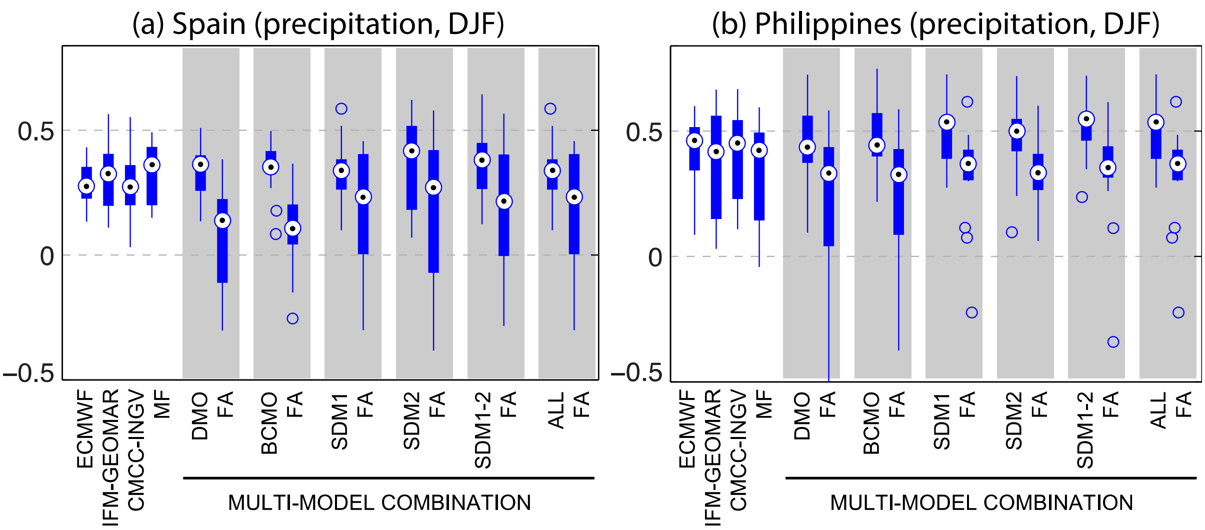

Figure 9: Performance (Pearson interannual correlation) of the direct model outputs from the four ENSEMBLES models and for the simple multi-model and the forecast assimilation obtained for the raw/calibrated/downscaled seasonal at the 17 gauges considered in Spain (see the text for details)

In order to further analyze this deficiency of the multi-model combination techniques, we performed the same experiment considering a stepwise linear regression for weighting each of the four models. We found that in most of the cases (particularly in Spain) the LOO cross-validated regression models did only include the predictors of

one of the models, indicating either a redundancy of the information and/or a lack of skill in the models.

Finally, we run the linear regression models with no cross-validation and obtained very high correlation values (over 0.9 in most of the cases), indicating a clear overfitting and a lack of generalization in this case. Therefore, simple model combination using equal weighting is preferable in those cases with no previous knowledge of the problem. This simple technique would be particularly appropriate if the global models are carefully assessed in order to remove bad performing models which do not provide any added value to the ensemble.

Table 2: Number of times (out of a total of 25, the number of years entering in the LOO cross-validation) that each one of the four global model outputs enter in the weighting linear regression for each of the stations (in rows) for Spain (left) and the Philippines (right). A stepwise regression has been used in this case, with significances 0.01 and 0.05 for entering and leaving the model, respectively. Note that a vector (25,25,25,25) would indicate that all global models have been used in all cases.

Spain (precipitation, DJF)			
ECMWF	IFM	CMCC	MF
0	23	0	0
0	0	0	0
0	0	1	22
0	0	1	0
0	0	23	0
0	0	1	0
0	1	0	18
0	0	0	24
0	0	0	1
1	0	23	0
14	4	0	0
0	22	0	0
0	4	0	18
0	0	0	3
0	24	0	0
0	0	24	0
0	6	0	0

Philippines (precipitation, DJF)			
ECMWF	IFM	CMCC	MF
25	3	0	1
0	0	0	0
0	0	0	0
0	0	10	15
0	0	0	0
0	0	1	25
1	24	1	1
0	2	0	23
0	22	0	4
1	10	16	0
1	0	5	0
1	24	0	0
0	0	26	0
0	26	0	0
0	11	15	0
0	8	16	0
1	1	6	0

3.3.3 Model assessment and combination with EUROSIP data

AEMET has focused its activity for this deliverable on estimating the skill of state-of-the-art operational seasonal forecast models. Direct precipitation and temperature outputs from four models —ECMWF System4, Meteo-France System 3, Met Office System3 and National Center for Environmental Prediction CFSv2— were available

for AEMET from the EUROSIP consortium for the common period 1988-2008 period and have been downloaded and homogenized over the Iberian Peninsula and other Mediterranean regions (France, Italy and the Balkans). Model combination using the forecast assimilation approach (Stephenson et al. 2005) has been also performed. Note that this work complements the assessment studies reported in previous sections considering a state-of-the-art hindcast of operational seasonal forecasting systems over Spain.

The E-OBS gridded dataset has been used as observational data (Haylock et al. 2008). The resulting anomaly values have been upscaled to 1º x 1º horizontal resolution. The multi-model forecast assimilated results have been computed considering LOO cross-validated series, using the same software of the previous sections. A set of deterministic and probabilistic verification scores has been computed (RPSS, ROC area and BSS) additionally to deterministic scores. Also statistical significance of the results has been quantified with a p value estimated using a nonparametric bootstrap method.

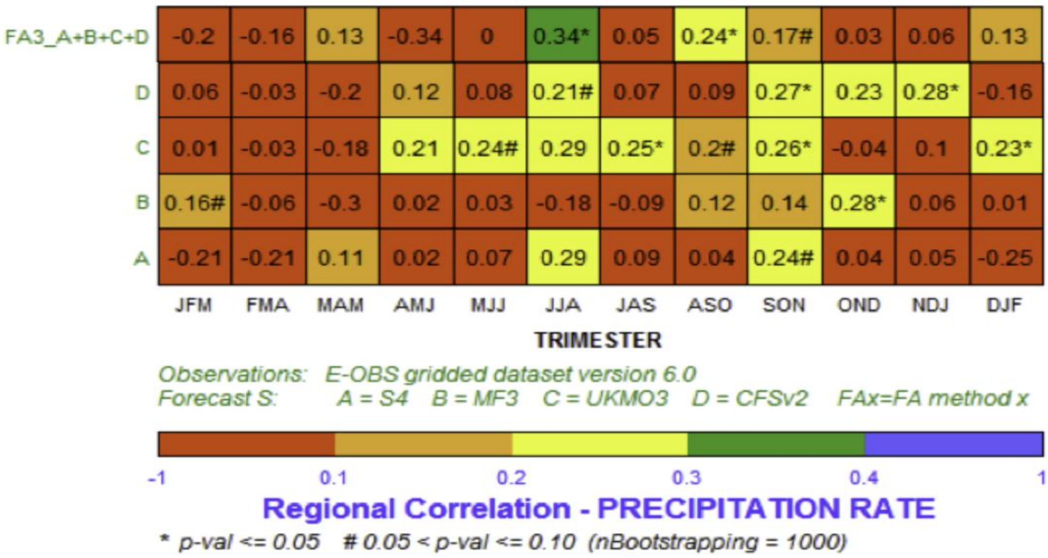

Figure 10: Pearson correlation coefficient computed for the Iberian Peninsula domain using: direct output seasonal forecasts from four models (A: ECMWF System4, B: MF3, C: UKMO3, D: CFSv2) and forecast assimilated (FA) outputs (first row). Significant values are also estimated (* p-value 0.05, # p-value 0.10)

Figure 10 shows an example of the results obtained from this study for the case of precipitation (extended information of this work is available in the EUPORIAS wiki[7]). In this case Pearson correlation coefficients for one-month lead-time seasonal precipitation forecasts in Iberia are displayed for the four models and for the forecast

---

[7] https://euporias.wikidot.com/local--files/wp32-meeting20140401/WP32_Progress_report_AEMET_Apr2014.pdf

assimilated multi-model. From this figure it can be seen that, in general, the model combination technique does not improve the performance of the individual models, with the exception of the summer months. Note that, as opposite to the previous sections, in this case the validation is performed at a regional level and, therefore, the aggregation can influence the validation scores (by increasing or decreasing the signal, if any).

Therefore, no clear benefit has been obtained when using forecast assimilation as compared to the individual model outputs and to the simple multi-model combination.

3.4 Peer reviewed articles and planned future publications

Currently, two publications (listed below) have been produced with partial results from this deliverable, both in collaboration with SPECS. Moreover, there are plans to submit two additional publications during the coming months: 1) results from the forecast assimilation in Spain using EUROSIP hindcasts (led by AEMET); 2) validation of the downscaled results and model combination (led by UC).

MANZANAS, R. AND FRIAS, M. D. AND COFIÑO, A. S. AND GUTIERREZ, J. M., 2014b: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill, Journal of Geophysical Research: Atmospheres, 119, 1708–1719, doi:10.1002/2013JD020680

MANZANAS, R. AND BRANDS, S. AND SAN MARTÍN, D. AND LUCERO, A. AND LIMBO, C. AND GUTIE´RREZ, J. M., 2014a: Statistical Downscaling in the Tropics is Sensitive to Reanalysis Choice, Under review in Journal of Climate.

3.5 References

ABAURREA, J AND ASIN, J, 2005: Forecasting local daily precipitation patterns in a climate change scenario, CLIMATE RESEARCH, 28, 183–197.

BRANDS, S. AND HERRERA, S. AND SAN-MARTIN, D. AND GUTIERREZ, J. M., 2011: Validation of the ENSEMBLES global climate models over southwestern Europe using probability density functions, from a downscaling perspective, Cli- mate Research, 48, 145–161, doi:10.3354/cr00995.

BRANDSMA, T AND BUISHAND, TA, 1997: Statistical linkage of daily precipitation in Switzerland to atmospheric circulation and temperature, JOURNAL OF HYDROLOGY, 198, 98–123.

CHANDLER, RE AND WHEATER, HS, 2002: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, WATER RESOURCES RESEARCH, 38, 38, doi:{10.1029/2001WR000906}.

COE, R. AND STERN, R. D., 1982: Fitting models to daily rainfall data, Journal of Applied Meteorology, 21, 1024–1031, doi:10.1175/1520-0450(1982)021⟨ 1024:FMTDRD⟩ 2.0.CO;2.

CUBASCH, U AND VONSTORCH, H AND WASZKEWITZ, J AND ZORITA, E, 1996: Estimates of climate change in Southern Europe derived from dynamical climate model output, CLIMATE RESEARCH, 7, 129–149.

DEE, D. P. AND UPPALA, S. M. AND SIMMONS, A. J. AND BERRISFORD, P. AND POLI, P. AND KOBAYASHI, S. AND ANDRAE, U. AND BALMASEDA, M. A. AND BALSAMO, G. AND BAUER, P. AND BECHTOLD, P. AND BELJAARS, A. C. M. AND VAN DE BERG, L. AND BIDLOT, J. AND BORMANN, N. AND DELSOL, C. AND DRAGANI, R. AND FUENTES, M. AND GEER, A. J. AND HAIMBERGER, L. AND HEALY, S. B. AND HERSBACH, H. AND HO´ LM, E. V. AND ISAKSEN, L. AND KA˚LLBERG, P. AND KO˙HLER, M. AND MATRICARDI, M. AND MCNALLY, A. P. AND MONGE-SANZ, B. M. AND MORCRETTE, J.J. AND PARK, B.K. AND PEUBEY, C. AND DE ROSNAY, P. AND TAVOLATO, C. AND THE´PAUT, J.-N. AND VITART, F., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. R. Meteorol. Soc., 137, 553–597.

DELSOLE T, YANG X, TIPPETT MK (2012) Is unequal weighting significantly better than equal weighting for multi-model fore- casting? Q J R Meteorol Soc 139:176–183

DOBLAS-REYES FJ, HAGEDORN R, PALMER TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. Tellus A 57:234–252

DOBLAS-REYES, F. J. AND WEISHEIMER, A. AND PALMER, T. N. AND MURPHY, J. M. AND SMITH, D.: Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts, Tech. Rep. 621, 2010.

FEALY, ROWAN AND SWEENEY, JOHN, 2007: Statistical downscaling of precipitation for a selection of sites in Ireland em- ploying a generalised linear modelling approach, Int. J. Climatol., 27, 2083–2094, doi:10.1002/joc.1506.

HAYLOCK MR, HOFSTRA N, KLEIN TANK AMG, KLOK EJ, JONES PD, NEW M, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res 113:D20119. doi:10.1029/2008JD010201

HERTIG, E. AND SEUBERT, S. AND PAXIAN, A. AND VOGT, G. AND PAETH, H. AND JACOBEIT, J., 2013: Changes of total versus extreme precipitation and dry periods until the end of the twenty-first century: statistical assessments for the Mediterranean area, Theoretical And Applied Climatology, 111, 1–20, doi:10.1007/s00704-012-0639-5.

LORENZ, EN, 1969: ATMOSPHERIC PREDICTABILITY AS REVEALED BY NATURALLY OCCURRING ANA- LOGUES, JOURNAL OF THE ATMOSPHERIC SCIENCES, 26, 636–&.

MANZANAS, R. AND BRANDS, S. AND SAN MARTIN, D. AND LUCERO, A. AND LIMBO, C. AND GUTIERREZ, J. M., 2014a: Statistical Downscaling in the Tropics is Sensitive to Reanalysis Choice, Submitted to Journal of Climate.

MANZANAS, R. AND FRIAS, M. D. AND COFIÑO, A. S. AND GUTIERREZ, J. M., 2014b: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill, Journal of Geophysical Research: Atmospheres, 119, 1708–1719, doi:10.1002/2013JD020680.

MANZANAS, R. AND SAN MARTIN, D. AND HERRERA, S. AND BRANDS, S. AND GUTIERREZ, J. M., 2014c: Regional projection of daily precipitation with an ensemble of statistical downscaling models: Partitioning the uncertainty sources, Preparing submission to Journal of Climate.

MORON, VINCENT AND ROBERTSON, ANDREW W. AND WARD, M. NEIL AND NDIAYE, OUSMANE, 2008: Weather types and rainfall over Senegal. part II: Downscaling of GCM simulations, J. Climate, 21, 288–307, doi:{10.1175/2007JCLI1624. 1}.

NELDER, JA AND WEDDERBURN, RWM, 1972: Generalized Linear Models, JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-GENERAL, 135, 370–&.

PIANI, C. AND HAERTER, J.O. AND COPPOLA, E., 2010: Statistical bias correction for daily precipitation in regional climate models over Europe, Theoretical and Applied Climatology, 99, 187–192, doi:10.1007/s00704-009-0134-9.

R.W. PREISENDORFER, 1988: Principal component analysis in meteorology and oceanography, Amsterdam, Elsevier, 1 edn. TEUTSCHBEIN, CLAUDIA AND WETTERHALL, FREDRIK AND SEIBERT, JAN, 2011: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale, Climate Dynamics, 37, 2087–2105, doi:10.1007/s00382-010-0979-8.

L.R.L RODRIGUES, F.J. DOBLAS-REYES, C.A.S. COELHO, 2014: Multi-model calibration and combination of tropical seasonal sea surface temperature forecasts. Clim Dyn, 42:597–616 DOI 10.1007/s00382-013-1779-8

SCAIFFE, A.A. et al., 2014: Skillful long-range prediction of European and North American winters. Geophysical Research Letters, 41, doi:10.1002/2014GL059637

DB STEPHENSON, CAS COELHO, FJ DOBLAS-REYES (2005) “Forecast assimilation: a unified framework for the combination of multi- model weather and climate predictions” Tellus A57:253-264. DOI: 10.1111/j.1600-0870.2005.00110.x

TIMBAL, B. AND JONES, D. A., 2008: Future projections of winter rainfall in southeast Australia using a statistical downscaling technique, CLIMATIC CHANGE, 86, 165–187, doi:10.1007/s10584-007-9279-7.

TIMBAL, B AND DUFOUR, A AND MCAVANEY, B, 2003: An estimate of future climate change for western France using a statistical downscaling technique, CLIMATE DYNAMICS, 20, 807–823, doi:10.1007/s00382-002-0298-9.

VAN OLDENBORGH, G. J.: Assessing the skill of seasonal forecasts, Tech. rep., KNMI Research Biennial Reports, 2003/04. KNMI, 2004.

VAN OLDENBORGH, G. J. AND BALMASEDA, M. A. AND FERRANTI, L. AND STOCKDALE, T. N. AND ANDERSON, D. L. T., 2005: Evaluation of atmospheric fields from

the ECMWF seasonal forecasts over a 15 year period, Journal of Climate, 18, 3250–3269, doi:10.1175/JCLI3421.1.

WANG B, LI J-Y, KANG I-S, SHUKLA J, PARK C-K, KUMAR A, SCHEMM J, COCKE S, KUG J-S, LUO J-J, FU X, YUN W-T, ALVES O, JIN E, KINTER J, KIRTMAN B, KRISHNAMURTI T, LAU N, LAU W, LIU P, PEGION P, ROSATI T, SCHUBERT S, STERN W, SUAREZ M, YAMAGATE T (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14 model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn 33:93–117

WEIGEL, A.P., R KNUTTI, M.A. LINIGER, C. APPENZELLER, 2010: Risks of Model Weighting in Multimodel Climate Projections. J. Climate, 23, 4175–4191. doi: http://dx.doi.org/10.1175/2010JCLI3594.1

WEISHEIMER, A. AND DOBLAS-REYES, F. J. AND PALMER, T. N. AND ALESSANDRI, A. AND ARRIBAS, A. AND DEQUE, M. AND KEENLYSIDE, N. AND MACVEAN, M. AND NAVARRA, A. AND ROGEL, P., 2009: ENSEMBLES: A new multi- model ensemble for seasonal-to-annual prediction. Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, Geophysical Research Letters, 36, 36, doi:10.1029/2009GL040896.

WETTERHALL, F AND HALLDIN, S AND XU, CY, 2005: Statistical precipitation downscaling in central Sweden with the analogue method, JOURNAL OF HYDROLOGY, 306, 174–190, doi:10.1016/j,jhydrol.2004.09.008.

YUN WT, STEFANOVA L, MITRA AK, KUMAR VV, DEWAR W, KRISHNAMURTI TN (2005) A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus 57A:280–289

ZORITA, E AND VON STORCH, H, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, JOURNAL OF CLIMATE, 12, 2474–2489.

4. Lessons Learnt

The most important lessons learnt from this deliverable are related to the most important findings: 1) In agreement with previous studies, the skill of the global models analysed is mainly located in the tropics (particularly for precipitation) and no significant skill is found in Europe. 2) Important seasonally varying biases are found in the global models, thus requiring bias-correction/downscaling to be suitable for impact applications. 3) Statistical downscaling methods can improve global model outputs in some cases (particularly when the model is more skillful with the circulation variables than with the target one); however, bias-correction methods do not improve model skill. 4) No clear benefit is observed when applying model combination techniques (neither to the global model outputs nor to the downscaled values); thus, simple model combination obtained by equal weighting is recommended in general.

5. Links Built

Parts of the work described in this deliverable have been done in collaboration with FP7 SPECS, in particular with WP62 (downscaling). Also, of importance it has been the collaboration with the data management tasks relative to the ECOMS-UDG server in order to automate the skill assessment work, so new models can be validated applying the same methodology when they become available.

Links have also been established with other European initiatives, such as the VALUE CORDEX action.

6. Versions and deviations from the initial plan

A first version of the deliverable was prepared in April 2014. However, the analyses of model combination undertaken by AEMET (using forecast assimilation) in European regions (the most relevant for EUPORIAS) showed very low seasonal prediction skill and, therefore, it didn't show an optimal performance. UC has been preparing statistical downscaling methodologies (bias correction and perfect prog), prioritizing bias correction to early meet other EUPORIAS users' needs and working in collaboration with SPECS to speed up this task. Some of our downscaling experiments (considering the ENSEMBLES multi-model hindcast) have been applied to tropical regions, where the skill is much higher than over Europe. In order to provide a complete report on the assessment and combination of seasonal predictions, a 3-months delay was approved by the EC project officer and a new submission deadline was agreed for 31st July 2014.