THEME ENV.2012.6.1-1

EUPORIAS

(Grant Agreement 308291)

European Provision Of Regional Impact Assessment on a

Seasonal-to-decadal timescale

Deliverable D21.1

Report on the skill of downscaled seasonal hindcasts

Deliverable Title	*Report on the skill of downscaled seasonal hindcasts*	
Brief Description	*Report on the skill of downscaled seasonal hindcasts as input to impact models and case studies over Europe*	
WP number	21	
Lead Beneficiary	*Jesús Fernández (UC)*	
Contributors	*Thomas Bosshard, Peter Berg, Ilias Pechlivanidis, Kean Foster (SMHI)* *Christian Viel, Anne-Lise Beaulant, Mathieu Papazzoni (MF)* *Laurent Dubus (EDF), Christiana Photiadou (KNMI)* *Jesús Fernández, Ana Casanueva, M Magariño, J. M. Gutiérrez (UC)*	
Creation Date Version Number Version Date	26-Feb-2016 1 29-Apr-2016	
Deliverable Due Date Actual Delivery Date	30-Apr-2016 29-Apr-2016	
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
0	04-Mar-2016	J. Fernández (UC)	Initial draft merging input from partners
1	31-Mar-2016	Christiana Photiadou (KNMI), Christian Viel (MF), Thomas Bosshard (SMHI), J Fernández, A Casanueva (UC)	Combining the feedback from several groups. Conclusions and Executive summary. First complete version of the document.
2	25-Apr-2016	Thomas Bosshard (SMHI), Laurent Dubus (EDF), J Fernández (UC)	Final changes prior to delivery to the management board.
3	29-Apr-2016	G. Nikulin (SMHI)	Small format changes

Table of Contents

1 Executive Summary 5
2 Objectives 6
2.1 Project Objectives 6
2.2 Work Package objectives 7
3 Detailed Report 8
3.1 Introduction 8
3.2 Bias correction 9
3.2.1 Pan-European bias correction for flood risk estimation using E-HYPE 10
3.2.2 SOSRhine case study 12
3.3 Statistical Downscaling 14
3.3.1 Flood volumes for hydropower in Swedish catchments 14
3.3.2 River flow in Southeastern France 15
3.3.3 Downscaling for the River Flow Forecast (RIFF) prototype 17
3.3.4 Temperature forecasts over Southern Italy and Southwest UK 18
3.4 Summary and conclusions 25
3.5 References 25
4 Lessons Learnt 27
5 Links Built 27

List of Tables

Table 1: Summary of the calibration and downscaling studies carried out in this Deliverable. See text for acronyms. 5

List of Figures

Figure 1: Mean monthly bias in temperature for the seasonal forecasts initialized in April and looking at the 3 months lead time. Shown are the bias in the uncorrected (left) and DBS corrected (right) data for the calibration period 1996-2010 (top) and the validation period 1981-1995 (bottom). 10
Figure 2: Mean monthly bias in precipitation for the seasonal forecasts initialized in April and looking at the 3 months lead time. Shown are the bias in the uncorrected (left) and DBS corrected (right) data for the calibration period 1996-2010 (top) and the validation period 1981-1995 (bottom). 11
Figure 3: Focusing on the Rhine basin: Mean monthly bias of ECMWF S4 for precipitation forecast initialized in June with lead time 4 months for 1981-2010 for (a) uncorrected forecast and (b) bias corrected forecast. 13
Figure 4: Mean monthly bias of ECMWF S4 for temperature forecast initialized in June with lead time 4 months for 1981-2010 for (a) uncorrected forecast and (b) bias corrected forecast. 13
Figure 5 (left) Domain of the large scale circulation variables used as predictors in the statistical downscaling and (right) location of the 84 sub-basins to which the LSCV were downscaled to. 15
Figure 6: Localization of the 33 watersheds. 15
Figure 7: Study description. 16
Figure 8: Tercile plot of T2m forecasts at Serre-Ponçon Watershed, Init date: November, Roc Skill Score. 17
Figure 9: ARPEGE-S5 daily Precipitation forecast, deciles 7, 9 and 10 (maximum), at 45.75N/8.25E (lat/lon). Symbols ”+” represent raw quantile data, black lines represent a 30-day running mean, red lines represent a LOESS-fitting. 18
Figure 10: Winter (DJF) minimum temperature biases (left) and interannual correlations (right) for the ECMWF System 4 raw output, this output interpolated (nearset neighbour) to the E-OBS grid (0.25º), downscaled by means of analogues and linear regression, and bias corrected by empirical quantile mapping (sequentially in rows). All downscaled output has been produced using a 5-fold random cross-validation. The observational reference is E-OBS in all cases. 20
Figure 11: As Figure 10, but for summer (JJA) mean temperature over southern Italy. The additional last column represents interannual correlation after detrending the series. 21
Figure 12: Tercile plots showing the probability of each temperature tercile forecast by each system (shades) along with the observed tercile (red dots). The upper panels show the original series, while the lower ones are detrended. 23
Figure 13: Reliability diagram for each forecast system. 24

1 Executive Summary

This deliverable focuses on the statistical downscaling technique to bridge the gap between coarse global climate model seasonal forecasts and local surface observations, required as input to impact models. Different groups have contributed to different climate service prototypes and case studies, which focus on different regions and target variables (Table 1). This report summarises their contributions.

Two main approaches have been followed: Perfect-prog statistical downscaling and bias correction (distribution-based model output statistics). The latter cannot improve the skill of a forecast, but only the statistical properties of the target variables to be fed to impact models. The former has potential to improve the skill, but results show that, over Europe, perfect-prog performs quite similarly to simpler bias correction methods. Only slight improvements in the reliability of the forecasts have been detected using the analog method for statistical downscaling.

The downscaled output data have been prepared and delivered to their relevant users within the project.

Target region	Target variables	Obs. data ref.	Global forecast	Method	Approach	Institution
Europe	T, P	WFDEI	ECMWF-S4	DBS	Bias correction	SMHI
Rhine basin "SOSRhine" Case study	T, P	E-OBS v12	ECMWF-S4	QM		KNMI
France RIFF prototype	T, P	SAFRAN (8km analysis over France)	Météo-France-S4	CMV, eQM	Interpolation + BC	MF
Swedish Hydropower basins	Spring flood volume	Swedish discharge measure ments	ECMWF-S4	SVD	Perfect Prog SD	SMHI
Eastern France	T, P		ECMWF-S4	Analog		EDF
Southwest UK	Tmin	E-OBS v11	ECMWF-S4	Analog, LR		UC
Southern Italy	T	E-OBS v11	ECMWF-S4	Analog, LR		UC

Table 1: Summary of the calibration and downscaling studies carried out in this Deliverable. See text for acronyms.

2 Objectives

2.1 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

2.2 Work Package objectives

The main objectives of this WP are (DOW, Workplan Tables 3, WP21):

1. To develop and apply a set of bias-correction and downscaling methods for use with seasonal to decadal forecasts;
2. To downscale standard climate variables and advanced climate indices for use in the EUPORIAS case studies;
3. To assess the uncertainty associated with the downscaling methods in collaboration with WP 33;
4. To make available, through the EUPORIAS web portal, downscaled data, along with a number of calibrated methodologies.

This deliverable contributes to objectives 1 and 2 above, using bias correction and statistical downscaling methodologies. The dynamical downscaling counterpart within these objectives is covered by D21.2, while the downscaling of advanced climate indices (objective 2) will be covered in the upcoming D21.3.

Objective 3 was considered in collaboration with WP32 (Uncertainty framework) and WP33 (Communicating levels of confidence) and the findings regarding statistical downscaling and bias correction can be found in D32.1.

Regarding objective 4, the ECOMS-UDG provides data, access tools and calibration and downscaling tools, thus making available regional-scale seasonal forecasts for impact modelling and other end users. These tools are documented elsewhere (see Milestones MS12 and MS15) and will not be covered in this Deliverable, which focuses on analyses of calibrated/statistically downscaled variables over target regions in Europe to specifically address the needs of different prototypes and case studies.

3 Detailed Report

Highlights

- Several bias correction and perfect-prog statistical downscaling methods successfully applied to provide input for a range of impact sectors in Europe.
- Similar performance of both approaches to remove biases and preserve forecast skill.
- Slight improvement in reliability by the analog method.
- Downscaled data delivered to climate service prototypes and case studies.

3.1 Introduction

Statistical downscaling is one of the means to bridge the gap between the coarse global model forecasts and the local information required for impact assessment (Wilby and Wigley, 1997; Maraun et al., 2010). It essentially considers a model (a transfer function or other algorithmic procedures) which links global model variables (predictors) to surface local variables (predictands). These Statistical Downscaling Models (SDMs) need to be calibrated (or trained) using observations before they can be applied. During this calibration phase, the SDM "learns" the relationship between predictors and predictands. This relationship is stored (e.g. as a given set of parameter values for the transfer function) and can later be applied to predictands out of predictors produced by a new forecast.
There are two approaches to calibrate an SDM:

- **Perfect Prognosis**: or Perfect "Prog" (PP, e.g. von Storch et al., 1993) uses only observations during the calibration phase, both for the predictors and predictands. The observed empirical relationship between both sets of variables is learnt by the SDM. Then the SDM is applied to a global model output variable, expecting it to produce predictors that resemble those observed (perfect prognosis).
- **Model Output Statistics** (MOS, e.g. Teutschbein and Seibert, 2012): uses model predictors and observed predictands during the calibration phase. This ensures the compatibility of the predictors during the SDM application phase. However, for long term model simulations, model output and observations differ in their day-to-day evolution and, therefore a MOS-SDM cannot rely on day-to-day correspondence for long-term prediction and has mainly been applied in short range NWP.

EUPORIAS focuses on seasonal to decadal time scales and, therefore, day-to-day correspondence between model forecasts and observations cannot be expected. The application of MOS-SDMs must then be restricted to SDMs not relying on this correspondence. Distribution-based SDMs build relationships between the probability distribution of a model predictand and a local, observed predictor. In this way, the order of the data does not enter the model and it can be used on S2D forecasts or even longer-term climate simulations.

Distribution-based MOS SDMs are most commonly referred to as "**bias correction**" techniques, especially when the model predictor variable is the same as the local predictand. The simplest approach is to relate the mean of the probability distribution of the predictand (e.g. modelled precipitation) to the mean of the predictor (observed precipitation). Even for this simple approach, the transfer function can take different shapes, and their adequacy depends on the variable. For instance, for precipitation, which has a lower bound and a gamma distribution, a multiplicative correction is usually applied (Wetterhall et al. 2012). However, for temperature, usually normally distributed, an additive shift is more appropriate (Casanueva et al. 2013). State-of-the-art bias correction techniques apply more sophisticated transfer functions between the predictor and predictand PDFs, usually mapping several sample quantiles (Déqué 2007) and/or assuming theoretical distributions for the variables (Piani et al. 2010). Therefore, these techniques need to be tailored to the different statistical nature of the target variables required by the end-users. The next section shows the performance of several of these techniques.

The dependence of the SDM on the predictand is even stronger under the PP approach. In this case, even the predictors might change according to the predictand. For each target variable and region, suitable predictors should be found. In return from this additional complexity, this approach usually takes advantage of the day-to-day correspondence in the observations. This fact, along with the use of large-scale predictors, provides a potential to improve the skill[1] of a global seasonal forecast (see also Deliverable D32.1). Global surface variable forecasts without skill due to unresolved processes can benefit from an SDM linking the processes again to large-scale skillful forecasts. This potential is absent in bias correction techniques, which use the surface variable from the global model and can only affect the PDF, but not the sequence of anomalies.

As any other statistical model trained with data, SDMs (either PP, MOS, or bias correction) can be subject to overfitting: a very good fit of the training data, but an inability to generalize the relations found to independent data. Overfitting can be avoided by using parsimonious models (low number of parameters) to prevent the model to fit even the noisy features of the training data. Overfitting can also be tested by using a proper cross-validation framework: test the model on independent data, not used for training (Kohavi, 1995; Gutiérrez et al, 2013).

As a result of all of the above, SDMs require good quality observations (to find meaningful relationships) and long records (for sufficient training and validation). Moreover, local forecasts can only be produced for those locations where observations are available. Therefore, this technique can only be applied in regions with good quality data coverage. Europe meets this requirement and, therefore, statistical downscaling is the main approach to bridge the gap between global forecasts and impact models.

3.2 Bias correction

[1] In this document, the word "skill" is reserved to the ability of a forecast system to correctly predict future anomalies of a given variable. More specific terminology used across this document can be found in the EUPORIAS Glossary of terms (http://www.euporias.eu/glossary).

Given the inability of this approach to improve upon the skill of a global seasonal forecast; the analyses shown in this section, focus on the performance of different bias correction methods under a cross-validation framework. Even if skill is not improved, the use of unbiased regional forecasts is important as input for the impact models. These models can be highly sensitive to input biases given that they are non-linear and include processes dependent on absolute thresholds. Even with very low seasonal forecast skill from the (necessary) atmospheric input, impact models can benefit from initial conditions of slow evolving processes (e.g. soil moisture, snow pack, …) as a source of predictability.

3.2.1 Pan-European bias correction for flood risk estimation using E-HYPE

SMHI has bias corrected ECMWF System 4 seasonal forecast data for the period 1981-2010. The DBS method (Yang et al., 2010) has been used to conduct the bias-correction. A preliminary analysis of the effect of the bias-correction on the bias in the seasonal forecasting data has been made (see below). A more in-depth analysis of the bias-correction effect will be made in close cooperation with the evaluation of the hydrological modeling part in WP23.

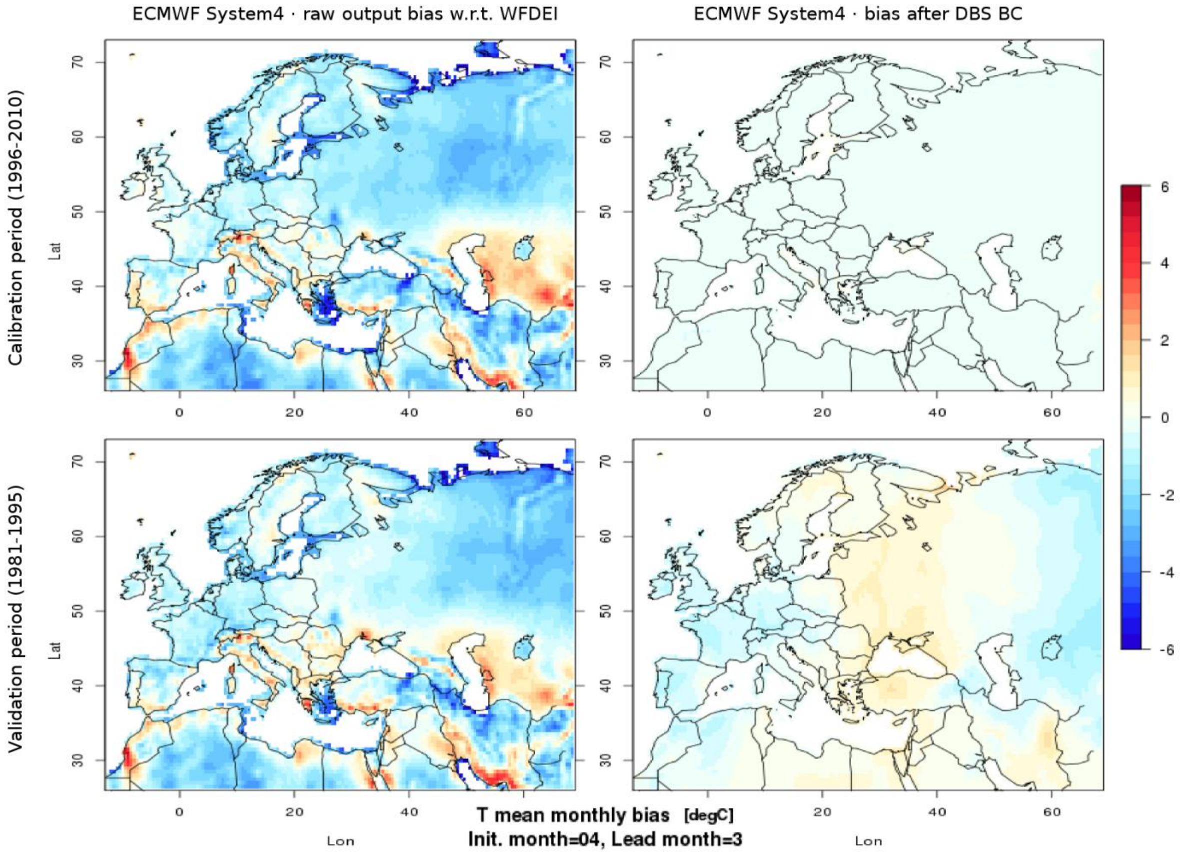

Figure 1: Mean monthly bias in temperature for the seasonal forecasts initialized in April and looking at the 3 months lead time. Shown are the bias in the uncorrected (left) and DBS corrected (right) data for the calibration period 1996-2010 (top) and the validation period 1981-1995 (bottom).

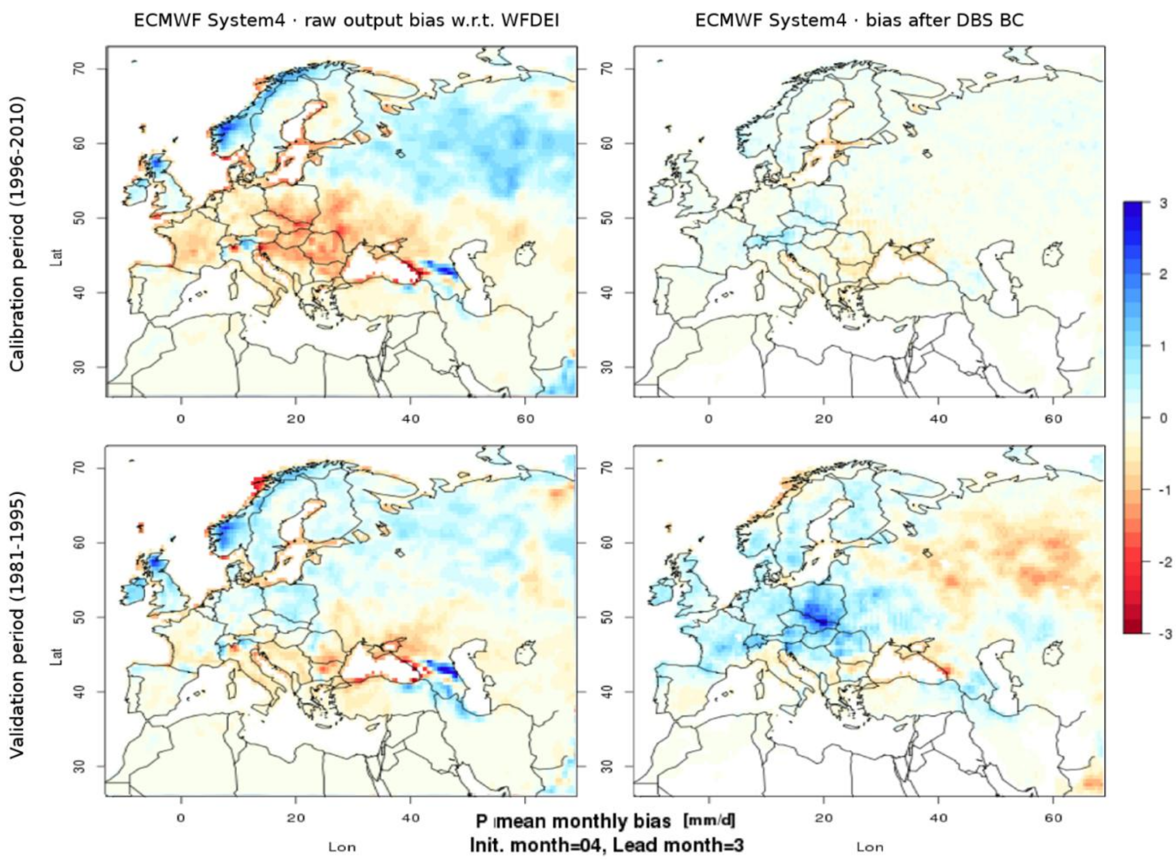

Figure 2: Mean monthly bias in precipitation for the seasonal forecasts initialized in April and looking at the 3 months lead time. Shown are the bias in the uncorrected (left) and DBS corrected (right) data for the calibration period 1996-2010 (top) and the validation period 1981-1995 (bottom).

The data has been split up into a calibration period (1996-2010) and a validation period (1981-1995). Figures 1 and 2 show the monthly mean bias in temperature and precipitation, respectively, for the forecast initialized in April and 3 months lead time. As reference data, the WFDEI data set has been used.

The uncorrected seasonal forecasts generally show a cold bias except for the northern part of the domain in the forecasts for the winter months (not shown). The bias in mean precipitation is more variable in space and time. In forecasts for the winter months, the mean precipitation is underestimated in most parts of the domain. In spring, there is a wide spread overestimation whereas in summer, there is a north-south pattern with overestimation in the northern and underestimation in the southern parts of the domain.

In the calibration period, the bias-correction using DBS effectively removes the bias in the mean monthly temperature and precipitation (top right panel in Figures 1 and 2). In the validation period, the performance of the bias-correction very much depends on the change in the bias in the raw data. In the example given here (see Figures 1 and 2), bias-correction shows good performance for temperature in the validation period. Comparing the bias pattern of the uncorrected data for the calibration and validation period, it is seen that they are quite similar; hence the bias-correction can be applied even outside the calibration period. In the case of precipitation, however, the bias-pattern changes between the calibration and validation period and thus there is still a remaining bias in the validation period after bias-correction, and in some cases, the biases might even be increased. This is generally due to overfitting of the data, i.e. the bias-correction tries to remove unsystematic biases (noise).

3.2.2 SOSRhine case study

KNMI used the ECMWF System 4 seasonal forecast data (Molteni et al., 2011) for the period 1981-2010 and bias corrected them using the Quantile mapping approach (Wilcke et al., 2013). The observation dataset was E-OBS v12.0 (Haylock et al., 2008) for both mean daily precipitation and mean daily temperature. The seasonal forecast data were first interpolated to the E-OBS grid resolution (0.25 deg. regular grid) and then bias corrected. The correction was done in connection to the case-study “SOSRhine” of DWD, KNMI and BFG (stakeholder), which aims to support logistic decisions for inland waterway transport on the River Rhine. The period of interest is in the low-flow seasons from early summer to late autumn as well as in ice early warnings for the impounded river sections in winter time. We proceeded with the bias correction of all seasons for all lead times and performed tests between calibration and validation periods (not shown here). Here we present results for precipitation and temperature (initialization in June at lead time 4).

In Figure 3 the mean monthly precipitation bias for the uncorrected (Figure 3a) and bias corrected (Figure 3b) precipitation forecast initialized in June with lead time 4 months are shown. Focusing on the Rhine and Meuse basin (over Netherlands, Germany, Switzerland, France, Luxembourg and Belgium) there is overall a wet bias, with North Italy having the strongest bias. However, there is no distinct spatial pattern of the bias, with the exception of south Germany where the dry bias is due to topography. After correction the bias is successfully removed.

For temperature (Figure 4); the uncorrected forecast (Figure 4a) shows a strong warm bias in southern Switzerland; again topographical effect but with no distinct spatial pattern over the rest of the basin. After bias correction (Figure 4b) the bias is successfully removed with the exception of north Germany where a warm bias is present and increased. This is probably due to the bias correction method where in some cases overfitting of the data can occur. Further analysis over calibration and validation periods (not shown here), showed that the quantile mapping was effectively removing the bias in both periods. Some seasons show different bias but in all cases the bias was effectively removed.

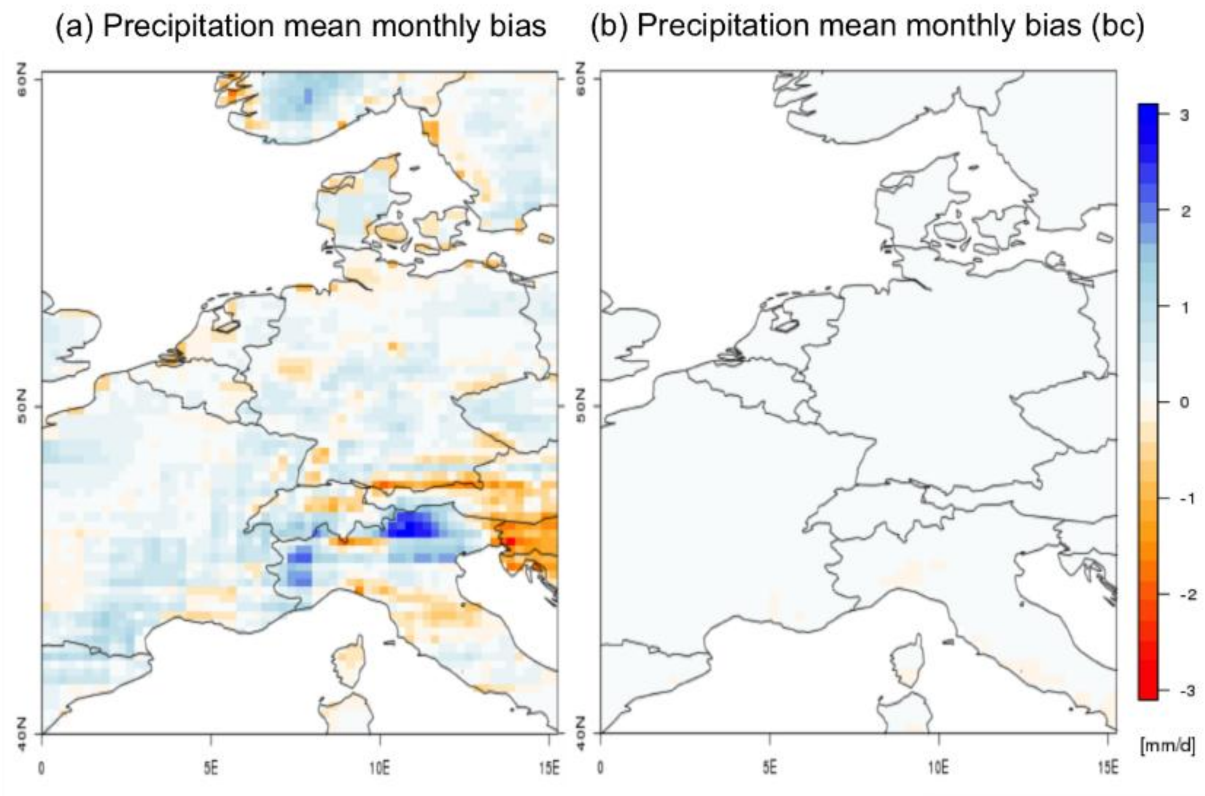

Figure 3: Focusing on the Rhine basin: Mean monthly bias of ECMWF S4 for precipitation forecast initialized in June with lead time 4 months for 1981-2010 for (a) uncorrected forecast and (b) bias corrected forecast.

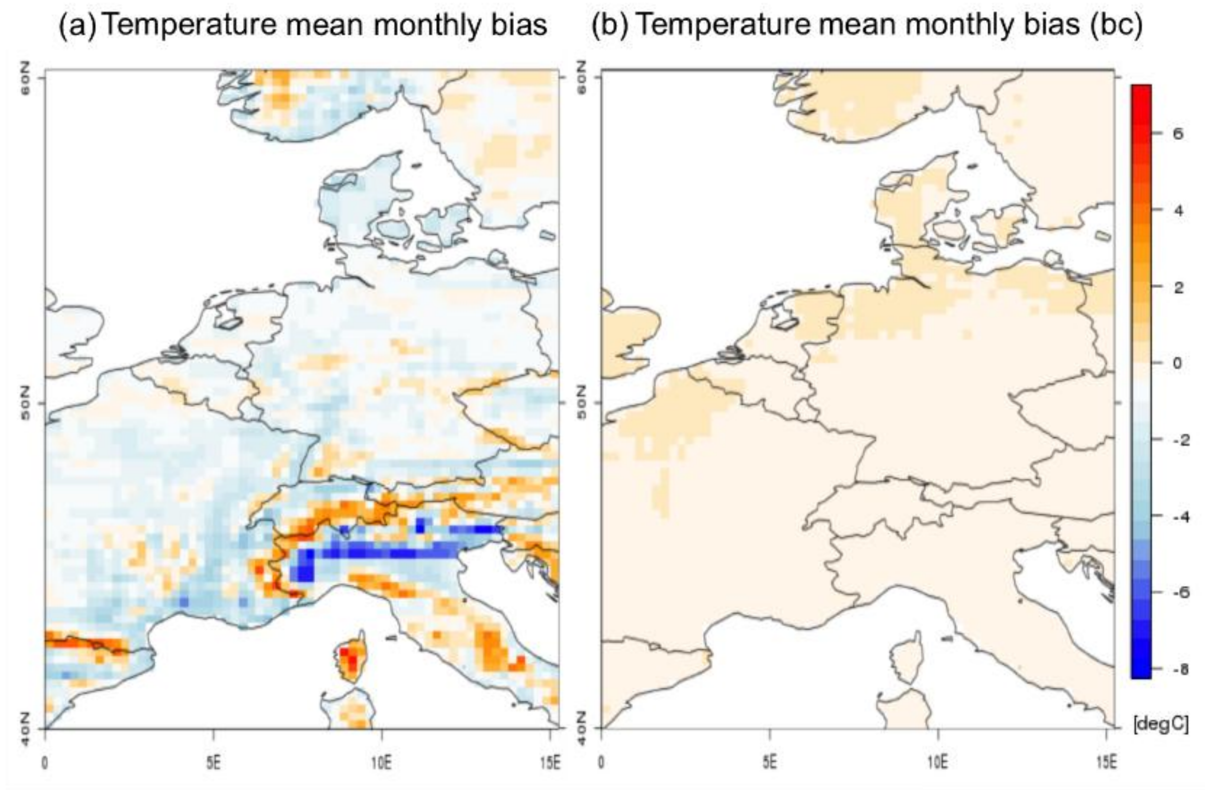

Figure 4: Mean monthly bias of ECMWF S4 for temperature forecast initialized in June with lead time 4 months for 1981-2010 for (a) uncorrected forecast and (b) bias corrected forecast.

3.3 Statistical Downscaling

Perfect Prog statistical downscaling has been applied to different regions in Europe in support of several case studies and climate service prototypes. A first study comparing PP and bias correction methods over regions of low (Spain) and high (Philippines) skill was carried out in Deliverable D32.1. More specific examples, targeted on specific case studies, can be found in this section.

In this case, the focus of the analyses is the skill of the downscaled forecasts.

3.3.1 Flood volumes for hydropower in Swedish catchments

SMHI has worked on the statistical downscaling of large scale circulation variable (LSCV) predictors from the ECMWF System 4 seasonal forecast system to accumulated spring flood volumes in selected Swedish catchments for the period 1981-2015. The statistical downscaling for 84 sub-basins in 7 of the leading hydropower producing catchments has been completed, including the 26 sub-basins in the Ångerman River. These data have been delivered for use in the hydropower prototype in WP42.

A singular value decomposition (SVD) based regression model (e.g. Uvo et al., 2001 and Paul et al., 2008) was developed to downscale large-scale circulation predictors to accumulated spring flood volumes. The downscaling was performed for five initializations, 1 January, 1 February, 1March, 1 April and 1 May, for the period 1981-2015. The LSCV predictors used were pressure field variables (geopotential height and wind components), mean sea level pressure, near-surface wind, radiation/temperature variables (t, 2t and sshf) and moisture related variables (q,slhf and tp) and modelled snow pack depth. The domain of these predictors is 75W-75E and 30N-80N (Figure 5). The predictands used were the observed accumulated spring flood volumes at 84 gauging basins.

LSCV predictors were selected by performing a k-fold cross-validated SVD analysis of the predictors and the predictands for each initialization date. This is done to improve the probability of selecting predictors that have robust skill while minimizing the risk of overfitting.

Evaluation of the downscaling will be done in coordination with WP42.

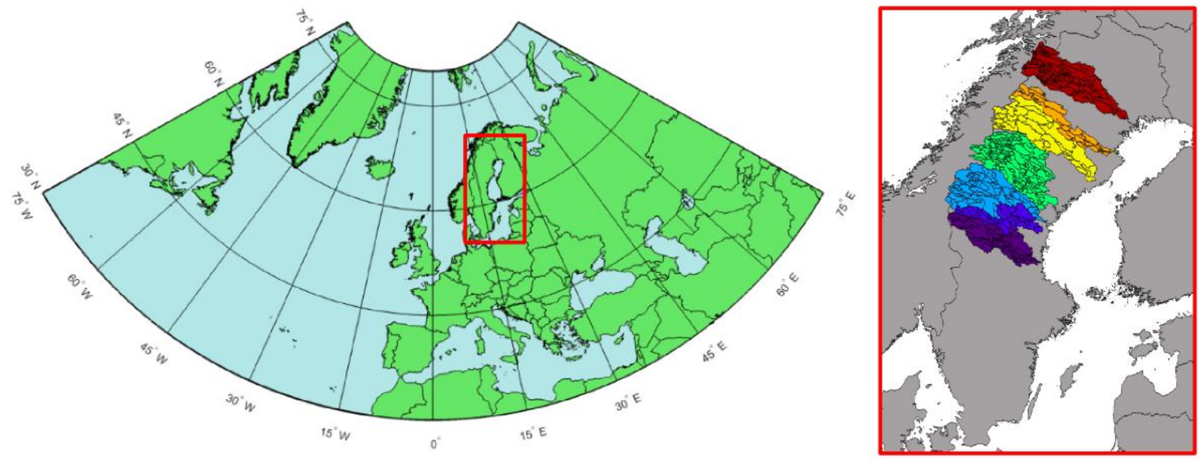

Figure 5 (left) Domain of the large scale circulation variables used as predictors in the statistical downscaling and (right) location of the 84 sub-basins to which the LSCV were downscaled to.

3.3.2 River flow in Southeastern France

EDF ran a study to downscale (WP21) ECMWF S4 hindcasts in order to then run its hydrological model MORDOR (WP23).

The downscaling process is based on the analog method, which uses as input data daily Z1000 & Z700 forecasts. The method, already tested on ECMWF monthly forecasts, provides T2m & precipitation forecasts over 33 watersheds (Figure 6)

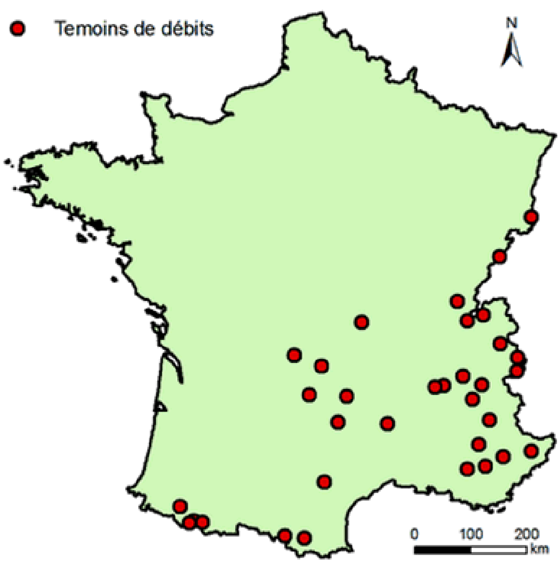

Figure 6: Localization of the 33 watersheds.

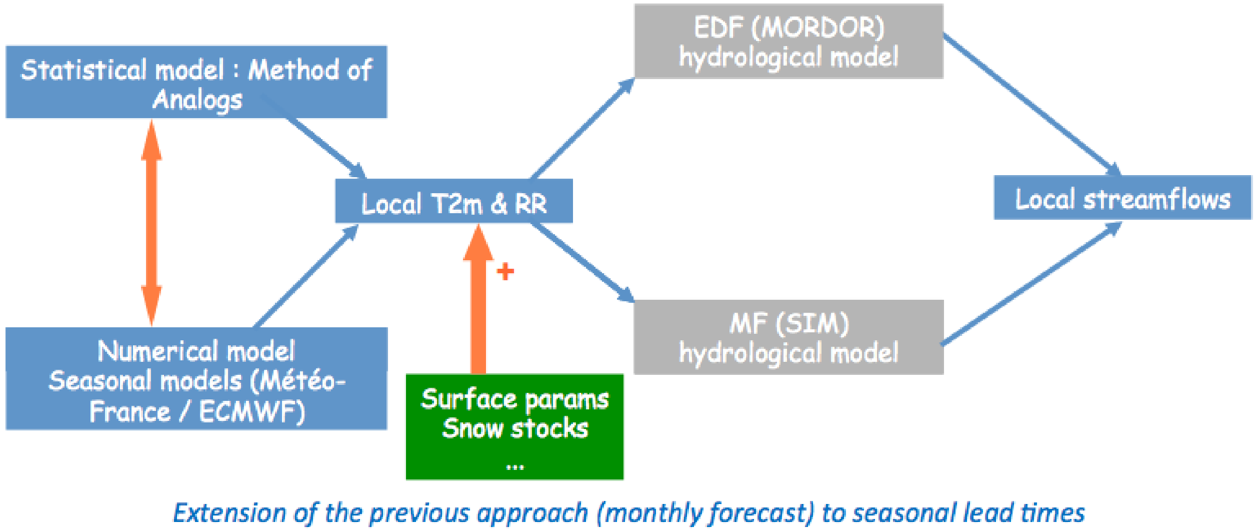

Figure 7: Study description

The analog method's results were compared to Météo-France S4 hindcasts for T2m and RR. However, in the following, the MF forecasts should not be considered as a bug was identified after the study itself (MF provided data at grid points whereas EDF uses data averaged on grid cells).

Regarding the temperature and rainfall forecasts, and considering the average performance of all models (System 4, ARPEGE-Climat, analogs method) regardless of the initialization dates and lead times, it appears that none of them is significantly better than another, whether in deterministic or probabilistic terms. The only difference lies in the analysis of probabilistic scores, which show that the reliability of the analogs method is above the others. Overall, it can only be said that on all basins, the models are well tuned because they are close to the climatology, but they do not really provide additional information on the forecasts for the whole year. The relevance of these seasonal forecast models does not lie in finding a general conclusion to identify recurring behaviors. It is rather in "exceptional" situations that they have an interesting contribution. Indeed, the forecasts studied here show that the System 4 model and the analogs method are able to detect cases of large rainfall deficits (summer drought from 1988 to 1992) and colder than normal temperatures (winters from 1985 to 1987). Those two particular cases are interesting for possible decisions for the management of hydraulic dams. Similarly, forecasts of the ARPEGE-Climat/SAFRAN model have clearly a contribution in spring. Nevertheless, these results are relative, because in selecting these special events, the sample size was reduced (from two to five years of data have been tested, instead of 30). Therefore, they do not rely on strong statistical basis, but are a good indicator for evaluating the potential of the models with this type of situation.

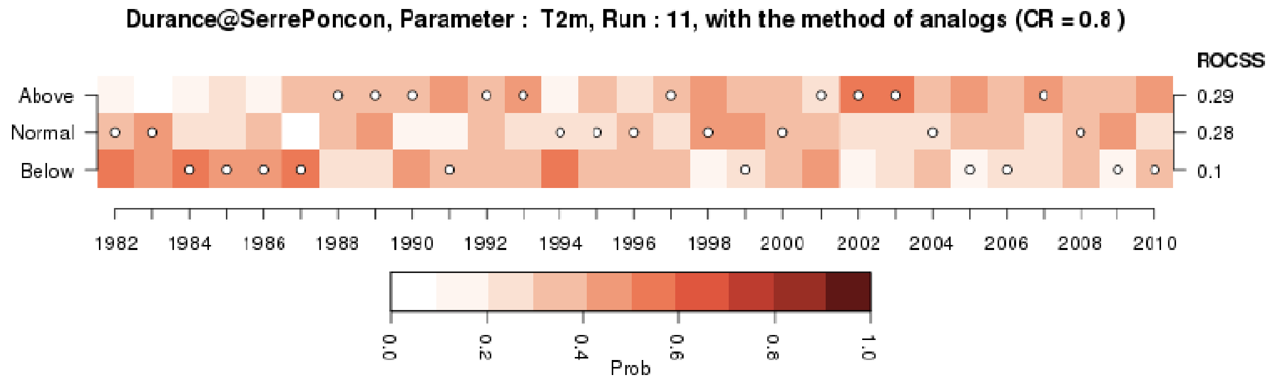

Figure 8: Tercile plot of T2m forecasts at Serre-Ponçon Watershed, Init date: November, Roc Skill Score

3.3.3 Downscaling for the River Flow Forecast (RIFF) prototype

To develop the prototype RIFF (dedicated to water resource management), Météo-France (MF) had to choose a method to downscale seasonal forecast at 8km resolution. In a previous work (Singla, 2012), MF tested 2 different downscaling methods of climate forcing, upstream from our hydrological model. A complex statistical method based on weather type classification and analogs (using DSCLIM, Pagé, 2009) has been compared with a simple spatial interpolation and calculation of standardized anomalies. In an operational perspective, we have decided to keep the most efficient one (balance between complexity and skill), which appears to be the simplest one: a simple correction of mean and variance (CMV). Concretely this method combines a bilinear interpolation to an 8 km grid and a basic correction, which shifts the distribution to remove the bias, and scales the distribution to adjust the variance. For precipitation, additionally, negative values where zeroed out. This method is applied to 6-hours temperature and to 24-hours total precipitation. Temperature is then linearly interpolated hourly, and daily precipitation simply equally spread in hourly amount.

In the last year of the project, we are refining the downscaling method we used. Indeed, to prepare a pre-operational version of RIFF, we will use the new MF Seasonal Forecast model (ARPEGE-S5, operational very soon), which has a higher resolution than the version we have used so far. We will also update our hydrological model. In this context, we will have the opportunity to test an empirical quantile mapping (eQM) to obtain a finer local adaptation of the atmospheric forcing. We have chosen to test the method presented in Mahlstein, 2015. It consists on calculating daily time-series of Seasonal Forecast and observed quantiles (Figure 9). Then we smooth them using a local polynomial regression fitting (LOESS). And finally we compute a quantile-quantile correction.

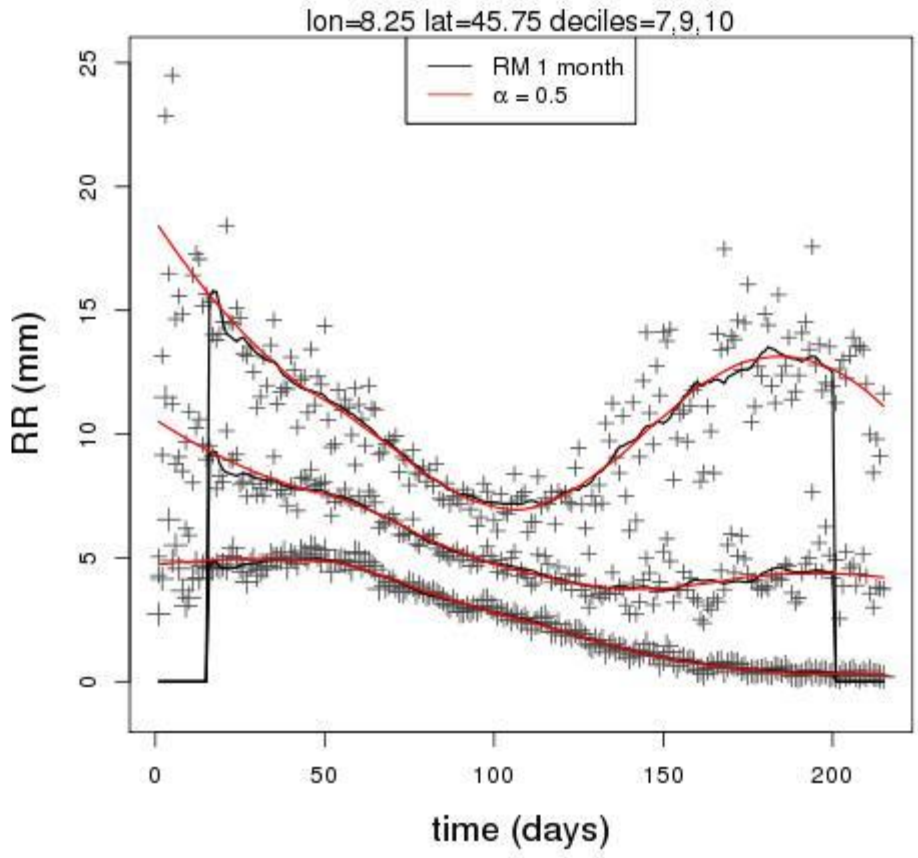

Figure 9: ARPEGE-S5 daily Precipitation forecast, deciles 7, 9 and 10 (maximum), at 45.75N/8.25E (lat/lon). Symbols "+" represent raw quantile data, black lines represent a 30-day running mean, red lines represent a LOESS-fitting.

3.3.4 Temperature forecasts over Southern Italy and Southwest UK

UC has provided statistically downscaled temperature forecasts to support the electricity demand case study over southern Italy and the Land Management Tool prototype over southwest UK. As a benchmark, the global forecasts were also bias corrected by means of an empirical quantile mapping (eQM).

In all cases, the global forecast system used was ECMWF's System4 (S4) and the surface observational database E-OBS v11, which has 0.25ºx0.25º spatial resolution. The full S4 period (1981-2010) was considered and we worked with 1-month lead time forecasts validated in a cross-validation framework. Three different frameworks were considered: (1) no cross validation, i.e. all data used for calibration and validation (2) 5-fold contiguous cross validation, i.e. 6-year periods were sequentially withdrawn from the sample before calibration and used as independent data for validation and (3) 5-fold random cross-validation, which is similar to the previous one, except that the 6-year blocks are not contiguous, but randomly chosen at daily time scale. The latter avoids potential stationarity problems of the contiguous k-fold in the presence of trends. In any case, the results from the 3 approaches are very similar so in this report we show only 5-fold random cross-validation results as an example.

In all cases, the SDMs are trained on daily data corresponding to each training set, applied at daily scale and validated as seasonal averages.

The SDMs considered were:

- Analogues: A single analog chosen to minimize the Euclidian distance in the space spanned by the 20 leading principal components (PCs) of the predictors on an extended domain.
- Linear Regression: Multiple linear regression using as covariables the 15 leading PCs of the predictors over an extended domain, plus the average of the predictors on the 4 nearest neighbours to the target point.

In both methods, ERA-Interim and E-OBS were used to train the models, using as predictors mean 2m temperature and MSLP. The predictor domains are 18W-4E 49-59N (UK) and 6-26E 36-46N (Italy). These methods and predictor sets are part of the EU COST action VALUE, where the ERA-Interim data was considered at 2º resolution. We followed this approach and interpolated the S4 forecasts to this coarse resolution to train the models. We also considered the nearest S4 gridpoint temperature (the simplest interpolation procedure), along with its bias correction by means of eQM (following Déqué, 2007). We plan to publish this analysis in an international peer-reviewed journal. An overview of the main results is provided next.

The target variable over southwest UK is minimum temperature and the target season was winter (DJF). Therefore, we used the November initialization (lead month 1).

The direct S4 output presents strong biases (Figure 10, left), which are efficiently removed by the PP SDMs and the eQM. Some slight biases remain after the analog method was applied. This is expected, given that this method does not minimize a cost function and relies on finding a close analog situation in the 24-year period database. In any case, remaining biases are small.

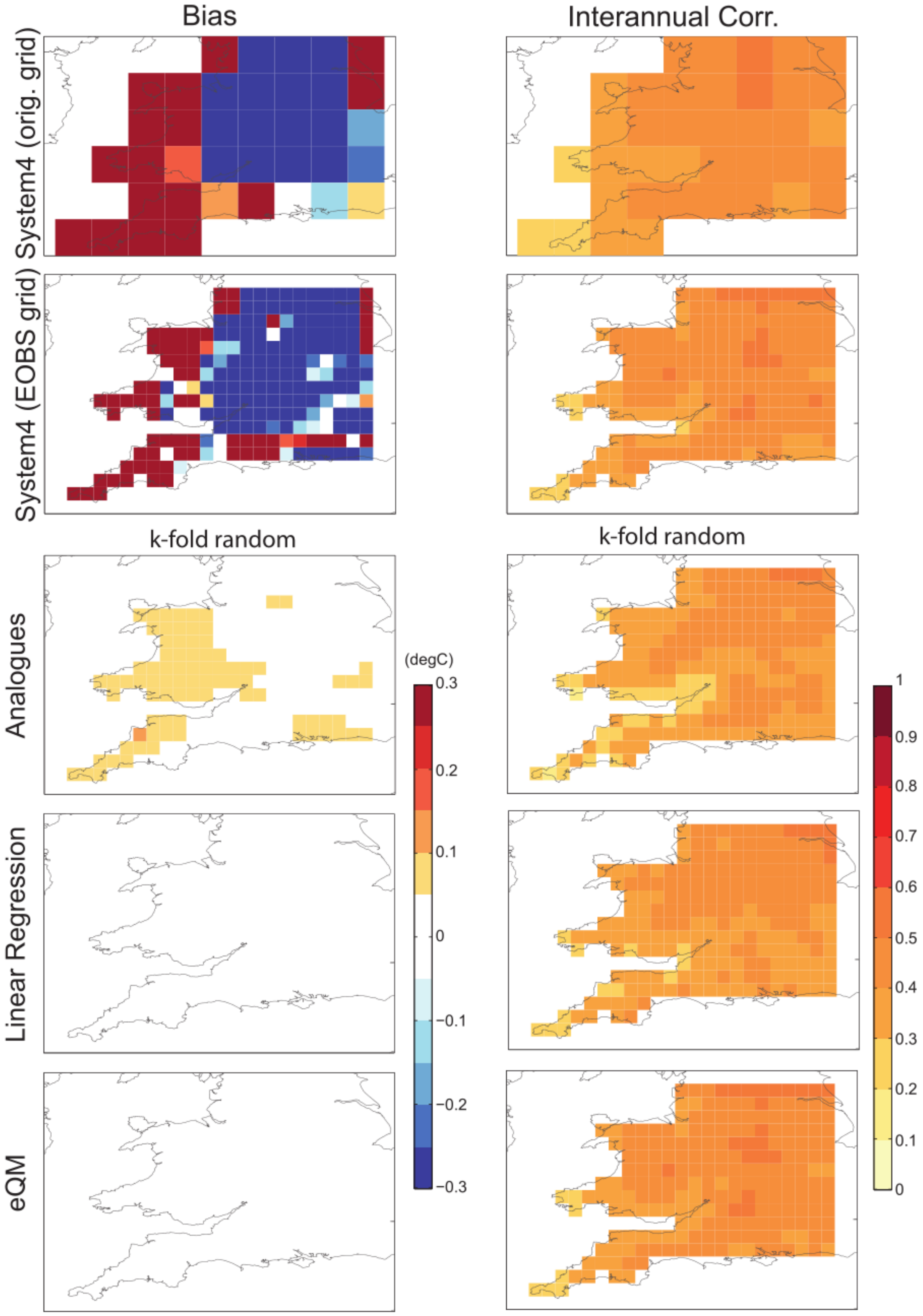

Figure 10: Winter (DJF) minimum temperature biases (left) and interannual correlations (right) for the ECMWF System 4 raw output, this output interpolated (nearset neighbour) to the E-OBS grid (0.25º), downscaled by means of analogues and linear regression, and bias corrected by empirical quantile mapping (sequentially in rows). All downscaled output has been produced using a 5-fold random cross-validation. The observational reference is E-OBS in all cases.

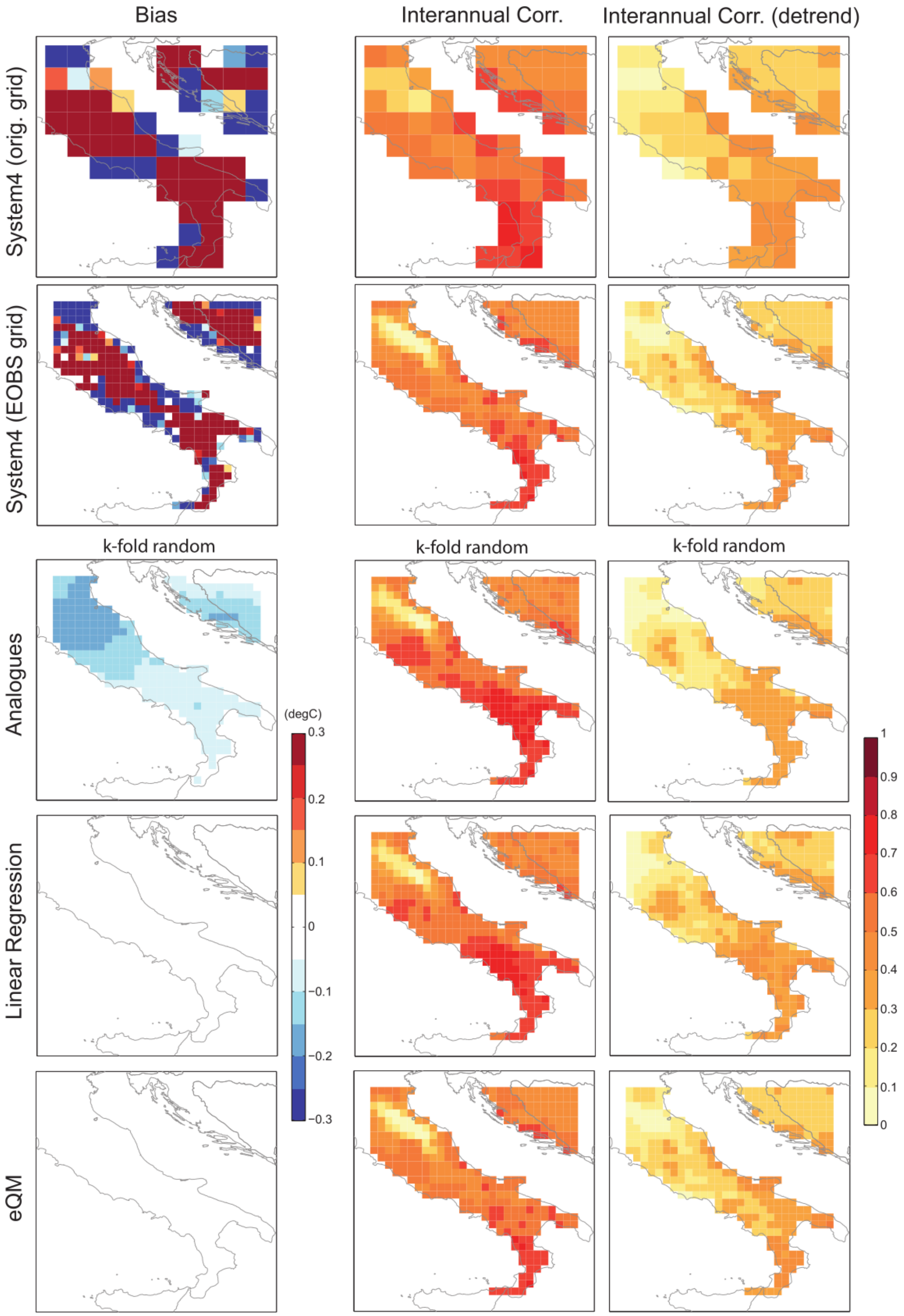

Figure 11: As Figure 10, but for summer (JJA) mean temperature over southern Italy. The additional last column represents interannual correlation after detrending the series.

In order to check the ability of the forecasts to actually predict seasonal anomalies, Figure 10 (right) shows the interannual correlation of forecast and observed anomalies. Correlations are mostly in the range 0.5-0.6 in S4 and drop a bit (0.4-0.5) after downscaling or bias correction. PP SDMs and bias correction (eQM) perform similarly in terms of interannual correlation.

A similar analysis was performed over southern Italy (Figure 11), where the target variable is mean temperature and the target season is summer (JJA). Therefore, we used the May initialization (lead month 1). Again, most biases are removed and, in this case, a slight cold bias remains after the analog method. Interannual correlations are surprisingly high over southern Italy (reaching 0.7-0.8). This is due to a trend in summer mean temperature over the area, given that correlations drop to 0.4-0.5 when the data are detrended.

Correlation is a deterministic accuracy measure, which summarizes the ability of the ensemble average to follow the observed seasonal anomalies. Tercile plots (Díez et al, 2011; Manzanas et al, 2014) show the probabilistic forecast for each season along with the observed anomaly (Figure 12). The forecasts show low sharpness (intermediate probabilities are assigned to all terciles, instead of high or low probabilities, which would ease decision-making). The trend in the data is also apparent in these plots, and many successful forecasts fail when the trend is removed (remarkably, the sharp 2002 forecast).

Trends need to be removed from the data, given that they are a source of artificial skill. When terciles are computed on a base period with a warming trend, new forecasts and observations tend to cluster in the upper tercile (they agree), but this is just a result of the trend, not skill of the forecast system.

A different aspect of probabilistic forecasts is illustrated in reliability diagrams (Figure 13). These show biases in probability space, i.e. the observed relative frequency of an event (Y-axis) that has been forecast to occur with a given probability (on the X-axis). These two quantities should match (i.e. lie on the diagonal) in order for a probabilistic forecast to be *reliable*. Skillful forecasts (relative to the climatology) lie on the blue shaded area.

S4 forecast is quite reliable. After a simple interpolation, reliability drops for the high probability forecasts. Bias correction is not able to improve reliability (eQM reliability diagram fully resembles that of the raw output interpolated to the observational grid). However, the analog method shows an improvement at high resolution. The unreliability of high forecast probabilities is improved by both PP SDMs. However, these never forecast very high probabilities and, therefore, they do not span the full forecast probability range.

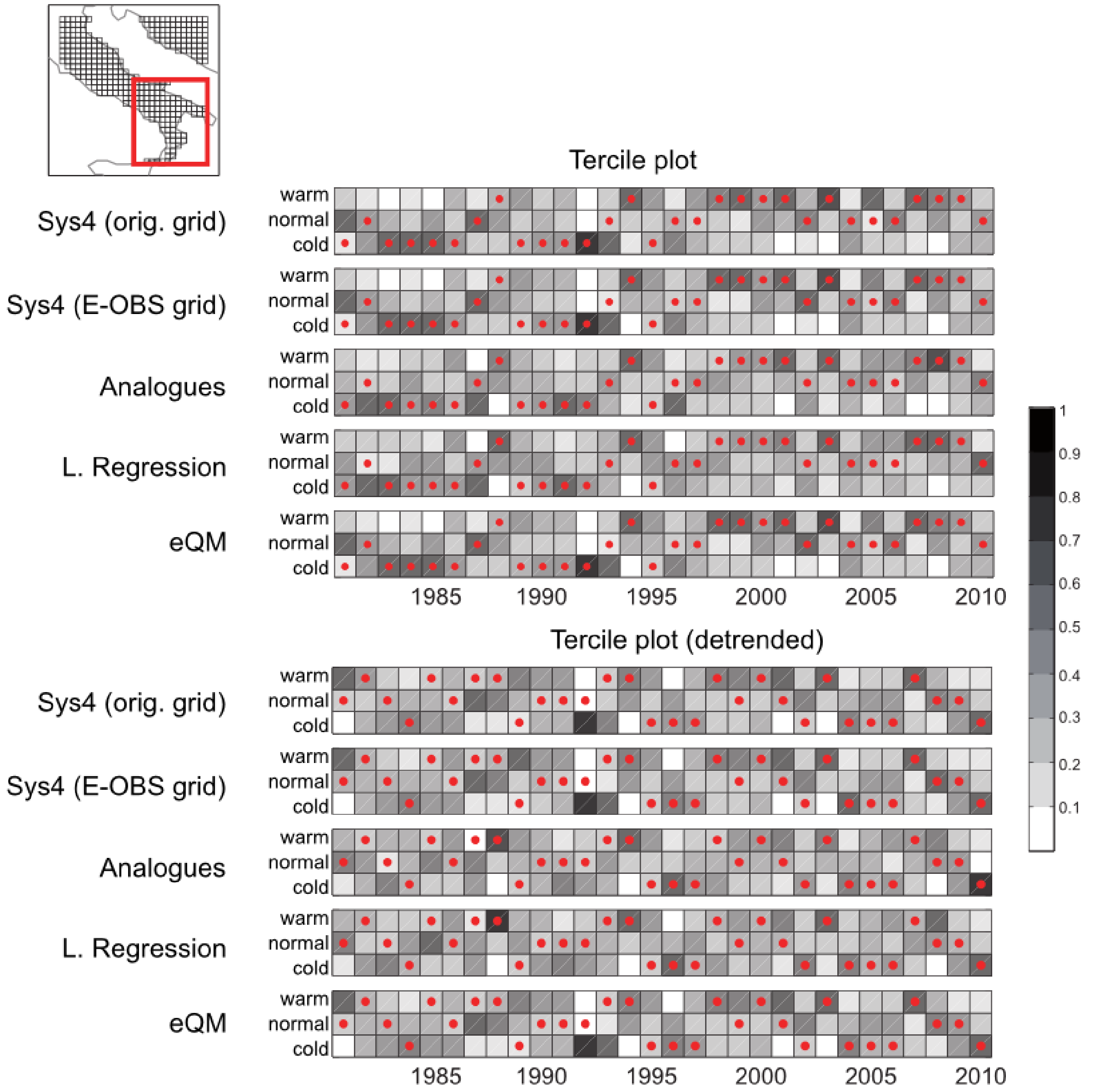

Figure 12: Tercile plots showing the probability of each temperature tercile forecast by each system (shades) along with the observed tercile (red dots). The upper panels show the original series, while the lower ones are detrended.

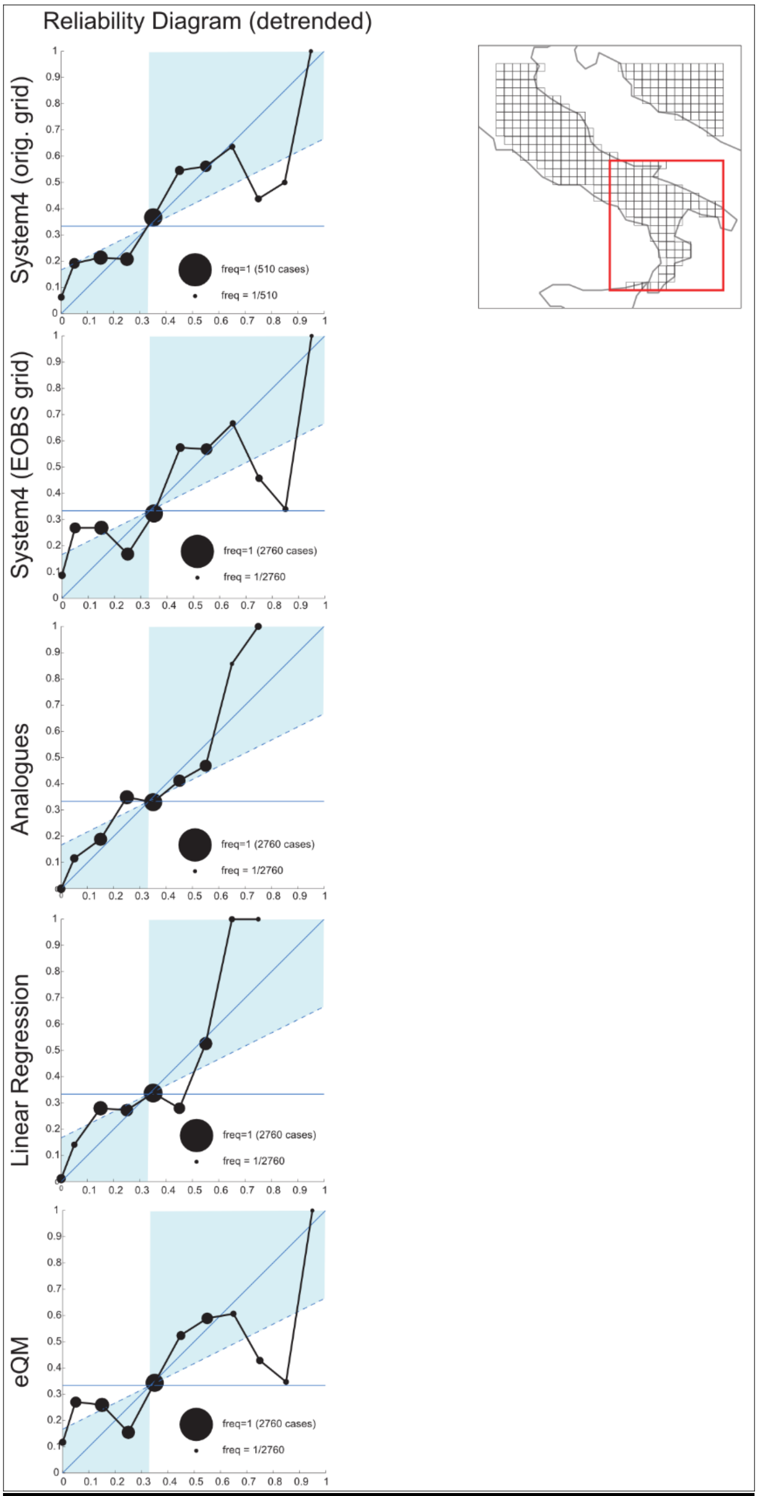

Figure 13: Reliability diagram for each forecast system.

3.4 Summary and conclusions

Global seasonal forecasts, mainly from ECMWF's System4, have been statistically downscaled by different groups using both MOS (bias correction) and PP approaches. The target variables were mainly temperature and precipitation.

We found temperature biases related to coarse GCM resolution, unable to resolve coastlines and complex orography. The pattern of precipitation biases is less connected to orography or continentality. In any case, bias correction and PP SDMs are both efficient in removing the biases, even during test periods not used in their calibration.

The skill of the PP downscaled forecasts is not improved even though the method has potential to improve the skill in other regions of the globe (see e.g. the example over Philippines in Deliverable D32.1). Only the analog method has shown a slight increase in reliability, as found in different analyses over France (EDF) and Italy (UC). However, they present other problems, such as a lack of high probability forecasts. Conditional forecasts such as those on ENSO years (Frias et al, 2010) or leading to extreme events could potentially bring more useful forecasts, but the sample is then dramatically reduced for the assessment of the statistical significance.

From a practical point of view, given the similar performance of BC and PP SDMs, some groups (e.g. MF) have relied on simple BC methods to bridge the spatial scale gap between global forecasts and impact models. Most groups are still connecting the output generated in this WP to impact models (WP23 and 42), where a more comprehensive analysis of the usefulness of downscaled forecasts can be addressed.

3.5 References

Casanueva, A., S. Herrera, J. Fernández, M. Frías, and J. Gutiérrez (2013): Evaluation and projection of daily temperature percentiles from statistical and dynamical downscaling methods. Nat. Hazards Earth Syst. Sci., 13, 2089-2099, doi:10.5194/nhess-13-2089-2013.

Déqué, M (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. Global and Planetary Change, 57 (1-2), 16-26, doi:10.1016/j.gloplacha.2006.11.030.

Díez, E., B. Orfila, M. D. Frías, J. Fernández, A. S. Cofiño, and J. M. Gutiérrez (2011) "Downscaling ECMWF Seasonal Precipitation Forecasts in Europe Using the RCA Model." Tellus A63. doi:10.3402/tellusa.v63i4.15857.

Frías, M. D., S. Herrera, A. S. Cofiño, and J. M. Gutiérrez (2010) "Assessing the Skill of Precipitation and Temperature Seasonal Forecasts in Spain: Windows of Opportunity Related to ENSO Events." Journal of Climate 23:209-220. doi:10.1175/2009JCLI2824.1.

Gutiérrez, J., D. San-Martín, S. Brands, R. Manzanas, and S. Herrera (2013): Reassessing statistical downscaling techniques for their robust application under climate change conditions. J. Climate, 26, 171-188

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New. (2008). A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

Kohavi, R. (1995): A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1137-1143, IJCAI'95.

Mahlstein, I. Spirig, C., Liniger, M. Appenzeller, C. (2015) : Estimating daily climatologies for climate indices derived from climate model data and observations. Journal of Geophysical Research: Atmospheres (120, 2808–2818, doi:10.1002/2014JD022327)

Manzanas, R., M. D. Frías, A. S. Cofiño, and J. M. Gutiérrez (2014) “Validation of 40 Year Multimodel Seasonal Precipitation Forecasts: The Role of ENSO on the Global Skill.” Journal of Geophysical Research: Atmospheres 119:1708-19. doi:10.1002/2013JD020680.

Maraun, D., et al. (2010): Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics, 48 (3), RG3003, doi:10.1029/2009RG000314.

Molteni F, Stockdale T, Balmaseda M, Balsamo G, Buizza R, Ferranti L, Magnusson L, Mogensen K, Palmer T, Vitart F (2011). The new ECMWF seasonal forecast system (System 4). ECMWF Technical Memorandum 656

Pagé, C., Terray, L., and Boé, J.: Dsclim: A software package to downscale climate scenarios at regional scale using a weather-typing based statistical methodology, Technical Report TR/CMGC/09/21, SUC au CERFACS, URA CERFACS/CNRS No1875, Toulouse, France, 2009.

Piani, C., J. O. Haerter, and E. Coppola (2010): Statistical bias correction for daily precipitation in regional climate models over Europe. Theoretical and Applied Climatology, 99 (1-2), 187-192, doi:10.1007/s00704-009-0134-9.

Paul, S., Liu, C. M., Chen, J. M., & Lin, S. H. (2008). Development of a statistical downscaling model for projecting monthly rainfall over East Asia from a general circulation model output. Journal of Geophysical Research: Atmospheres (1984–2012), 113(D15).

Singla, S., Céron, J.-P., Martin, E., Regimbeau, F., Déqué, M., Habets, F., and Vidal, J.-P.: Predictability of soil moisture and river flows over France for the spring season, Hydrol. Earth Syst. Sci., 16, 201-216, doi:10.5194/hess-16-201-2012, 2012.

Teutschbein, C. and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology, 456-457, 12-29.

Uvo, C. B., Olsson, J., Morita, O., Jinno, K., Kawamura, A., Nishiyama, K., ... & Nakashima, T. (2001). Statistical atmospheric downscaling for rainfall estimation in Kyushu Island, Japan. Hydrology and Earth System Sciences Discussions, 5(2), 259-271.

von Storch, H., E. Zorita, and U. Cubasch (1993): Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. Journal of Climate, 6 (6), 1161-1171, doi:10.1175/1520-0442(1993) 006h1161:DOGCCEi2.0.CO;2.

Wetterhall, F., Pappenberger, F., He, Y., Freer, J., Cloke, H.L. (2012). Conditioning model output statistics of regional climate model precipitation on circulation patterns. Nonlin. Processes Geophys. 19, 623–633. doi:10.5194/npg-19-623-2012

Wilby, R. L. and T. M. L. Wigley (1997): Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography, 21 (4), 530-548, doi:10.1177/030913339702100403.

Wilcke, R., T. Mendlik, and A. Gobiet (2013). Multi-variable error correction of regional climate models. Climatic Change, 120(4), 871-887

Yang, W., Andréasson, J., Phil Graham, L., Olsson, J., Rosberg, J., & Wetterhall, F. (2010). Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. *Hydrology Research*, *41*(3-4), 211. doi:10.2166/nh.2010.004

4 Lessons Learnt

The main lesson learnt is the similar performance of simple bias correction as compared to PP statistical downscaling approaches. This seems to be so due to the currently low skill of the global forecasts, both at the surface and in large-scale predictors over the area.

5 Links Built

All the results from this deliverable are inputs for WP23 and WP42. In dealing with the uncertainty associated to the BC/SD step, we worked with WP32 and WP33 (main results gathered in Deliverable D32.1).

The downscaling tools provided by the downscaleR package were developed in collaboration with the EU-funded project SPECS. BC methods were implemented in the framework of EUPORIAS and PP SDMs provided by SPECS (see Milestones MS12 and MS15).

The downscaling framework used by UC (predictors and cross-validation framework) agree with that developed in the EU COST Action VALUE.