THEME ENV.2012.6.1-1

EUPORIAS

(Grant Agreement 308291)

EUPORIAS

European Provision Of Regional Impact Assessment on a

Seasonal-to-decadal timescale

Deliverable D*22.2*

PROVIDE METHODOLOGY TO CALCULATE CIIS AND THEIR SKILL

Deliverable Title	*Provide methodology to calculate CIIs and their skill*	
Brief Description	*Provide methodology to calculate CIIs and their skill: Methodology to calculate CIIs will be provided, plus their skill for the selected climate service prototypes.*	
WP number	22	
Lead Beneficiary	*MeteoSwiss*	
Contributors	*Irina Mahlstein, Jonas Bhend, Christoph Spirig, Mark Liniger (MeteoSwiss), Verónica Torralba, Rachel Lowe, Markel Garcia, Joan Ballester, Xavier Rodo (IC3), James Creswick (WHO), Joaquín Bedia, Jesús Fernández, Maria Eugenia Magariño (UC), Nicola Golding, Carlo Buontempo (Met Office), Anne-Lise Beaulant (Meteo France), Christiana Photiadou (KNMI), Roxana Bojariu, Sorin Dascalu, Madalina Gothard (Meteo-Ro), Sandro Calmanti (ENEA)*	
Creation Date Version Number Version Date	02/06/2015 6 30/10/2015	
Deliverable Due Date Actual Delivery Date	31/10/2015 30/10/2015	
Nature of the Deliverable	*R*	*R - Report*
		P - Prototype
		D - Demonstrator
		O - Other
Dissemination Level/ Audience	*PU*	*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	02/06/2015	Irina Mahlstein	Template for partner contributions
1	21/09/2015	Irina Mahlstein	Added first partner contributions
2	25/09/2015	Irina Mahlstein, Jonas Bhend	Additional contributions and formatting
3	28/09/2015	Jonas Bhend, Irina Mahlstein	Edited summary information (executive summary, introduction, links built, lessons learned)
4	14/10/2015	Jonas Bhend	Added contribution by KNMI and UL-IDL, added revisions to contributions from IC3, Uni.Can., Met Office, and MeteoFrance
5	28/10/2015	Jonas Bhend	Included contributions by Meteo-Ro and ENEA, updated contribution of Uni.Can. / Met Office, copy editing
6	30/10/2015	Jonas Bhend, Mark Liniger	Further copy editing

Table of Contents

List of Tables

List of Figures

1. Executive Summary

The aim of this work package is to assess the value of seasonal forecasts for a collection of user-targeted climate information indices (CIIs, see Table 3.1.1). CIIs make it possible to provide decision relevant climate information to users in a very direct way. As such, CIIs represent an easy to use and relatively easy to produce alternative to more sophisticated impact models, but may also provide information of more relevance to the user than forecasts of meteorological variables usually produced by seasonal forecasting centres. Forecasts of CIIs evaluated in this work package are used to complement the prototypes and case studies developed as part of the EUPORIAS project. This deliverable documents the implementation of different CIIs and the calibration approaches applied to take into account systematic model errors. Estimates of the prediction skill of CII forecasts and its uncertainty also form part of the deliverable.

Table 3.1.1: List of climate information indices analysed in this deliverable.

Climate index	Stakeholder sector	Figure(s) in deliverable
Percentage of time with anomalous wind speed	Energy	Figure 3.2.1 - Figure 3.2.6
Heating and cooling degree days	Energy	Figure 3.3.1 - Figure 3.3.6
Growing season precipitation and suitability, hydrothermal, Selianinov, cool night, Huglin heliothermal, and Winkler indices	Agriculture (wine production)	Figure 3.4.1 - Figure 3.4.7
Frost days	Agriculture, transport	Figure 3.5.1 - Figure 3.5.2
Heavy precipitation days and total precipitation on wet days	Agriculture, hydrology, insurance	Figure 3.6.1 - Figure 3.6.8
Intense precipitation	Hydrology, civil protection, agriculture	Figure 3.7.1 - Figure 3.7.5
River flow and soil moisture	Agriculture, hydrology, insurance	Figure 3.8.1 - Figure 3.8.6
Palmer drought severity index	Agriculture, hydrology	Figure 3.9.1 - Figure 3.9.3
Water requirement satisfaction index	Agriculture	Figure 3.10.1 - Figure 3.10.2
Fire weather	Civil protection, forestry	Figure 3.11.1 - Figure 3.11.7
Temperature related mortality	Health	Figure 3.12.1 - Figure 3.12.3

We analyse different approaches to calibration of daily forecast time series to derive CII indices. Generally, CII forecast skill is found to be rather insensitive to the choice of calibration method. In some cases, however, using more sophisticated calibration methods (e.g. quantile mapping compared to mean de-biasing) improves the skill in CII forecasts. For specific applications such as forecasting impacts on health, it may be relevant to consider multiple indices (e.g. peak timing of excess mortality, maximum incidence at the peak, etc.) which ideally are calibrated consistently across the different indices. We further find that indices defined with respect to percentiles of the forecast and observed distribution respectively (e.g. percentage of time with wind speed above the 90th percentile) are

advantageous in that such indices are less prone to systematic model errors than indices defined with respect to absolute thresholds (e.g. frost days). However, having in mind that most indices operate on daily values, the estimation of relative thresholds must be handled with care.

The forecast lead times for seasonal forecasts are too long to be able to deterministically forecast extreme events at a daily to weekly resolution. To account for the inherent uncertainty in long-range forecasts, such forecasts are therefore framed probabilistically. Consequently, it is important that climate indices (and impact models) are computed using all available ensemble members rather than the ensemble mean. Also, calibration methods have to be applied in a manner that allows for the uncertainty information contained in the ensemble forecast to be retained.

In general, forecast skill of seasonal forecasts of CII in Europe is limited with enhanced skill in summer compared to winter and generally higher skill for indices related to temperature than to precipitation. However, skill varies strongly by season, region, lead time, climate index, and spatio-temporal aggregation; some regions show skill in specific seasons which may open a window of opportunity for forecasts to be useful in specific cases. Further research, however, is needed to understand variations in the skill in forecasting CIIs. Such further research should also target the temporal variability of predictability and thereby forecast skill.

Also, forecasts of CII are found to be at most as skilful as the forecasts of seasonal means of the underlying meteorological variables. In contrast, enhanced skill is found in forecasts of river flow based on a hydrological model. Forecasts from the hydrological model driven with atmospheric parameters from seasonal forecasts, however, are only marginally more skilful than forecasts driven with climatological input. The skill in forecasts of river flow and soil moisture is thus mainly due to the long-term memory of initial conditions for the hydrological model including soil moisture and snow. This highlights the presence of sources of predictability not included in the operational seasonal forecasting system that may be exploited for targeted applications.

Apart from long-term memory from initial conditions determining interannual variability, external forcing such as increasing greenhouse gases are a potential source of predictability on seasonal timescales. Preliminary analysis suggests that long-term trends are an important source of predictability in winter, whereas predictability in summer seems to be mainly independent of long-term trends. Further research, however, is needed to better understand the sources of predictability in Europe. Also, the impact of calibration on the representation of trends (and thereby predictability) in CII forecasts needs further study to develop trustworthy, well-calibrated forecasts of CIIs.

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. Introduction

In the following we present the methodology and skill analysis for a selection of climate information indices relevant for different sectors. These indices include

- measures of extreme wind speed relevant for wind turbine operation,
- heating and cooling degree as a proxy for energy demand for heating and cooling,
- the growing season precipitation and suitability, and the hydrothermal, Selianinov, cool night, Huglin heliothermal, and Winkler indices relevant for wine production,
- frost days relevant to agriculture and transportation,
- precipitation anomalies to estimate water availability for agriculture,
- measures of intense precipitation to forecast floods,
- river flow and the soil wetness index, the Palmer drought severity index and soil moisture anomaly, and the water requirement satisfaction index to forecast the water balance and droughts,
- fire weather index to forecast years and seasons with anomalous fire danger, relevant to a range of decision makers including forestry, land management, and civil protection sectors, and
- heat and cold related mortality relevant for the health sector.

3.2. Wind

3.2.1. Definition and Equation

Two indices based on 6-hourly 10-m wind speed are used in order to characterize the low and high wind speeds in a particular season. The first index is the percentage of time where the average wind speed is lower than the 10th percentile (sfcWindq10nd) and the second one is defined as the percentage of time where the wind speed is greater than the 90th percentile (sfcWindq90nd).

The two indices can be expressed as:

$$sfcWindq10nd = \frac{1}{n}\sum_{i=1}^{n} d_i \;\; d_i = \begin{cases} 0 \; sfcWind_i > q_{10th} \\ 1 \;\; sfcWind_i < q_{10th} \end{cases}$$

$$sfcWindq90nd = \frac{1}{n}\sum_{i=1}^{n} d_i \;\; d_i = \begin{cases} 0 \; sfcWind_i < q_{90th} \\ 1 \;\; sfcWind_i > q_{90th} \end{cases}$$

where $sfcWind_i$ is the wind speed for a particular time step (using six-hourly data in this case) in a month, the q_{10th} and q_{90th} are the 10th and 90th climatological percentiles (respectively) for each month and n is the number of days in a month.

These indices have been estimated for each month and then averaged across the season. The methodology followed was proposed by Pepler *et al.* (2015) in the first place and developed further in Prodhomme *et al.* (2015).

To obtain the indices, the climatological 90th and 10th percentiles are estimated from the six-hourly wind speeds over the period 1981-2012. The climatological percentile is then used to calculate the percentage of time in which the four values per day of the wind speed exceed the 90th percentile threshold (or are below the 10th percentile threshold). Then, the frequency of days over and under the corresponding climatological percentile in a month is estimated. The methodology has been applied separately for the ERA-Interim reanalysis data (Dee *et al.*, 2011) and for the ECMWF S4 predictions (Molteni *et al.*, 2011) because the percentiles in the reference reanalysis and in the simulations can be very different.

3.2.2. Bias Correction

The sfcWindq90nd and sfcWindq10nd indices have been computed from the ECMWF S4 predictions of the six-hourly 10-m wind speeds with 51-member ensembles over the period 1981-2012. The forecasts considered for illustration are those issued on the 1st of November, for which three-month statistics for the December-January-February (DJF, also known as one-month lead seasonal forecast) period are used. The analysis of the wind speed extremes focuses on the DJF boreal winter since this season was found to have the largest wind speed variability in the Northern Hemisphere.

The sfcWindq10nd and sfcWindq90nd indices were compared with the corresponding variables computed from the six-hourly 10-m wind speed from the ERA-Interim reanalysis.

To identify possible systematic model errors, maps of bias of both sfcWidnq10nd and sfcWindq90nd indices have been computed (Figure 3.2.1a and b, respectively). These maps have been calculated as the difference between the climatology of the indices obtained from ECMWF S4 predictions and that of the reference dataset.

The largest bias is found for sfcWindq10nd index, corresponding to low wind speeds (Figure 3.2.1a). In tropical regions the differences between the predictions and observations reach -0.12, indicating that the prediction system shows a 12% less time under the 10th percentile than ERA-Interim. For the sfcWindq90nd index (Figure 3.2.1b) a negative bias is found in some regions, such as South-eastern Asia or Western Russia, revealing that in those regions the prediction system provides less time over the 90th percentile than the reanalysis dataset over the DJF season.

These differences between the bias for the two indices suggest that in the regions where the systematic error is present, the climatological probability density function relative to the six-hourly wind speed predictions does not match with the probability density function of ERA-Interim as well as the associated percentiles values.

In general, the differences between the indices (sfcWindq10nd and sfcWindq90nd) in the predictions and those computed for the reanalysis are low and mainly located over oceanic regions. This result suggests that the bias correction is a not critical step to correct the wind extreme indices, because they are based on relative thresholds (10th and 90th percentiles) that are computed separately for the predictions and the reference, taking into account implicitly the bias correction.

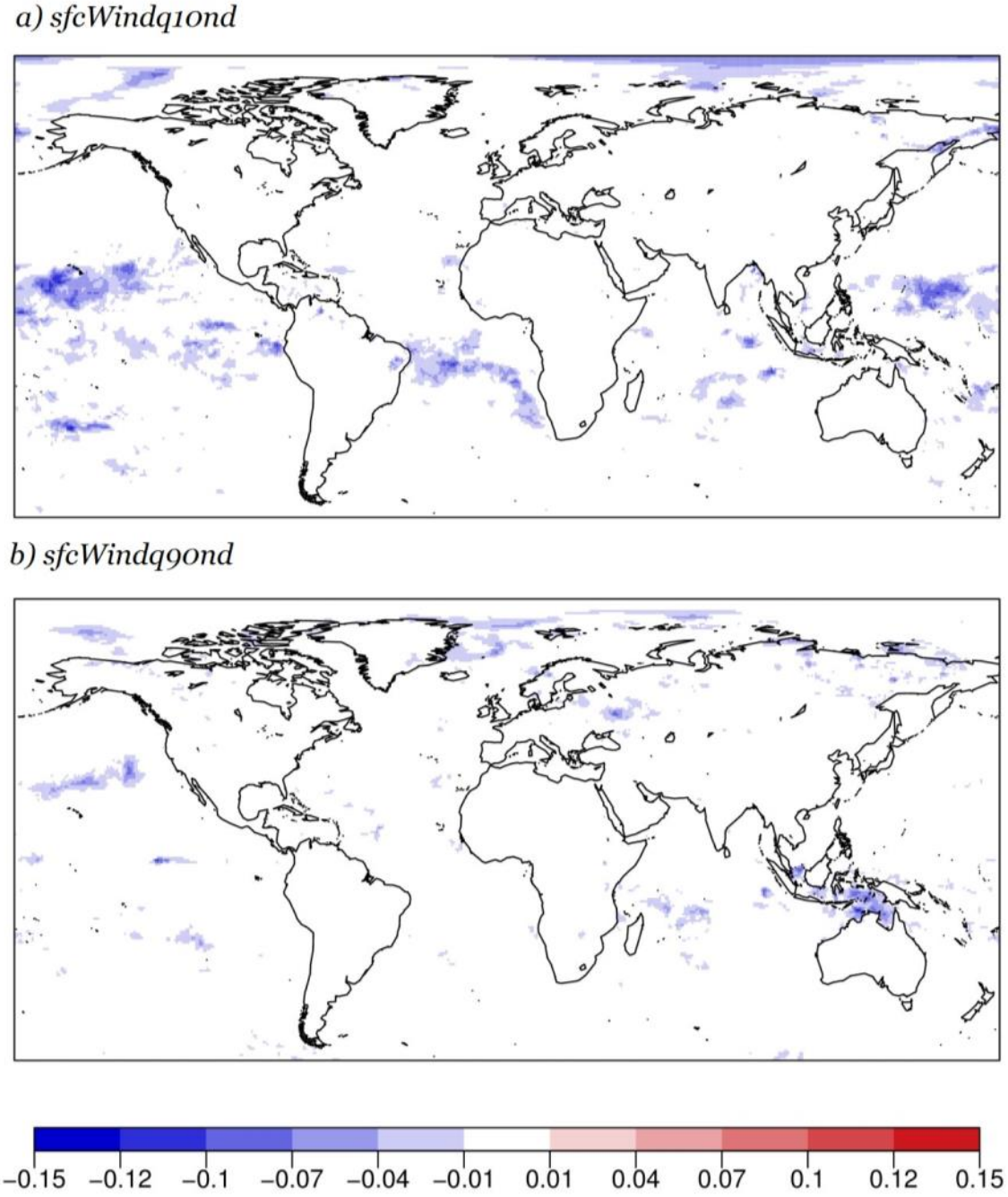

Figure 3.2.1: Bias of wind speed extreme indices: a) Percentage of time under the climatological 10th percentile and b) Percentage of time over the climatological 90th percentile mean in winter (DJF). The bias is computed from ECMWF System 4 and ERA-Interim. Forecasts from System 4 have been initialized the 1st of November in the period of 1981-2012.

3.2.3. Forecast Quality

3.2.3.1. Anomaly Correlation

The anomaly correlation measures the correspondence between the observed and predicted anomalies and it is useful to quantify the potential skill, which is the maximum skill that can be achieved for an index in a particular region given a forecast system.

Figure 3.2.2 shows the correlation between ensemble-mean predictions of the wind speed extreme indices (sfcWindq10nd and sfcWindq90nd) and the indices obtained from ERA-Interim. A perfect agreement along time between the reference and simulated indices would give 1.0. Red areas show that the forecast are on average better than a naive climatological forecast, while blue areas appear where the predictions are worse than the climatology.

The sfcWindq10nd (Figure 3.2.2a) and sfcWindq90nd (Figure 3.2.2b) indices show a positive and significant skill at the 95% confidence level over tropical and oceanic regions. Positive values are located in Northern South-America, North-eastern Brazil, Western Africa and South-eastern Asia. The highest skill in the tropical region is related to anomalies in the tropical sea surface temperatures (SST), in particular in the El Niño-Southern Oscillation (ENSO) region, which is the main source of predictability at seasonal time scales (Kirtman and Pirani, 2009).

Although the predictability of the seasonal prediction systems is limited in extra-tropical latitudes, regions such as Canada or Central North America exhibit also significantly positive correlation for both the sfcWindq10nd and sfcWindq90nd indices. These values can be associated with the ENSO teleconnections and with other sources of seasonal to interannual predictability such as the persistence of the North Pacific decadal oscillation (Lienert *et al.*, 2011).

In Northern Europe the sfcWindq10nd shows positive skill, although it is only significant in the North Sea, Baltic Sea, Northern Scandinavia and South West of the Iberian Peninsula. However, the sfcWindq90nd only displays significant skill in Scandinavia, North East of the Iberian Peninsula and East of the British Islands. This illustrates the asymmetry of the skill between the two indices.

Skilful forecasts of sfcWindq10nd and sfcWindq90nd can help users to save costs related with the vulnerabilities and risks associated with the wind extremes. Therefore, the potential skill of sfcWindq10nd and sfcWindq90nd found in some key regions for the wind industry, with a number of wind farms located there, such as North America, the North Sea or the Northeast of Brazil, demonstrate that there could be useful information for the decision making processes in wind energy operations. However, potential skill does not imply that the forecasts are either useable or useful because they also need to be reliable.

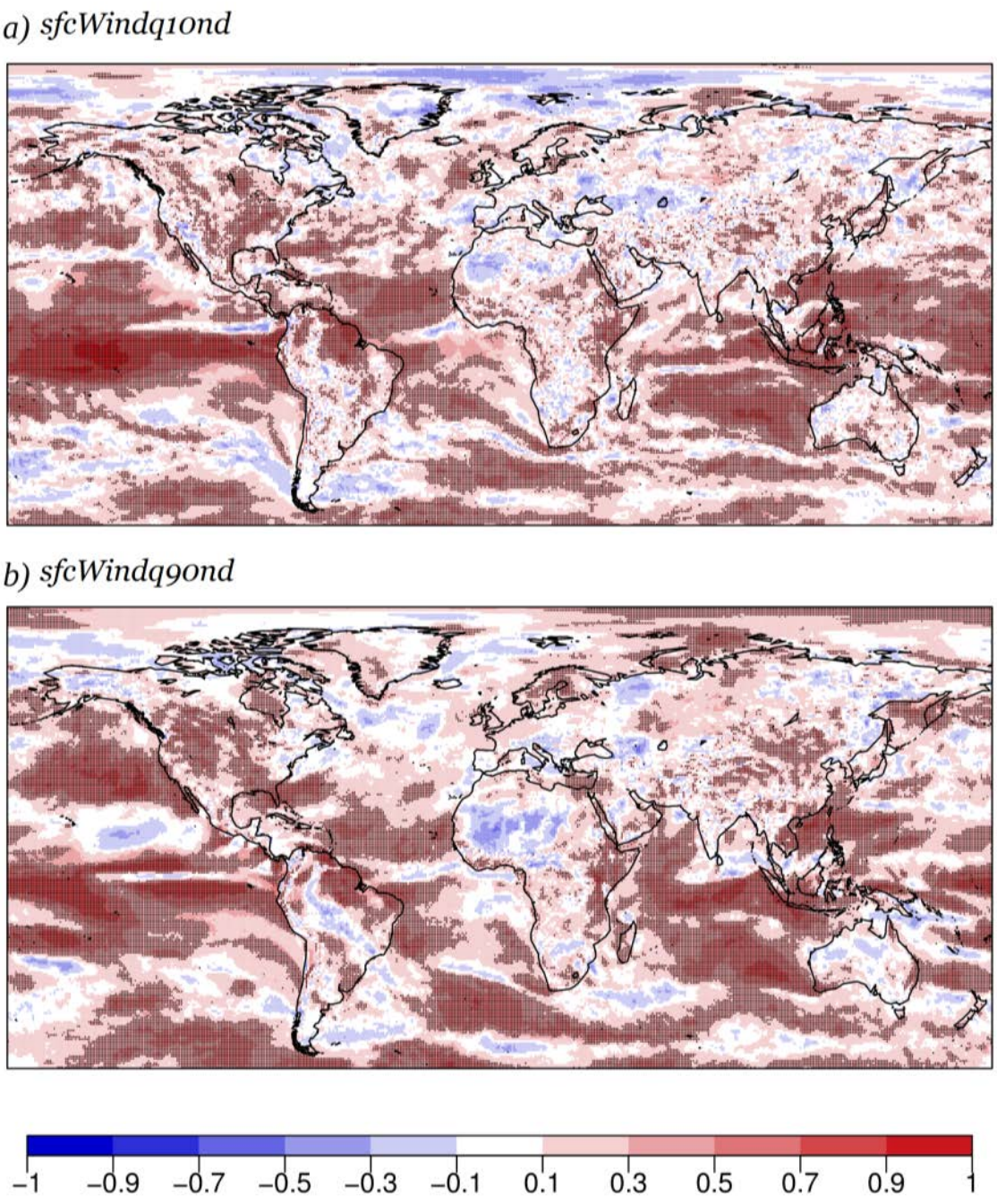

Figure 3.2.2: Correlation of forecasted ensemble mean of: a) Percentage of time under the climatological 10th percentile; b) Percentage of time over the climatological 90th percentile, in winter (DJF) from ECMWF System 4 and Era Interim. Forecasts with System 4 have been initialized on the 1st of November over the period 1981-2012. The dots mark the areas where the skill is significant at the 95% confidence level (i.e. System 4 is significantly more skilful than a climatological forecast).

3.2.3.2. ROC Area

The area under the ROC curve characterises the quality of a forecast system by describing the system's ability to discriminate correctly between occurrence and non-occurrence of events. Two dichotomous events have been considered: below-normal category and above-normal category. The normal category has not been included in this analysis because the

signal-to-noise ratio is small for this category and the climatological distribution becomes nearly as effective as accurately defining slight deviations from climatology (Kumar *et al.*, 2009).

The highest ROC area skill score appears in the tropical regions for both below-normal (Figure 3.2.3) and above-normal (Figure 3.2.4) categories, similar to the correlation (Figure 3.2.2). The ROC skill score of sfcWindq90nd (Figure 3b) is higher than sfcwWindq10nd (Figure 3.2.3a) in North America for the below-normal category. The sfcWindq90nd also exhibits values significantly larger than zero in Northern Europe, particularly in the British Islands, Northern Scandinavia and the North Sea. Nevertheless, the sfcWindq10nd index displays less skill than sfcWindq90nd in Northern Europe for such event.

The above-normal category map (Figure 3.2.4) shows positive skill for North America and East Asia. The sfcWind10nd for such event (Figure 3.2.4a) displays positive skill in some regions of Europe, and the sfcWindq90nd exhibits skilful regions in the Northern part of the Iberian Peninsula, Scandinavia and Eastern part of Mediterranean Sea.

The differences in the ROC area skill score between the below and above normal category (Figure 3.2.3 and Figure 3.2.4) indicate that asymmetry in skill for each wind speed extreme index exist, with regions where the below-normal category is more skilful than those regions for the above-normal category. In addition differences between the sfcWindq10nd and sfcWindq90nd show that asymmetry in skill is also present for the wind speed distribution.

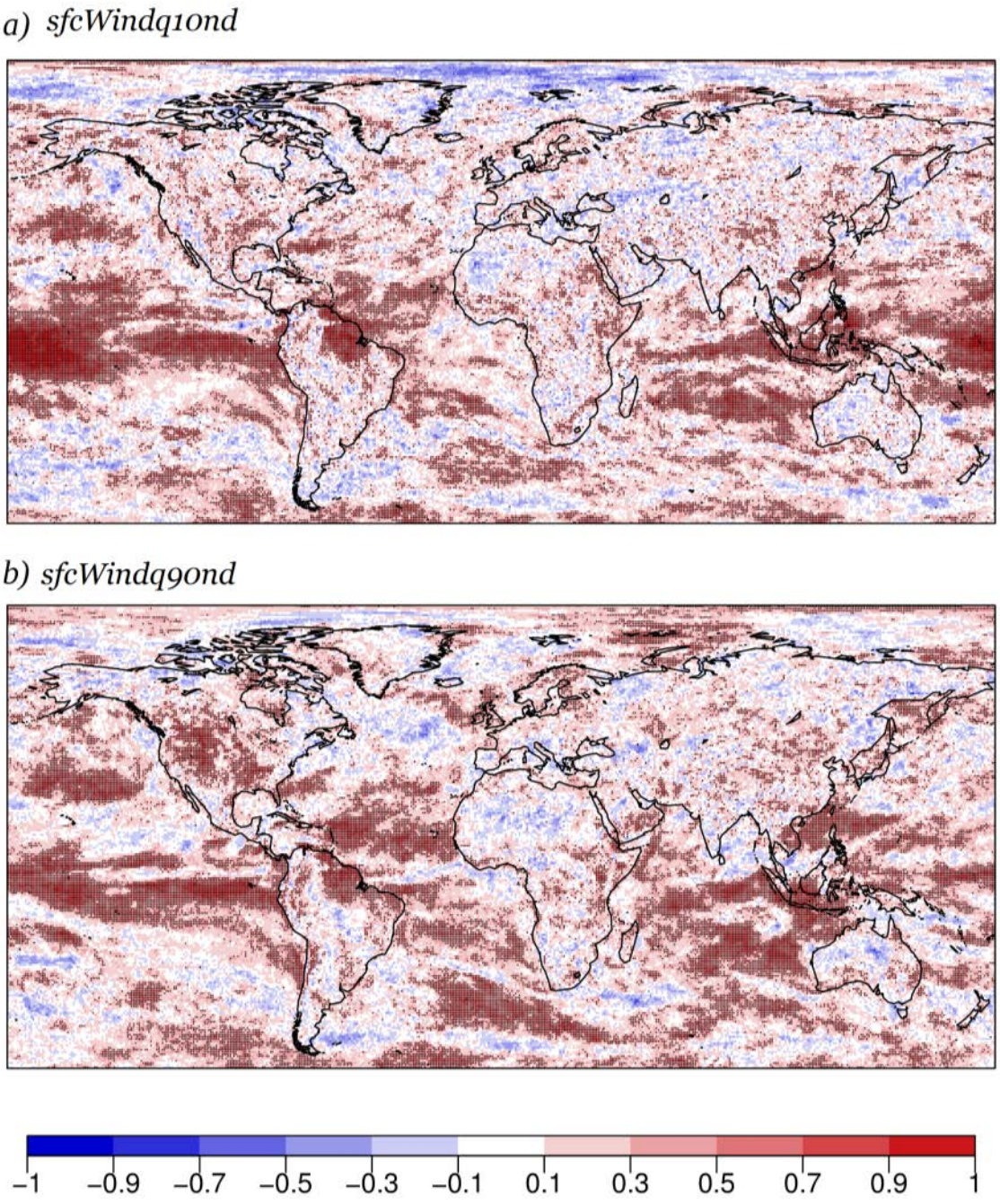

Figure 3.2.3: ROC area skill score for below-normal category of forecasted: a) Percentage of time under the climatological 10th percentile; b) Percentage of time over the climatological 90th percentile, in winter (DJF) from ECMWF System 4 and ERA Interim. Forecasts with System 4 have been initialized on the 1st of November over the period 1981-2012. The dots mark the areas where the skill is significant at the 95% confidence level (i.e. System 4 is significantly more skilful than a climatological forecast).

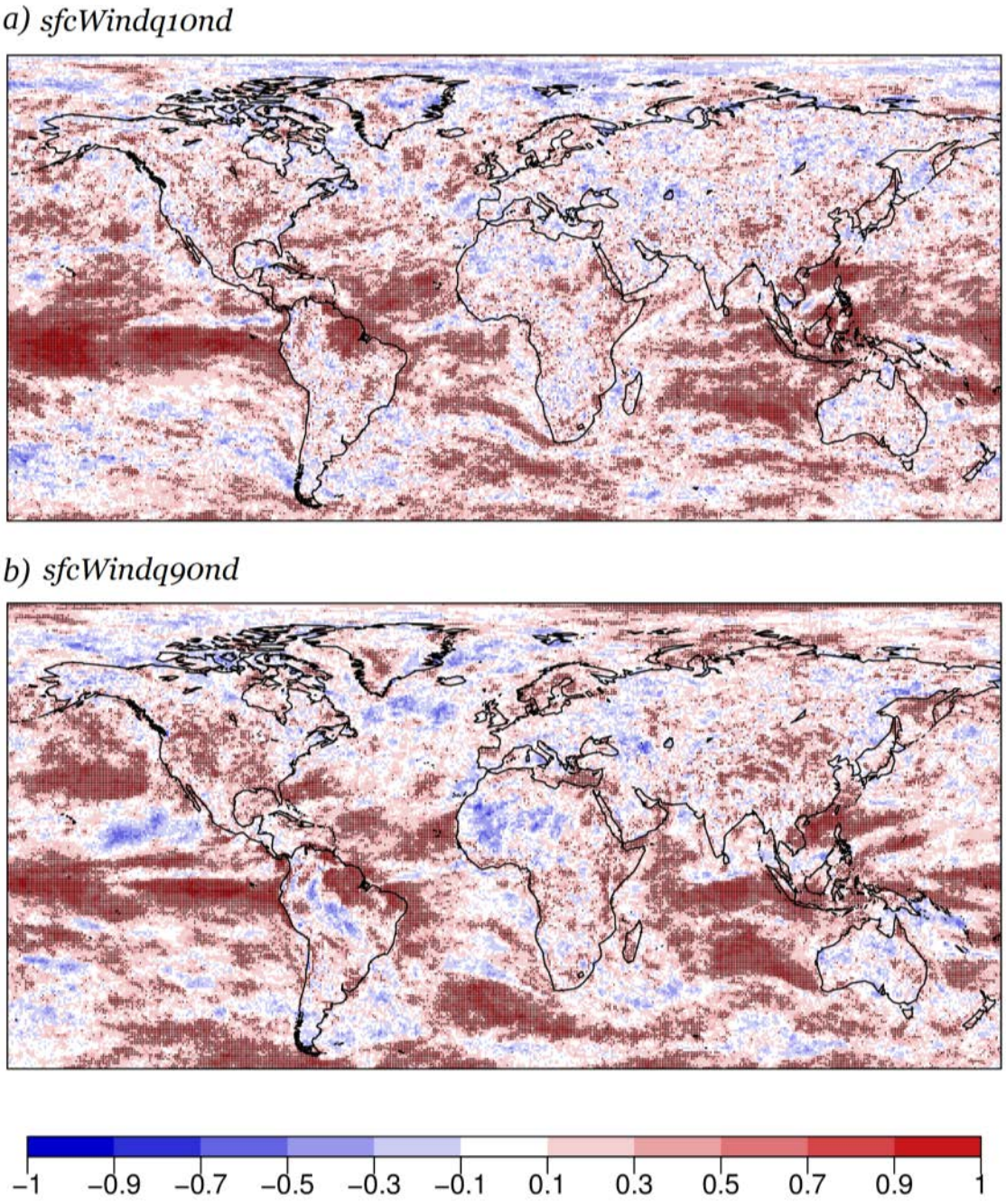

Figure 3.2.4: Same as Figure 3 but for the above-normal category.

3.2.3.3. Spread-error Ratio

The spread-error ratio is a useful estimate to quantify the ability of the ensemble to represent the forecast error in a statistical sense. It is computed as the square root of the ratio between the ensemble variance and the mean squared error of the ensemble mean with the verifying observations. A fair version of the spread-error ratio skill score has been computed (Figure 3.2.5) as in Weigel (2012). For this fair version of the spread-error ratio, the intra ensemble sample variance is inflated to account for the finite ensemble size. In the resulting maps, the red values correspond to a ratio higher than 1, indicating that over-dispersion (under-confidence) of the ensemble exists. The ratio lower than 1 (blue colours), corresponds to under-dispersion (over-confidence).

The spread-error ratio of forecasted sfcWindq10nd (Figure 3.2.5a) and sfcWindq90nd (Figure 3.2.5b) indices display values lower than 1 in several regions around the tropics, as in the tropical Pacific, Southern America, central Africa and South-eastern Asia. In those regions the ensemble is overconfident for both indices. The regions with over-dispersion (under-confidence) in the ensemble appear in the Caribbean Sea, Northern America, China, Western Africa and Eastern Europe for both indices.

Differences in the spread-error ratio are found for the two indices. Under-dispersion of the ensemble appears for the sfcWindq10nd index in Northern Europe and Northern Asia. This can be interpreted as the ensemble to be too narrow to represent the forecast uncertainty, which implies that additional perturbations should be introduced for the ensemble to be representative of the true uncertainty. Under-confidence is found over the Iberian Peninsula and the North of Scandinavia for the sfcWindq10nd index (Figure 3.2.5a). These differences show an asymmetry in the skill for the 6-hourly wind speed distribution from which the sfcWindq10nd and sfcWindq90nd indices have been computed.

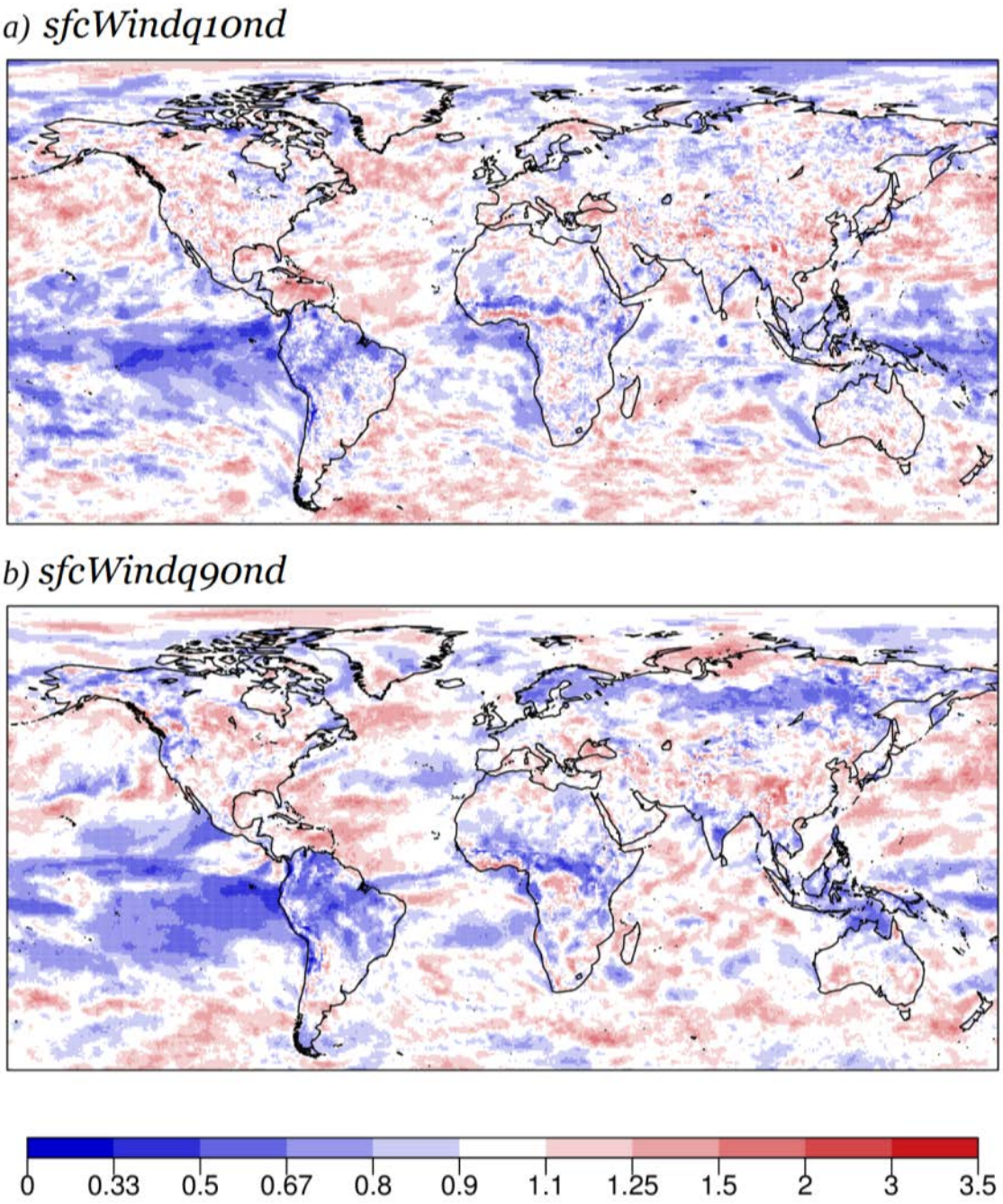

Figure 3.2.5: Fair spread-error ratio of a) Percentage of time under the climatological 10th percentile; b) Percentage of time over the climatological 90th percentile, in winter (DJF) from ECMWF System 4 and ERA Interim. Forecasts with System 4 have been initialized on the 1st of November over the period 1981-2012.

3.2.3.4. CRPSS

The continuous ranked probability skill score (CRPSS) is a skill score based on the CRPS that is defined as the integrated squared difference between the cumulative forecast and observation distribution functions. It can be interpreted as a probabilistic generalisation of the mean absolute error. A perfect forecast would have a CRPSS equal to 1. Negative values of CRPSS (blue colours) imply that the skill of the estimated forecast probabilities is worse than

the use of climatological frequencies as forecast. Positive values of CRPSS (red colours) indicate that the forecast is better than the climatological probabilities.

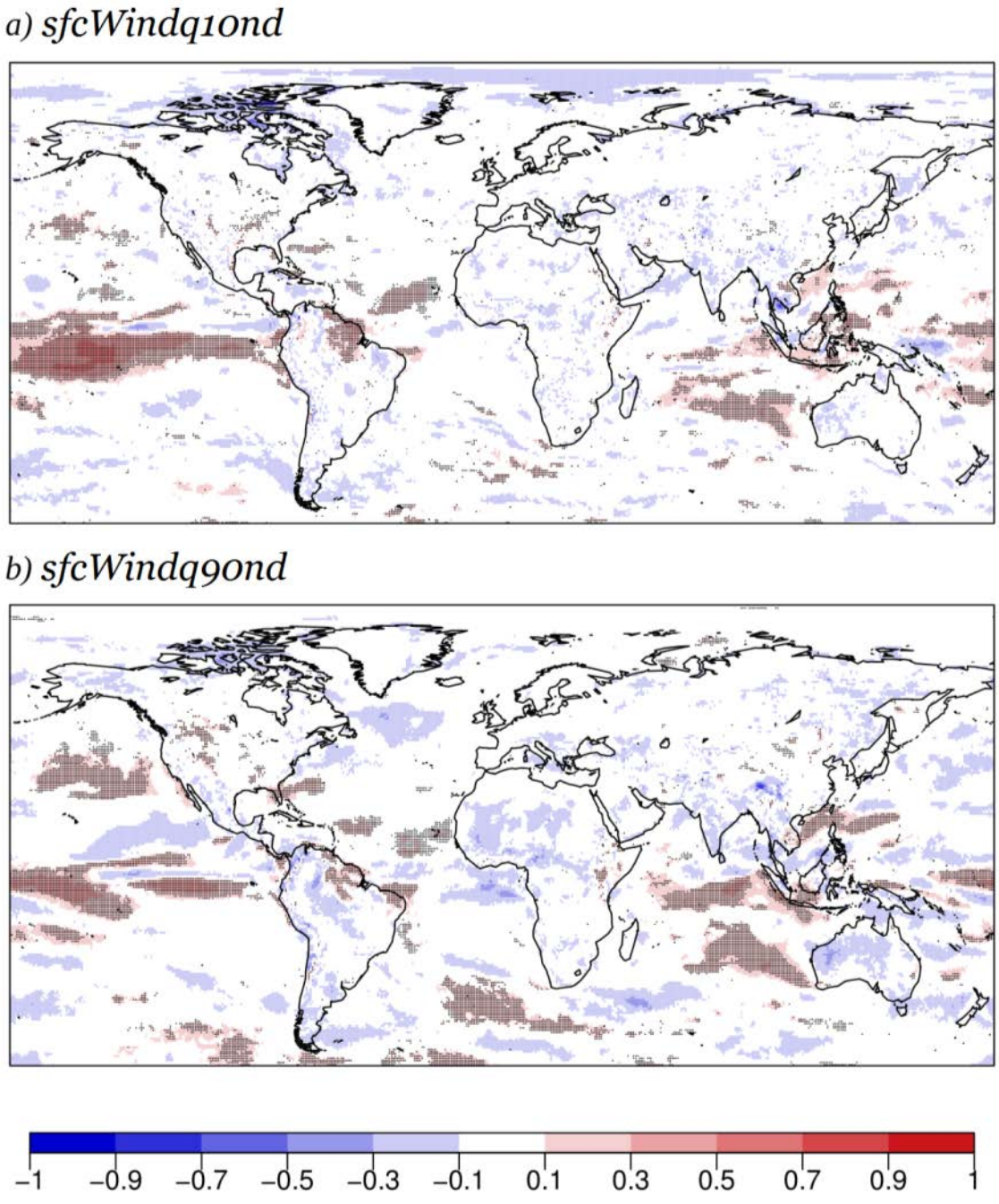

Figure 3.2.6: Fair continuous ranked probability skill score (CRPSS) of: a) Percentage of time under the climatological 10th percentile; b) Percentage of time over the climatological 90th percentile, in winter (DJF) from ECMWF System 4 and ERA Interim. Forecasts with System 4 have been initialized on the 1st of November for the period 1981-2012. The dots mark the areas where the skill is significant at the 95% confidence level (i.e. System 4 is significantly more skilful than a climatological forecast).

In this section a fair version of CRPSS was used to reward ensembles that behave as though their members and the verifying observation are sampled from the same distribution (Ferro, 2014).

As in the previous examples, the tropics are the region with highest CRPSS (up to 0.5), particularly northern South America and the Northeast of Brazil. Some of central North America retains a positive and significant skill at the 95% confidence level for both sfcWindq10nd and sfcWindq90nd wind speed extreme indices (Figure 3.2.6). But nearly all other continental regions, including Europe, have not significant values.

The differences that have been found between Figure 3.2.6 and Figure 3.2.2 indicate that although the potential skill to predict the wind extremes indices is high in many regions around the world, predicting the full distribution of the indices is still a challenge. We are currently working on the calibration of the predictions to reduce the negative skill values.

We have introduced two modifications from the DOW in this deliverable:

- Both the upper and lower percentile wind speeds have been investigated, despite that in the DOW only the upper wind speeds are mentioned. As the wind turbine operating thresholds are instantaneous wind speed values, they cannot be compared directly to seasonal mean values. For this reason, the upper and lower extreme wind speed percentiles were used.
- IC3 collaborates with several stakeholders from the wind industry which have shown interest on the wind speed extreme indices. Therefore, in this deliverable we only focus our analysis on these wind speed extremes. However such indices can be estimated from different variables such as solar radiation and different thresholds, an option that will be explored in the future.
- Two Climate Information Indices (CIIs) have been developed for the renewable energy sector with a focus on the upper and lower wind speed thresholds, as suggested by the industry. A forecast quality assessment has been made of the wind speed CIIs corresponding to the percentage of time in a season under the 10th climatological percentile and exceeding the 90th climatological percentile.

The Climate Forecasting Unit at IC3 assessed the CIIs relevant for wind power stakeholders, which are predominantly based on wind speed. The percentage of time below and above the 10th and 90th percentiles respectively (considered as extremes values) show a small bias because the CII has been computed based on a relative threshold from the model's climatology. For that reason, the bias correction is not essential for the predicted indices. An assessment of the skill of the percentage of time above and below the 10th and 90th percentile was made using several forecast quality measures. The differences in the skill between the two indices reveal that an asymmetry in the skill for the wind speed distribution is present. As expected, the more robust the assessment, the lower the observed skill, with CRPSS showing the lowest values. In general, the highest skill is found over the tropics and the central US region for both indices.

Key Points: Assessing Skill in Wind Speed

- Indices with thresholds defined relative to the simulated and observed climatology respectively are advantageous in that such indices are less affected by systematic model biases than indices defined relative to absolute thresholds.
- While there is considerable potential skill in predicting wind speed, predicting the full distribution is still challenging. Statistical recalibration of forecasts may help to improve predictions of wind speed related indices.

3.3. Heating and cooling degree days

3.3.1. Data and Methods

We present seasonal forecasts of the European Centre for Medium-range Weather Forecasts' (ECMWF) System 4 (Molteni *et al.,* 2011) forecasting system for the summers (JJA) from 1981-2014 and winters (DJF) from 1981-2013. We derive seasonal heating degree day (HDD) and cooling degree day (CDD) forecasts from bias corrected daily temperature time series of ECMWF System 4 initialized in May for summer forecasts and November for winter forecasts. The daily temperature time series are bias corrected against daily temperature from the ERA Interim reanalysis (Dee *et al.,* 2011). Daily bias correction is estimated using a local linear regression smoothing of the daily observed and forecast climatologies as suggested by Mahlstein *et al.* (2015). Bias correction is performed in leave-one-out cross validation mode using data from 1981-2010 to estimate the correction factors.

We use the following definitions for heating and cooling degree days:

$$HDD = 1/n \sum_{i=1}^{n} \begin{cases} 18 - T_i, & \text{if } T_i < 12 \\ 0, & \text{otherwise} \end{cases}$$

$$CDD = 1/n \sum_{i=1}^{n} \begin{cases} T_i - 22, & \text{if } T_i > 22 \\ 0, & \text{otherwise} \end{cases}$$

In Figure 3.3.1, we show the bias in winter HDD and summer CDD. Biases in the indices prevail after bias correction of the daily time series due to systematic biases in the distribution of forecast daily temperature and daily temperature variability not accounted for by the bias correction. These biases are in general larger for HDD than for CDD. Please note, however, that also the climatological mean for HDD is larger than for CDD due to the different definition and thresholds. Areas with positive and negative bias in HDD and CDD are forecast roughly cancel out, indicating that the daily bias correction works well apart from sampling uncertainty and the above sources of errors not accounted for.

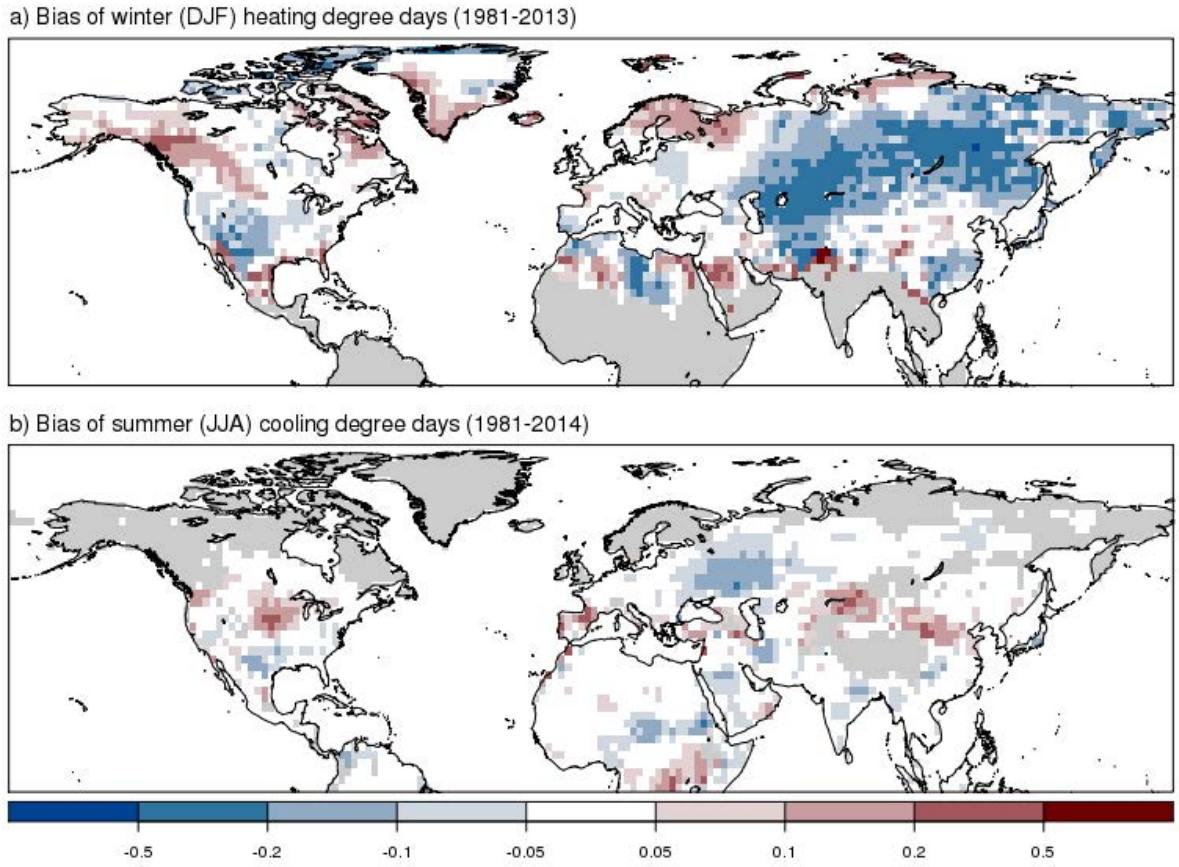

Figure 3.3.1: Bias of winter (DJF) heating degree days (a) and summer (JJA) cooling degree days (b) in units of degree days / day. Areas where zero HDD or CDD are observed for more than one third of the years are masked in grey.

We validate seasonal time series of HDD and CDDs against HDDs and CDDs derived from daily temperature time series of the ERA Interim reanalysis. Areas with no HDD and CDD occurrences are masked out as the forecasts are trivial and the climatological forecast will always be at least as good as a dynamical forecast. Therefore, we only present forecasts quality metrics for areas where zero HDD (CDD) are observed in at most one third of all the years. Using this criterion we avoid difficulties in interpreting forecasts of rare events and we also ensure that the terciles for tercile forecasts are well defined.

3.3.2. Seasonal Forecast Quality of Indices Related to Energy Consumption

3.3.2.1. Correlation

In Figure 3.3.2 we show the correlation of the ensemble mean seasonal HDD and CDD time series with the respective time series from ERA-Interim. We find significant correlation (at the 5% level) in large parts of the northern hemisphere and only limited areas with correlations below zero (indicating negative correlation skill compared to a climatological forecast without interannual variability). However, correlations over land rarely exceed 0.6 for HDD in winter, and regions of strong correlation over land in summer for CDDs are constrained to subtropical and tropical areas. In Europe we find significant but weak correlation for HDD in winter over the British Isles and parts of central western Europe. In

summer, southern and eastern Europe exhibits significant correlation exceeding 0.6 in the eastern Mediterranean.

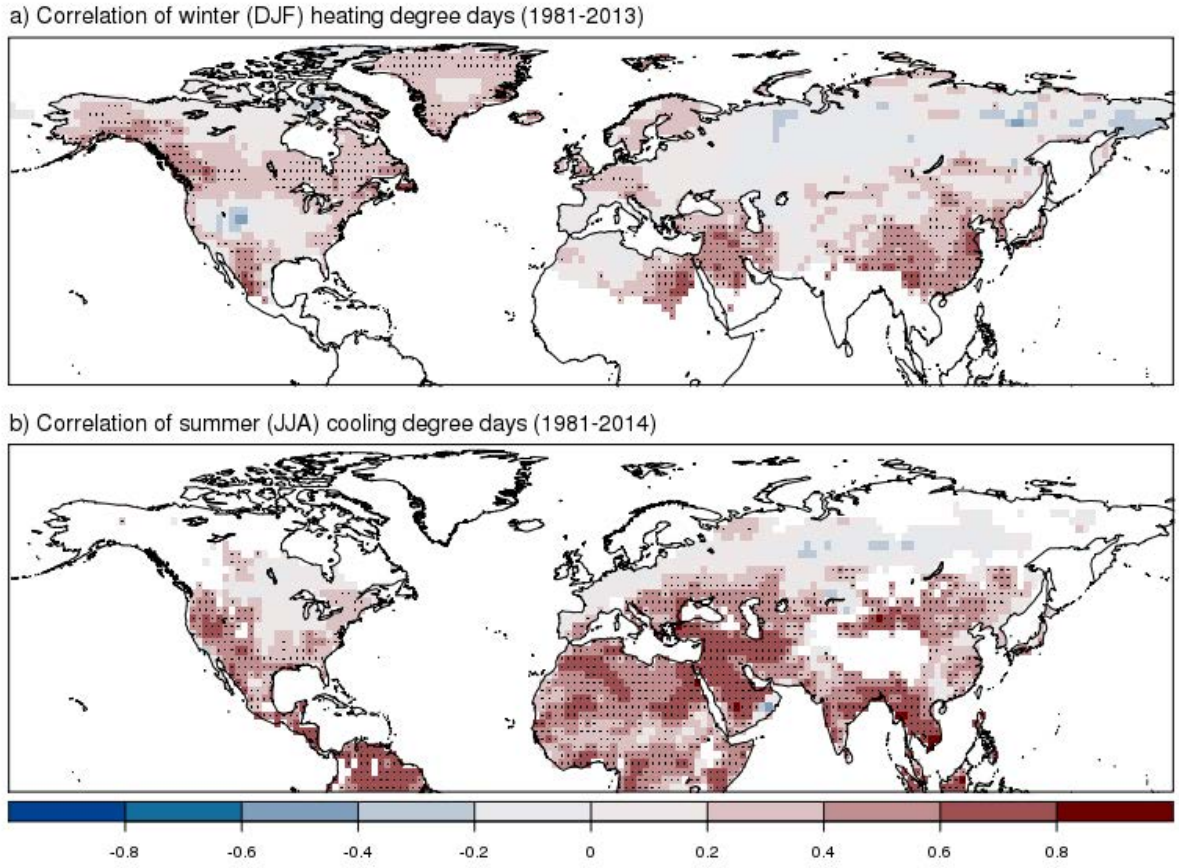

Figure 3.3.2: Correlation of winter (DJF) heating degree days (a) and summer (JJA) cooling degree days (b). Stippling indicates correlations significantly (at the 5% level) larger than zero.

3.3.2.2. Ranked Probability Score

In Figure 3.3.3 we show the ranked probability score (RPSS) for tercile forecasts. The RPSS characterizes to what extent we are able to probabilistically forecast above normal, normal and below normal conditions. RPSS larger than zero indicates forecasts that are skilful compared to a climatological forecast (i.e. a constant 33.3% probability for the three tercile categories), RPSS smaller than zero indicates forecasts that are worse than a climatological forecast. In analogy to the correlation results shown in Figure 3.3.2, we find limited areas of significant RPSS for HDD in winter (Figure 3.3.3a). For CDD in summer, we find significant RPSS throughout the subtropics and tropics as well as in the western USA and in China (Figure 3.3.3b). In Europe, significant RPSS is found mainly in the eastern Mediterranean in summer.

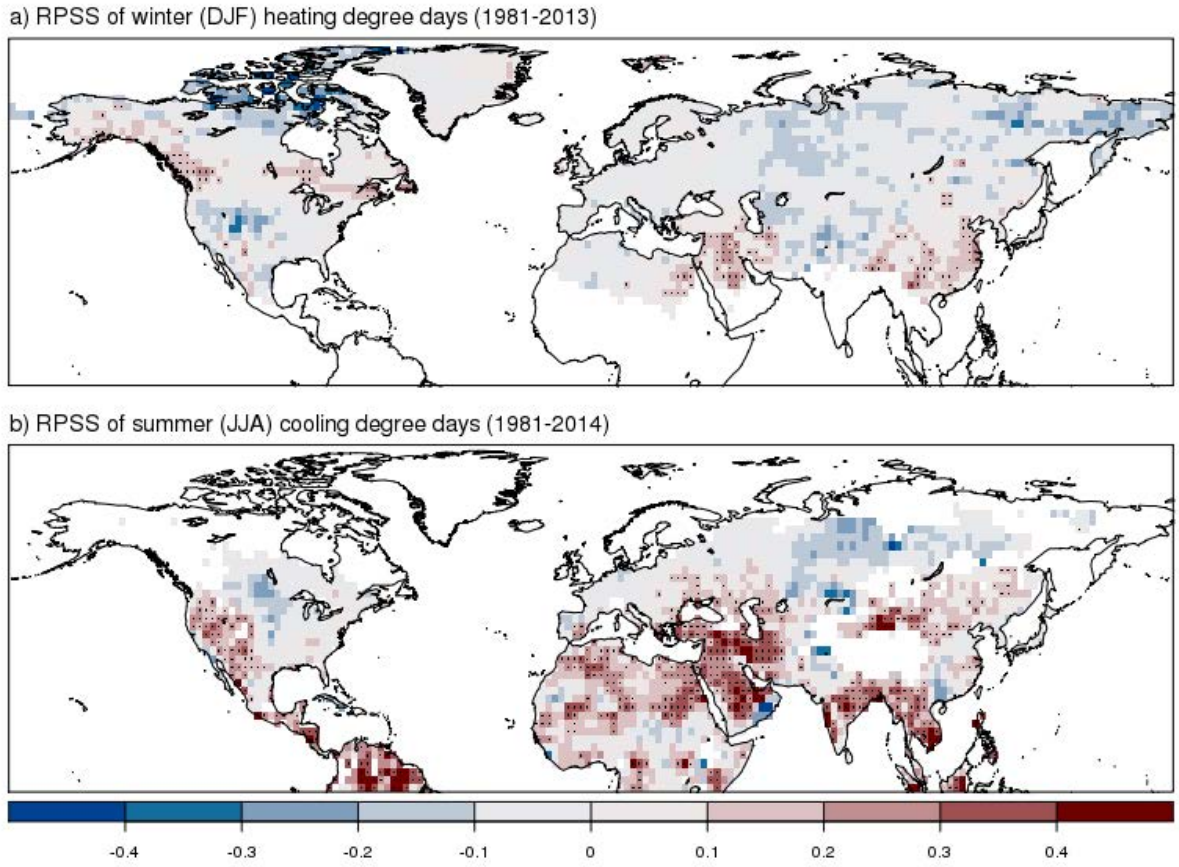

Figure 3.3.3: Ranked probability skill score for tercile forecasts of winter (DJF) heating degree days (a) and summer (JJA) cooling degree days (b). Stippling indicates correlations significantly (at the 5% level) larger than zero.

3.3.2.3. Continuous Ranked Probability Score

In Figure 3.3.4 we show the continuous ranked probability score (CRPSS). In contrast to the RPSS that compares probabilistic category forecasts, the CRPSS measures the absolute discrepancy between forecast probability distribution and observations and is thus sensitive to both the bias and the lack of reliability of the forecast. Therefore, the CRPSS is a more strict score for which it is harder to score well.

CRPSS for winter HDD forecasts is positive only in a few areas in the northern subtropics, whereas two distinct regions with strongly negative CRPSS are found in the western US and central Asia (Figure 3.3.4a). In contrast, CRPSS for summer CDD is significantly positive for large areas in the subtropics and tropics, interspersed with smaller regions of negative CRPSS (Figure 3.3.4c).

The areas of negative CRPSS for HDD and CDD forecasts are due to a combination of remaining biases in the forecast indices (Figure 3.3.1) and lack of correlation (Figure 3.3.2). Both of which can in principle be accounted for by a recalibration of the HDD and CDD forecasts. Such a recalibration should then result in at least zero skill. Here we use the climate conserving recalibration of Weigel *et al.* (2009). Indeed, most of the areas of strongly negative CRPSS disappear when using the recalibrated forecasts (Figure 3.3.4b, d). As we compute the recalibration (like the bias correction) in cross-validation mode, there are remaining areas of negative CRPSS in particular for CDD forecasts in Siberia. These are

due to the uncertainty in estimating the recalibration parameters from a finite sample. Also CRPSS skill is generally slightly reduced for the recalibrated forecasts compared to the bias corrected forecasts due to the additional sampling uncertainty of the parameters for the recalibration. This illustrates that post-processing of forecasts tends to affect the skill, positively through the reduction of systematic biases but also negatively through the sampling uncertainty of the parameters of the post-processing (see also Gangstø *et al.*, 2012)

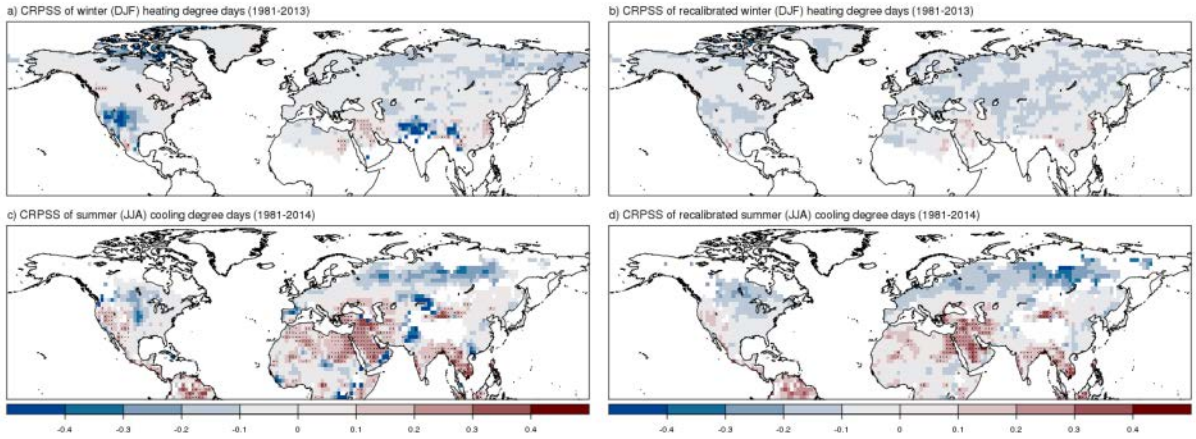

Figure 3.3.4: Continuous ranked probability skill score for winter (DJF) heating degree days (a) and summer (JJA) cooling degree days (b). Stippling indicates correlations significantly (at the 5% level) larger than zero.

3.3.2.4. Spread to Error Ratio

To assess if the forecasts of indices are reliable, that is if forecast probabilities correspond to observed frequencies, we compute the spread to error ratio. The spread to error ratio is only a necessary but not sufficient condition for reliability. Also it has been found that spread error correlation is low even for perfectly reliable ensembles (e.g. Hopson, 2014) and thus the spread to error ratio will be subject to considerable sampling uncertainty. Nevertheless, we use it here as a quick and simple metric to assess forecast reliability.

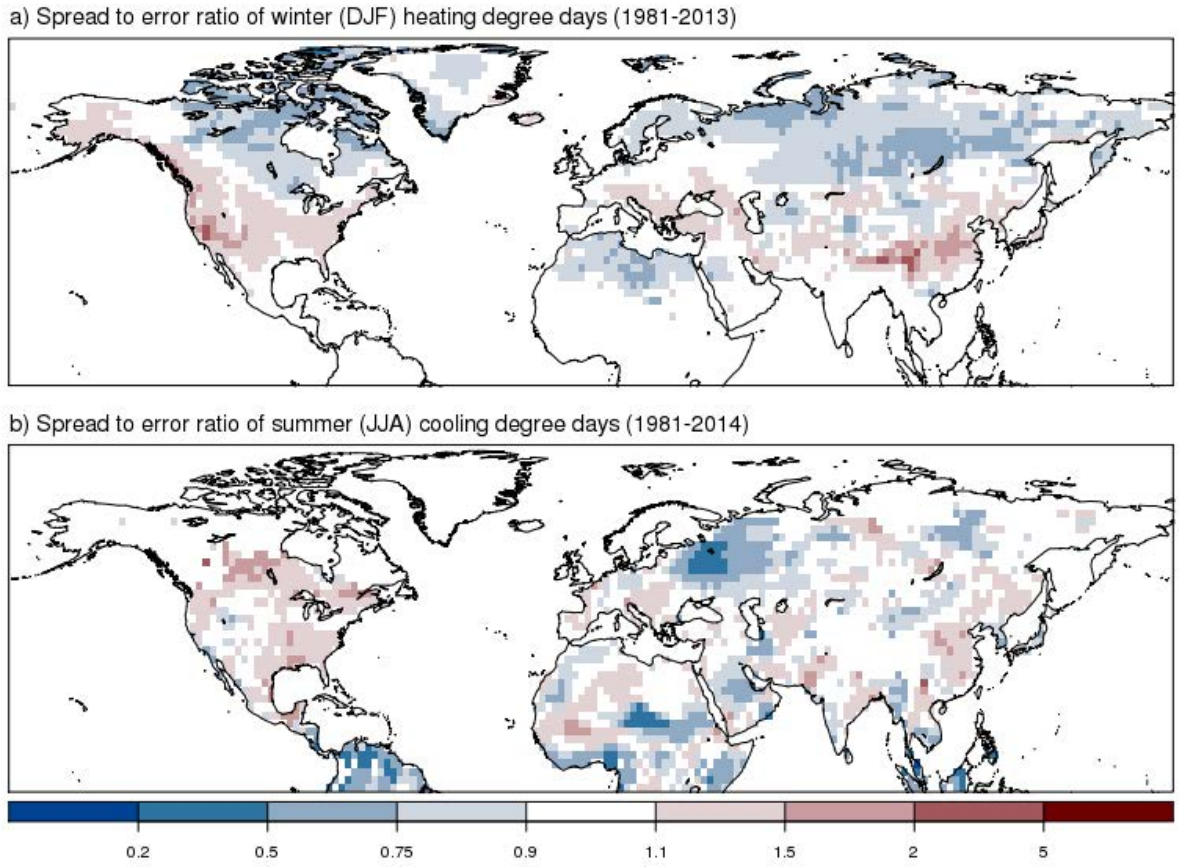

Figure 3.3.5: Spread to error ratio for winter (DJF) heating degree days (a) and summer (JJA) cooling degree days (b).

In Figure 3.3.5 we show the spread to error ratio of HDD forecasts for the winter season and CDD forecasts for summer. HDD forecasts are found to be overconfident (i.e. spread to error ratios smaller than 1) in large parts of the northern mid-latitudes, whereas in the western US and central Asia, forecasts are found to be over-dispersive. In contrast, the pattern of spread to error ratios for CDD forecasts for the summer season is much patchier. We find indication of overconfidence in north-western Russia and tropical areas, and indication of over-dispersion elsewhere. The regions of overconfidence in HDD and CDD forecasts do not directly relate to regions of remaining biases in HDD and CDD (see Figure 3.3.1). Therefore, the lack of reliability cannot be explained by remaining biases in the system, but is either due to sampling variability or is in fact a property of the forecasting system.

3.3.3. Sources of Predictability

In the following, we analyse to what extent predictability in HDD and CDD is driven by a trend in mean temperature and thus a trend in the indices. Such a linear trend is reflective to the global warming signal and is likely brought about by changes in external forcings. To analyse this, we contrast correlation of HDD and CDD with the correlation of detrended HDD and CDD. Skill in detrended forecasts is reflective of skill in predicting interannual variability independent of such low frequency changes. While our aim is to predict interannual variability, it is important to note that there is intrinsic value in ensemble forecasts even in the absence of interannual skill as such forecasts help to better characterize the current climate

and current climate risks than a climatological assessment that does not account for long-term trends.

We find significant correlations in detrended HDD throughout large parts of the northern hemisphere (Figure 3.3.6a). In particular, correlations of detrended HDD are in most parts comparable to correlation of HDD with trends (Figure 3.3.2), and therefore suggest that most of the predictability in winter HDD in the northern hemisphere is due to long-term trends (and thus likely due to changes in external forcings). For CDD in summer, however, the situation is remarkably different (Figure 3.3.6c). While the correlations of detrended CDD in the tropics and areas of the northern mid-latitudes are comparable to the correlations of CDD with trends, in large areas in the subtropics, Eastern Europe and central Asia and the western US, this is not the case (red shading in Figure 3.3.6d). In these areas we can thus conclude that there is predictability beyond the long-term trend. Further investigation will have to reveal whether sources of this interannual predictability can be identified (the El Niño - Southern Oscillation being a likely candidate, see Pepler *et al.*, 2015).

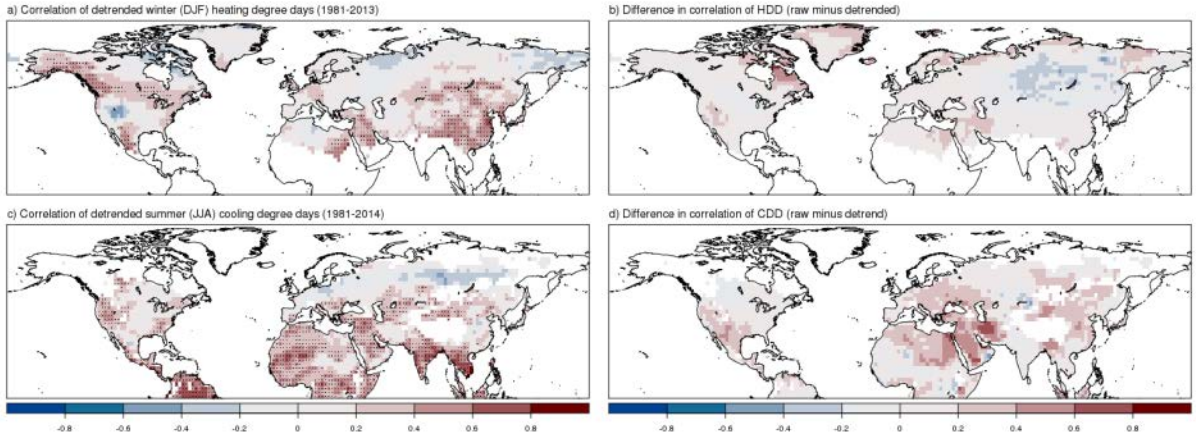

Figure 3.3.6: Correlation skill for detrended winter (DJF) heating degree days (a) and detrended summer (JJA) cooling degree days (c) along with the difference in correlation between the correlation in the index with trend (see also Figure YY) minus the correlation in the detrended index (b, d). Stippling in a and c indicates correlations significantly (at the 5% level) larger than zero.

3.3.4. Conclusions

Heating and cooling degree days are defined with respect to absolute temperature thresholds. Therefore, the time series of daily mean temperature as forecast by the forecasting system have to be bias corrected (calibrated) previous to analysis. The skill in forecasting heating and cooling degree days, however, is mostly independent of the bias correction method used (not shown) and simple bias correction methods are sufficient.

While there is some skill in forecasting cooling degree days in southern Europe, elsewhere in Europe and for heating degree days in winter, skill is limited. In other areas, e.g. the tropics and the western US, considerable prediction skill has been identified. Remaining areas of negative skill (Figure 3.3.3 and Figure 3.3.4) indicate the potential for further improvement of the forecasts. In areas of negative skill, we should rather resort to a constant, climatology-based forecast than issue misleading forecasts. Post-processing of the forecasts of climate indices can help to achieve that.

Key Points: Cooling and Heating Degree Days

- Bias correction of daily forecast series is important for climate indices defined with respect to absolute thresholds.
- Skill in forecasting heating and cooling degree days is limited with enhanced skill for cooling degree days in southern Europe in summer.

3.4. Indices Relevant for Wine Production

A specialized set of CII relevant for the wine producing stakeholders were calculated (Table 3.4.1) from ERA-Interim reanalysis (Dee *et al.*, 2011) and 15 members of System 4 seasonal forecasts (S4 for the remainder of this section, Molteni *et al.*, 2011). These CII were computed for the period 1981-2010 and for the European domain at 0.7º resolution. The indices are mostly based on temperature and/or precipitation accumulated between April and September, thus the seasonal forecast initialized in March (i.e. with a lead time of one month) were used to derive the CII forecasts.

Table 3.4.1: Climate indices for wine production. Definition and usefulness

Index	Definition	Utility
Growing season precipitation (GSP)	$GSP = \sum_{Apr}^{Sep} P$	One of the most discriminating climatic variables (Blanco-Ward *et al.* 2007)
Hydrothermal Index (HyI)	$Hyl = \sum_{Apr}^{Sep} T_{mean} \times P$	Considers both precipitation and temperature regimes for estimating the risk of downy mildew disease (Branas *et al.* 1946) HyI < 2500°Cmm - low HyI > 5100°Cmm - high
Selianinov Index (SI)	$SI = \sum_{Apr}^{Sep} P / (T_{\mathrm{mean}} - 10)$	Measure of hydric regime, i.e. effectiveness of precipitation in the growing season (Magalhães 2008). SI<1 - insufficient 1< SI < 3 - normal SI >3 - excessive

Cool night index (CI)	$CI = \sum_{Sep} {}^{T_{min}}/_{30}$	Provides a relative measure of ripening potential, being equal to the average minimum temperature during the month before harvest (Tonietto *et al.* 2004) • very cool nights (CI ≤ 12°C), • cool nights (12 < CI ≤ 14°C) • temperate nights (14 < CI ≤ 18°C) • warm nights (CI > 18°C)
Growing season suitability (GSS)	$GSS = {}^{\sum_{Apr}^{Sep} d_{Tmean>10ºC}}/_{\sum_{Apr}^{Sep} d}$	Useful as a zoning tool (Santos et al 2012)
Huglin Heliothermal Index (HI)	$HI = \sum_{Apr}^{Sep} \frac{(T_{mean} - 10) + (T_{\max} - 10)}{2} d$	Useful as a zoning tool (Huglin 1978) HI ≤ 1500 too cool 1500<HI ≤ 2100 temperate 2100<HI ≤ 2700 warm 2700<HI ≤ 3000 very warm HI > 3000 too warm
Growing Degree Day (GDD) or Winkler index	$GDD = \sum_{Apr}^{Sep} (T_{mean} > 10ºC - 10)$	Useful as a zoning tool to differentiate between grape varieties and climate (Winkler *et al.* 1974)

3.4.1. Bias Correction

Systematic errors were identified through the analysis of maps of bias determined by the difference between the indices calculated with S4 ensemble mean and ERA-Interim 30 year climatology. Given the large bias for all indices, bias correction was carried out using quantile mapping of the ensemble mean daily precipitation, maximum, minimum and mean temperature. For each day, a 31 day moving window (i.e. ± 15 days centred on the day of interest) is used as in Wilcke *et al.* (2013). Thus each ECDF is determined from 930 values, taking into account autocorrelation and the interannual variability of each day. In the case of

precipitation, the ensemble mean dry day frequency is corrected to be equal to the dry day frequency of ERA Interim. In addition to the ensemble mean correction, the ensemble spread is also corrected following Barnston *et al.* (2015).

3.4.2. Forecast Quality

3.4.2.1. Growing Season Precipitation

The Growing Season Precipitation (GSP), as the accumulated precipitation from April to September, i.e. the growing season, can be used as a discriminating index for vineyard distribution. Areas with GSP above 600mm are usually considered excessively humid, whereas areas with GSP bellow 200mm are extremely dry. Although considerably reduced in relation to a simple bias correction with the climatology mean, significant biases still prevail after the bias correction of the daily time series (Figure 3.4.1a). Whereas the uncorrected forecast showed a dry bias, a wet bias is found after bias correction. This illustrates the difficulty in bias correction of the precipitation. The anomaly correlation (Figure 3.4.1b), illustrates S4's low skill in describing the correspondence between ERA Interim and predicted anomalies. The quantile mapping bias reduction technique usually does not improve correlation and in some instances it may even degrade it. Thus, no improvement in relation to the simple bias correction is observed. Positive but low values are found in the western part of the Iberian Peninsula and British Isles. The highest correlation skill is found in northern France and southern Sweden. The western Iberia, the Mediterranean and central Europe all have negative anomaly correlations.

Positive skill in forecasting the lower GSP tercile is found in western Iberia and north-western France, indicating some ability to forecast drought in these areas (Figure 3.4.1c). In the remaining areas in Europe the forecast has no or very little skill. North-western France also shows some predictability for precipitation seasons wetter than the 3rd tercile (Figure 3.4.1d). The spread to error ratio (Figure 3.4.1e) displays values well above 1 for most of the continent indicating over-dispersion/under-confidence of the ensemble, thus the ensemble is too broad and over-emphasizes forecast uncertainty. In central and southern Iberian Peninsula under-dispersion occurs. The negative continuous ranked probability skill score (Figure 3.4.1f) implies that using forecast probabilities is worse than using climatological frequencies as a forecast. This lack of skill is likely due to residual biases and improved post-processing of the forecasts may result in continuous ranked probability skill scores closer to potential skill as indicated by the correlation (Figure 3.4.1b).

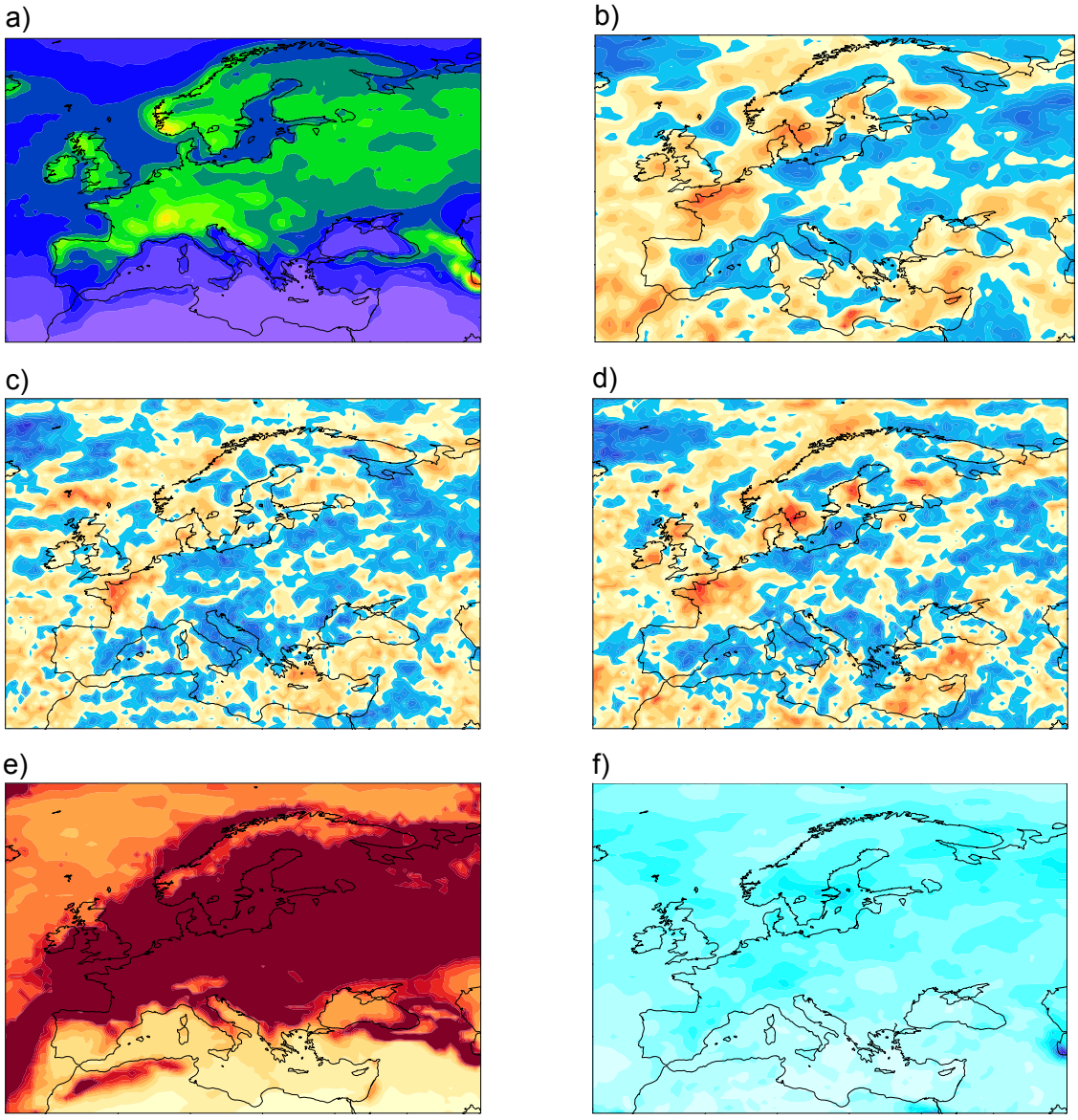

Figure 3.4.1: Growing season precipitation (a) Bias, (b) Anomaly Correlation, ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio, (f) Continuous ranked probability skill score

3.4.2.2. Hydrothermal Index

According to Branas *et al.* (1946) the hydrothermal index can be used to determine the risk of the development of mildew. According to the climatology, all of central Europe has a high risk of mildew occurrence and only southern Greece, Sicily, Sardinia and southern Iberia Peninsula are low risk areas. A relatively low bias, circa 10%, occurs from Iberia to central Europe. Higher biases occur in Scandinavia, Italy and the Balkans (Figure 3.4.2a). The anomaly correlation and the ROCs maps (Figure 3.4.2b, c and d) are very similar to the GSP anomaly correlations and ROCs, indicating that the precipitation distribution dominates the index skill. The spread to error ratio (Figure 3.4.2e) displays values between 0.4 and 0.5 for most of the continent indicating under-dispersion/over-confidence of the ensemble, thus the ensemble is too narrow and does not represent forecast uncertainty. As in GSP, the continuous ranked probability skill score (Figure 3.4.2f) is negative.

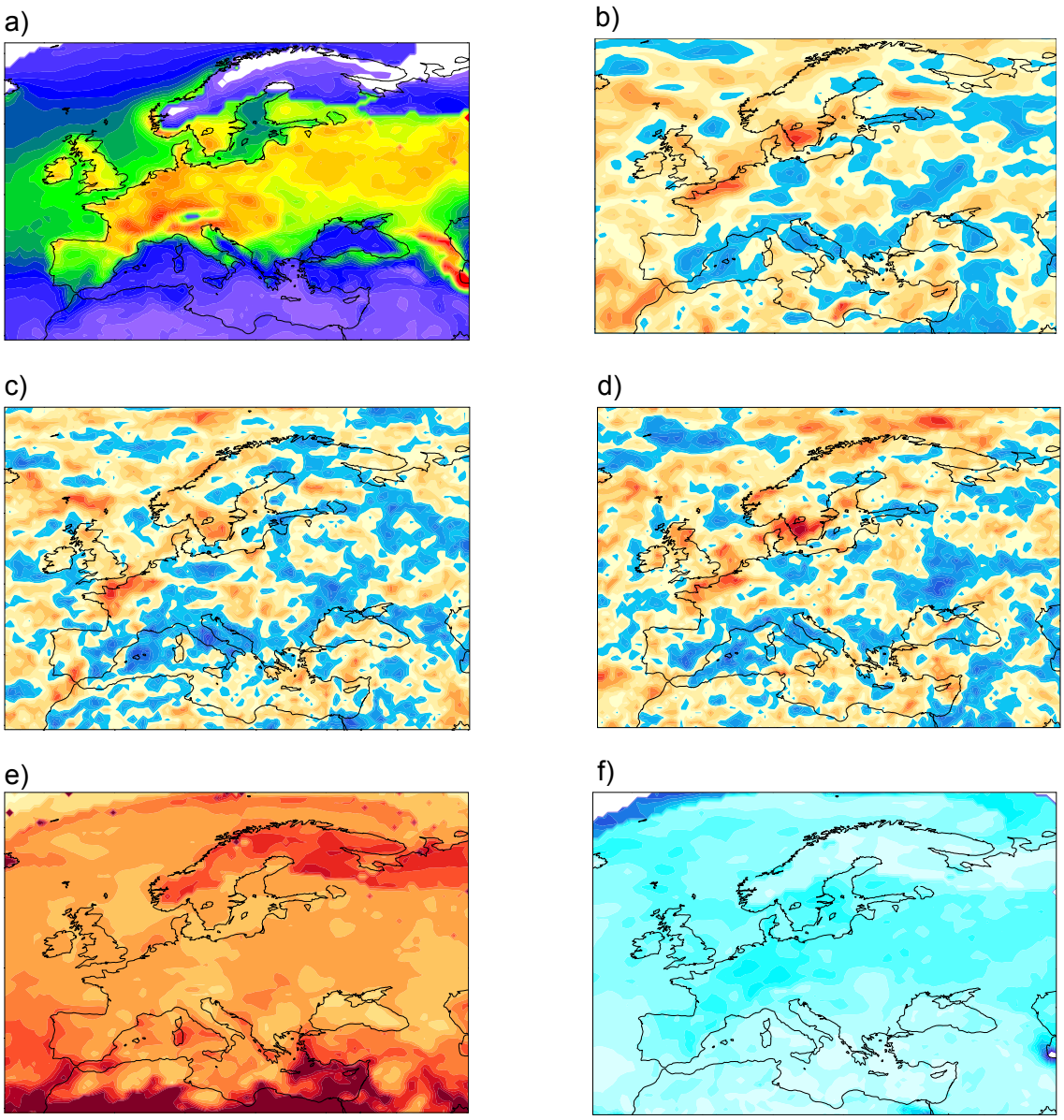

Figure 3.4.2: Hydrothermal Index (a) Bias, (b) Anomaly Correlation, ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio, (f) Continuous ranked probability skill score

3.4.2.3. Selianinov Index

A measure of the effectiveness of precipitation in the growing season is presented through the Selianinov index (SI). According to Magalhães (2008), an SI lower than 1 indicates that the precipitation, during the growing season, in that region is insufficient for wine grape production. Due to the non-linear interactions between precipitation and temperature, the index shows high biases and low skill in all of the forecast skill measures (Figure 3.4.3).

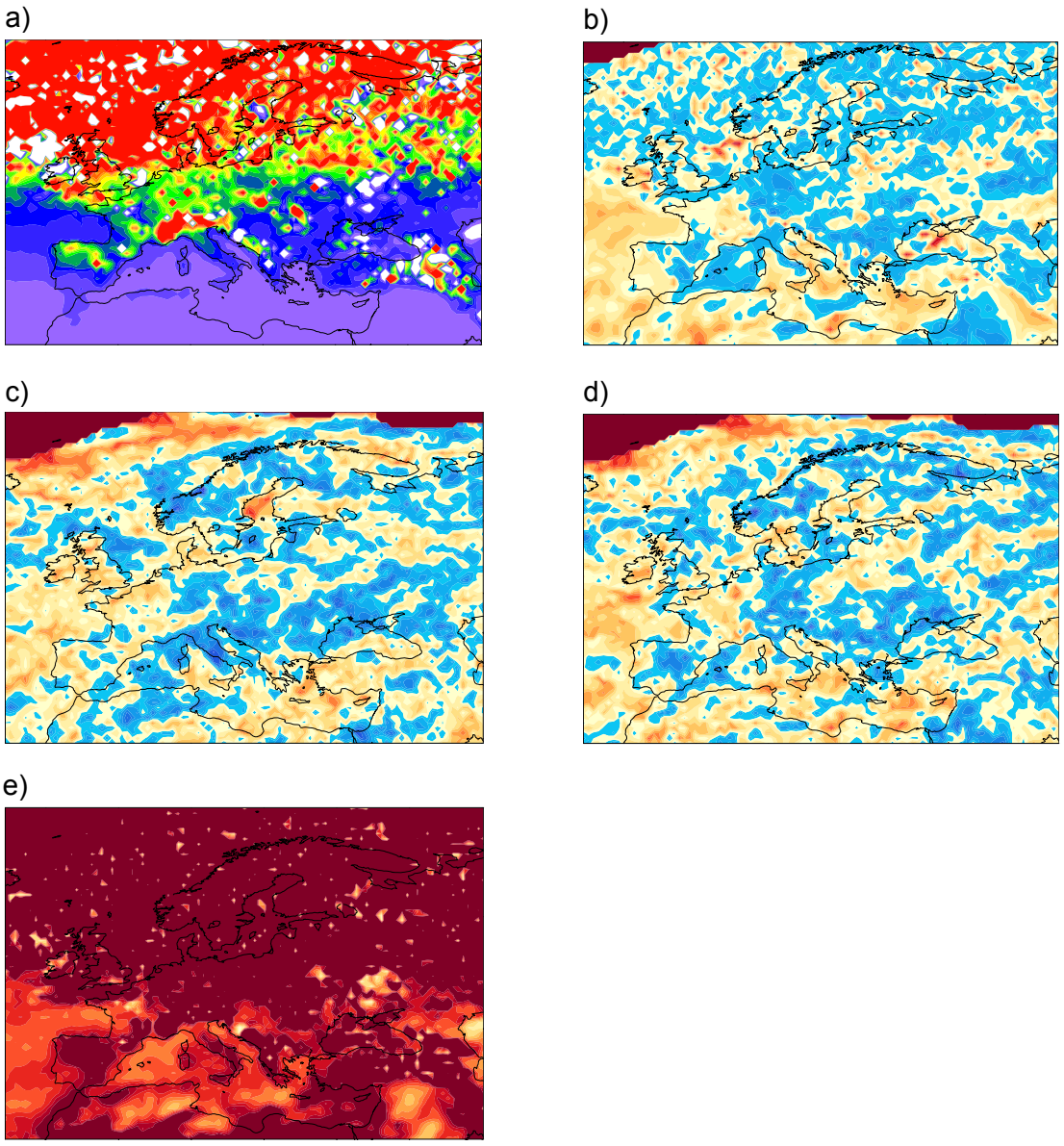

Figure 3.4.3: Selianinov Index (a) Bias, (b) Anomaly Correlation, ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio

3.4.2.4. Cool Night Index

The Cool Night Index mean pattern exposes that almost all of Europe exhibits very cool to temperate nights (average minimum temperatures in September lower than 18°C), meaning that excessively warm nights in later maturity stages of grapes are relatively rare in Europe. On average, the S4 ensemble mean is 0.5ºC colder than the climatology mean (Figure 3.4.4a), yet the anomaly correlation is negative or near zero in all of Europe, indicating no skill in representing the observed anomalies (Figure 3.4.4b). The ROCs score for the first tercile (Figure 3.4.4c), displays some skill in northern Germany, Eastern Europe and the Iberian Peninsula indicating the ability to forecast the colder September mean minimum temperatures. In contrast and with the exception of northern Germany, in these regions the S4 forecast has no skill in the upper ROCs tercile. This can be found in France, Germany and Balkans. The spread to error ratio has values above 0.5 in Germany and higher than 0.6 in the other regions, indicating that the forecast ensemble is under-dispersive. The mildly

negative continuous ranked probability skill score (Figure 3.4.4f) also implies that the skill of the forecast probabilities is worse than the use of climatological frequencies as forecast.

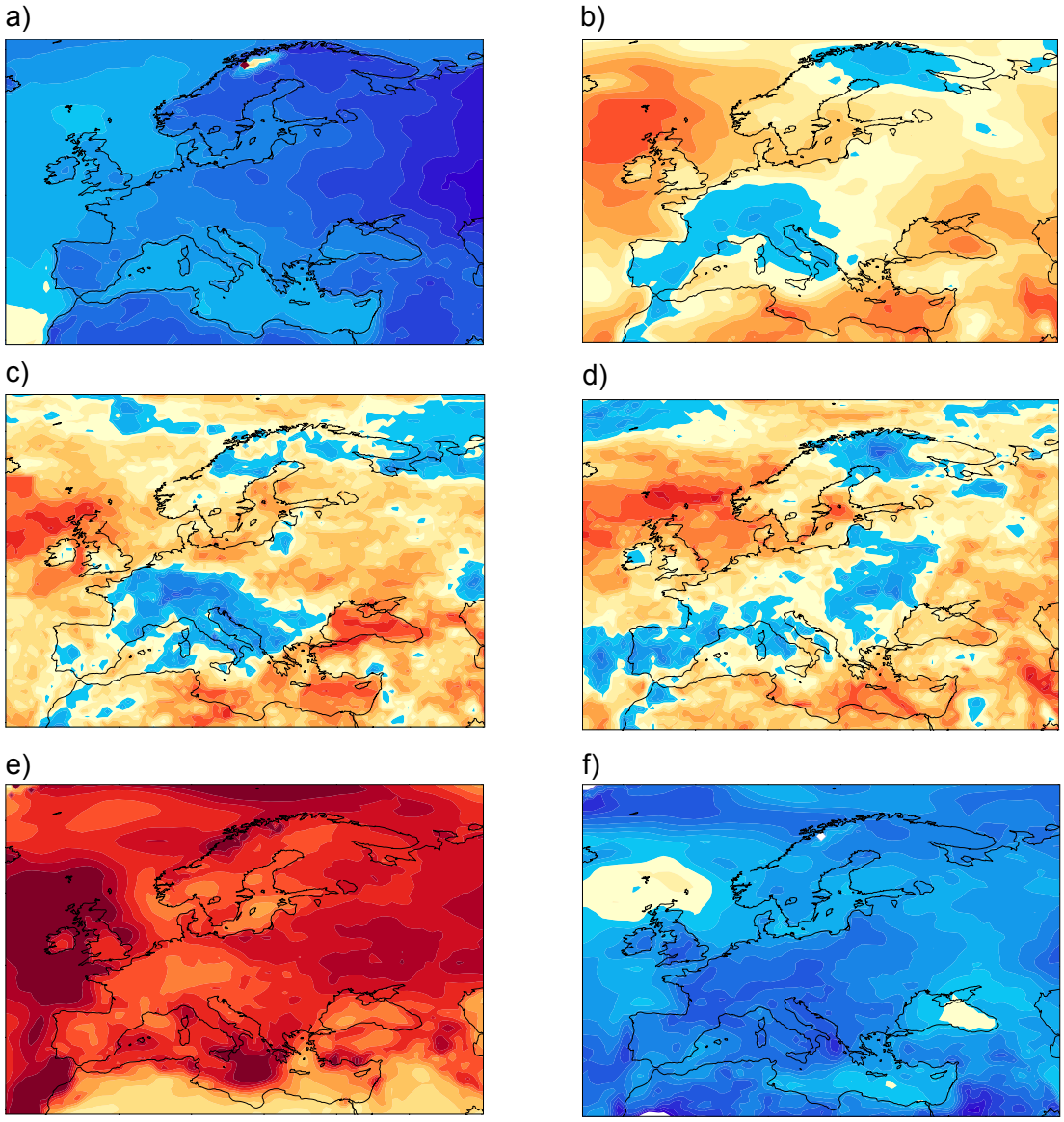

Figure 3.4.4: Cool Night Index (a) Bias, (b) Anomaly Correlation, ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio, (f) Continuous ranked probability skill score

3.4.2.5. Growing Season Suitability

From a thermal perspective, a given region can be considered suitable for grapevine growing when daily mean temperatures are predominantly above 10ºC (Winkler *et al.* 1974) during the growing season, i.e. April to September in Europe. Regions where at least 90% of the days meet this thermal requirement tend to be the most suitable for wine production. This is true for most of the Iberian Peninsula and Italy. Nonetheless, vineyards are also grown at higher latitudes and also altitudes, where the growing season is shorter and thus an 80% threshold will also be considered. Thus, the German and French wine producing regions are considered. Figure 3.4.5a shows a residual negative bias for all of the above mentioned areas as well as for other regions in Europe. The exceptions are the Alps and

Scandinavia where a positive bias is registered, indicating a forecast warmer than the climatology. Negative anomaly correlation is registered in the south of Iberia and the Alps, indicating that on average the forecast is disconnected from the observed anomalies in these regions (Figure 3.4.5b). In the other regions a positive correlation, albeit less than 0.5, is indicative of some skill. This is also evident in the ROCs positive scores for similar regions as the correlation (Figure 3.4.5c and d). The spread to error ratio is very similar to the Cool Night Index but with more skill.

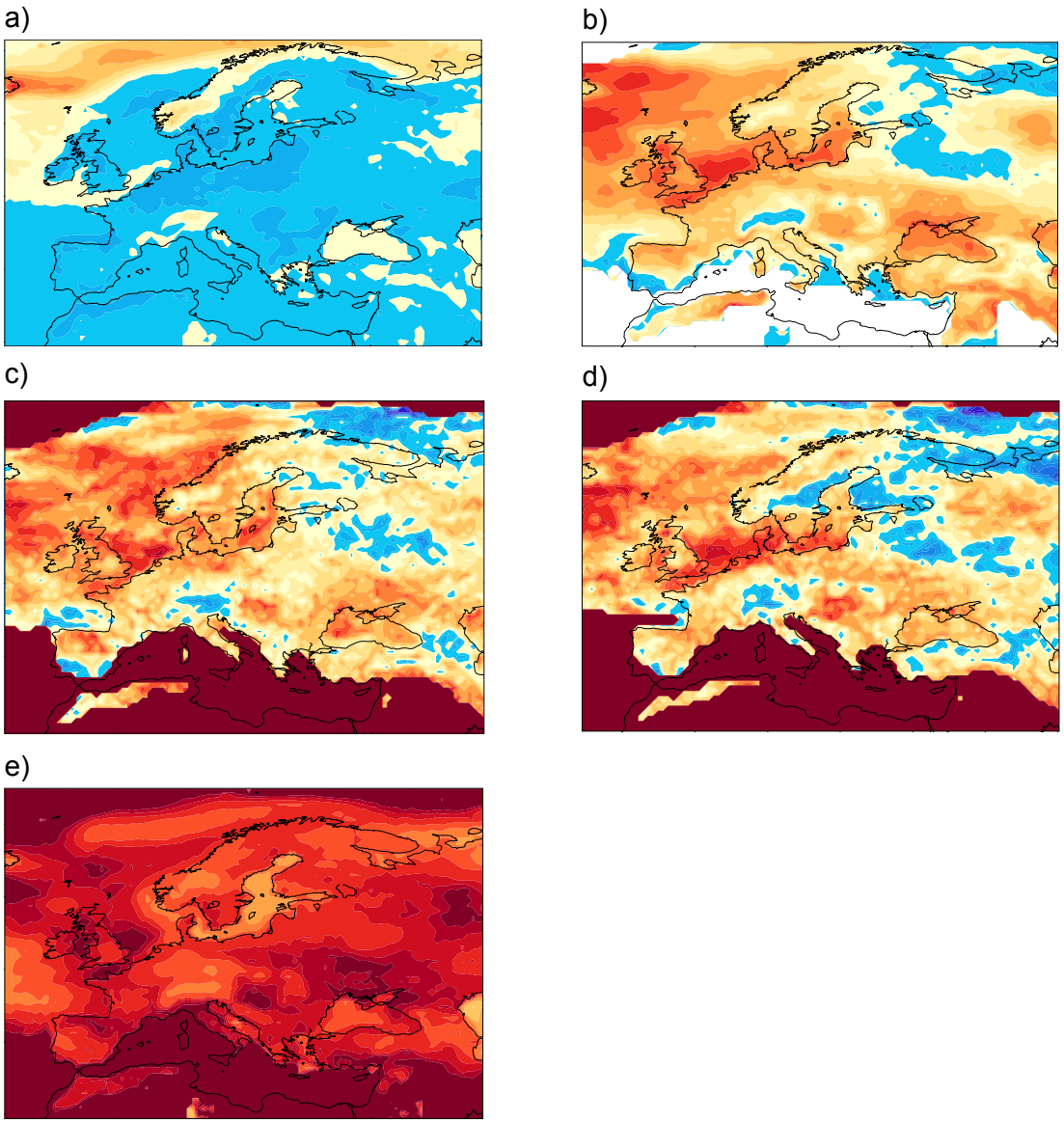

Figure 3.4.5: Growing Season Suitability (a) Bias, (b) Anomaly Correlation, ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio

3.4.2.6. Huglin Heliothermal Index

The Huglin index (HI) integrates the day-length into the temperature distribution, thus explaining the northern viticulture extension. The lower temperatures observed at higher latitudes are compensated by the longer insolation during the growing period. HI below 1500 is usually considered not suitable for viticulture. The area with HI>1500 in ERA Interim and in S4 bias corrected climatologies is in reasonable agreement with European grapevine

distribution. Although considerably reduced in relation to a simple bias correction with the climatology mean, some biases still prevail after the bias correction of the daily time series (Figure 3.4.6a). These biases are mostly associated to central European orography, i.e. the Alps and Carpathians. The anomaly correlation (Figure 3.4.6b) shows positive skill, although limited, in the Mediterranean area and no skill in central Europe.

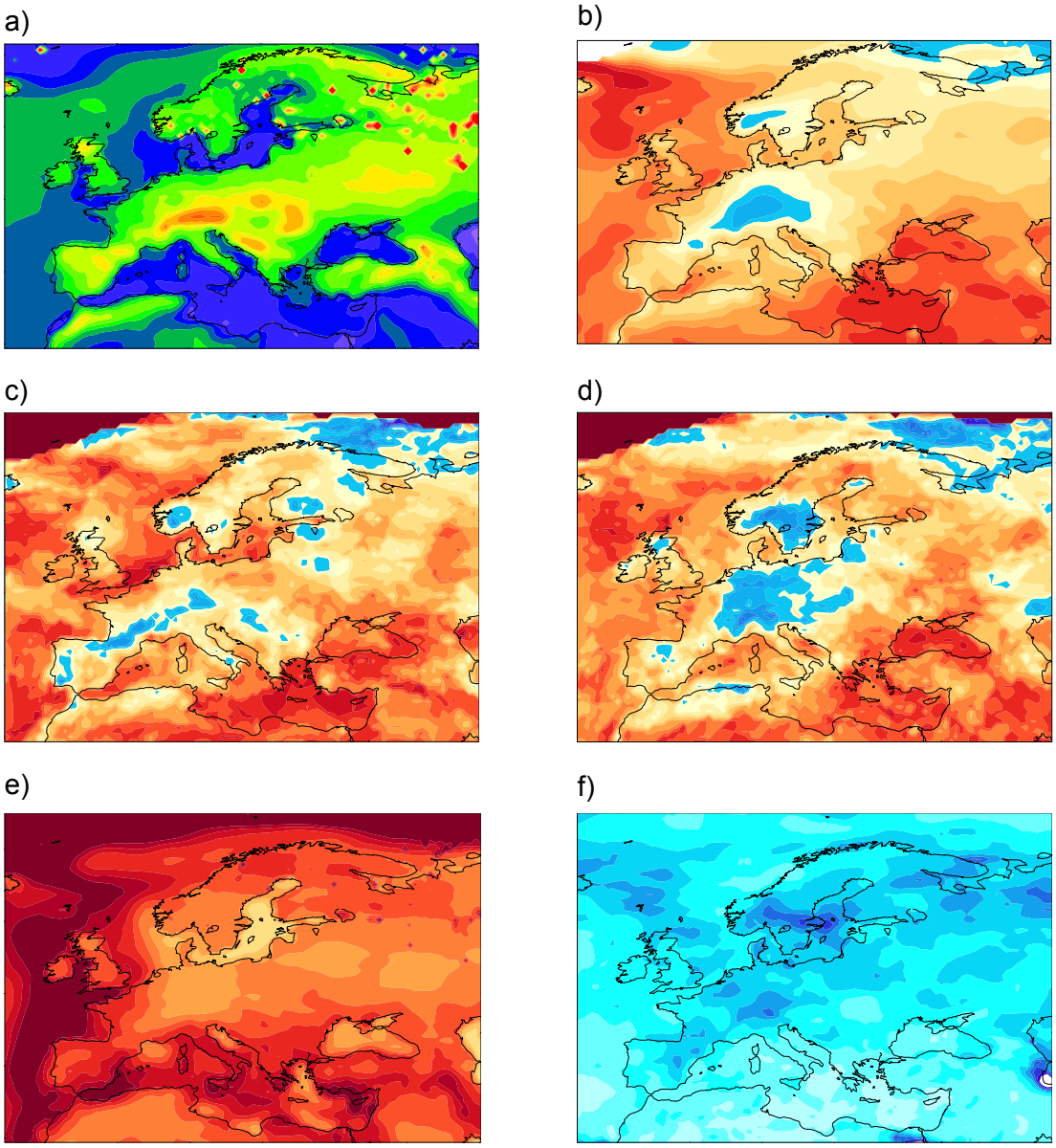

Figure 3.4.6: Huglin Heliothermal Index (a) Bias, (b) Anomaly Correlation, (ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio, (f) Continuous ranked probability skill score

Low ROCs skill was obtained in the lower tercile in Portugal, southern France and Germany mainly, indicating a difficulty in the forecast to determine occurrences in the lower tercile in these areas (Figure 3.4.6c). The highest skill occurs in Eastern Europe and Turkey. The skill for values in the upper tercile increases considerably in western Iberia and north-western France (Figure 3.4.6d). The forecast, however, has no skill for occurrences in the upper tercile in central Europe.

In the Mediterranean, the spread-error ratio (Figure 3.4.6e) is close to one indicating that the ensemble spread is able to represent the forecast uncertainty. Although to a lesser degree, the positive values in central Europe also show a similar signal. Yet, the continuous ranked probability score is negative for the entire continent and climatological frequencies would provide a better forecast (Figure 3.4.6f). Again, the negative CRPSS is likely due to the presence of biases and over-confidence of the forecast.

3.4.2.7. Growing Degree Day

The Huglin (Figure 3.4.6) and the Growing Degree Day (Figure 3.4.7) indices show very similar patterns and thus similar analysis.

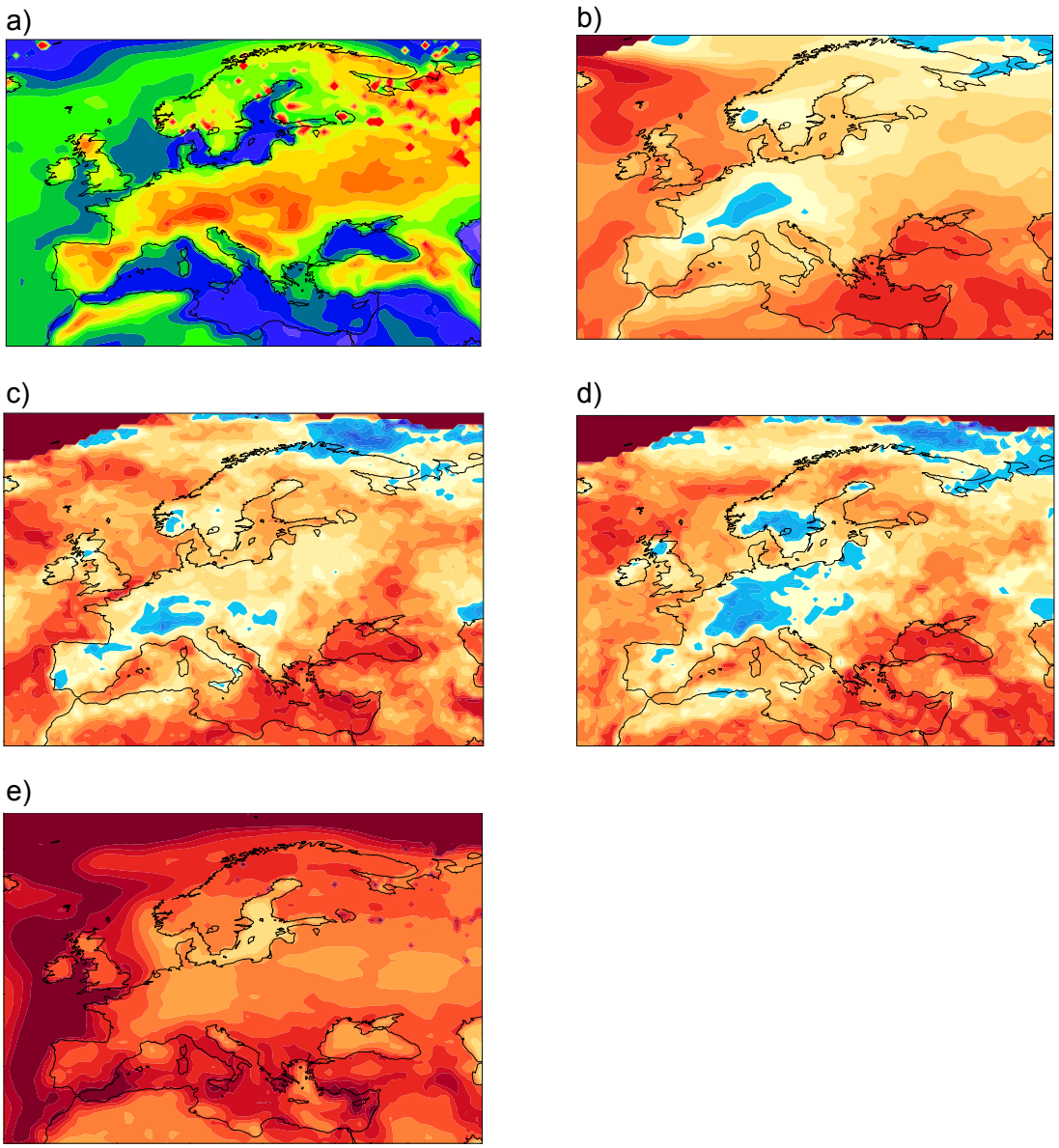

Figure 3.4.7: Growing Degree Day (a) Bias, (b) Anomaly Correlation, (ROC area skill score for the (c) lower tercile and (d) upper tercile, (e) Spread to Error ratio

3.4.3. Conclusions

Although the bias correction with quantile mapping and the spread correction reduced the biases considerably and brought the spread to error ratio to values closer to 1, biases in the indices remain. Therefore, the forecast skill (as measured by the continuous ranked probability score) is negative and using forecast probabilities is worse than using climatological frequencies as a forecast. To overcome this limitation, an attempt to correct the indices directly should be considered.

Key Points: Indices Relevant for Wine Production

- There is limited potential skill for forecasting indices relevant for wine production in Europe. Skill varies strongly by region and index.
- Bias correction of the input quantities used to compute the indices reduced biases considerably, but some biases remain. These biases negatively affect forecast skill. An additional calibration of the index should be attempted to reduce the effect of biases on forecast skill.

3.5. Frost Days

3.5.1. Data and Methods

We present seasonal forecasts of the European Centre for Medium-range Weather Forecasts' (ECMWF) System 4 forecasting system (Molteni *et al.*, 2011) for the winters (DJF) from 1981-2013. We derive the number of frost days (FD), defined as days with minimum temperatures below zero degree C, from bias corrected daily minimum temperature time series of ECMWF System 4 initialized in November. The daily minimum temperature time series are bias corrected against daily minimum temperature from the ERA Interim reanalysis (Dee *et al.*, 2011) using quantile mapping (Panofsky and Brier, 1968). To account for the seasonal cycle and lead time dependence of systematic errors, quantile mapping is carried out using a moving window of 31 days centred on the day of interest as in Themeßl *et al.* (2012) and Wilcke *et al.* (2013). In addition, we carry out quantile mapping in leave-one-out cross validation mode using data from 1981-2010 to estimate the correction factors.

We use quantile mapping instead of a simple bias correction with the climatological mean such as for heating and cooling degree days (Section 3.3.1). With quantile mapping, residual biases in frost days are small (Figure 3.5.1a), whereas FDs are underestimated by up to 5 days per season when using mean debiasing (Figure 3.5.1b).

In the following we present the evaluation of time series of forecasted winter frost days against winter frost days derived from the ERA Interim reanalysis. In regions where almost every day is a frost day or where frost days rarely occur, the year to year variability in frost days is too limited to meaningfully interpret forecast quality. Therefore, we analyse forecast quality only in regions where at least 5% and less than 95% of the days per winter are classified as frost days on average (see mask in Figure 3.5.1). These regions are roughly confined to the US, western and central Europe, the Middle East, and eastern China.

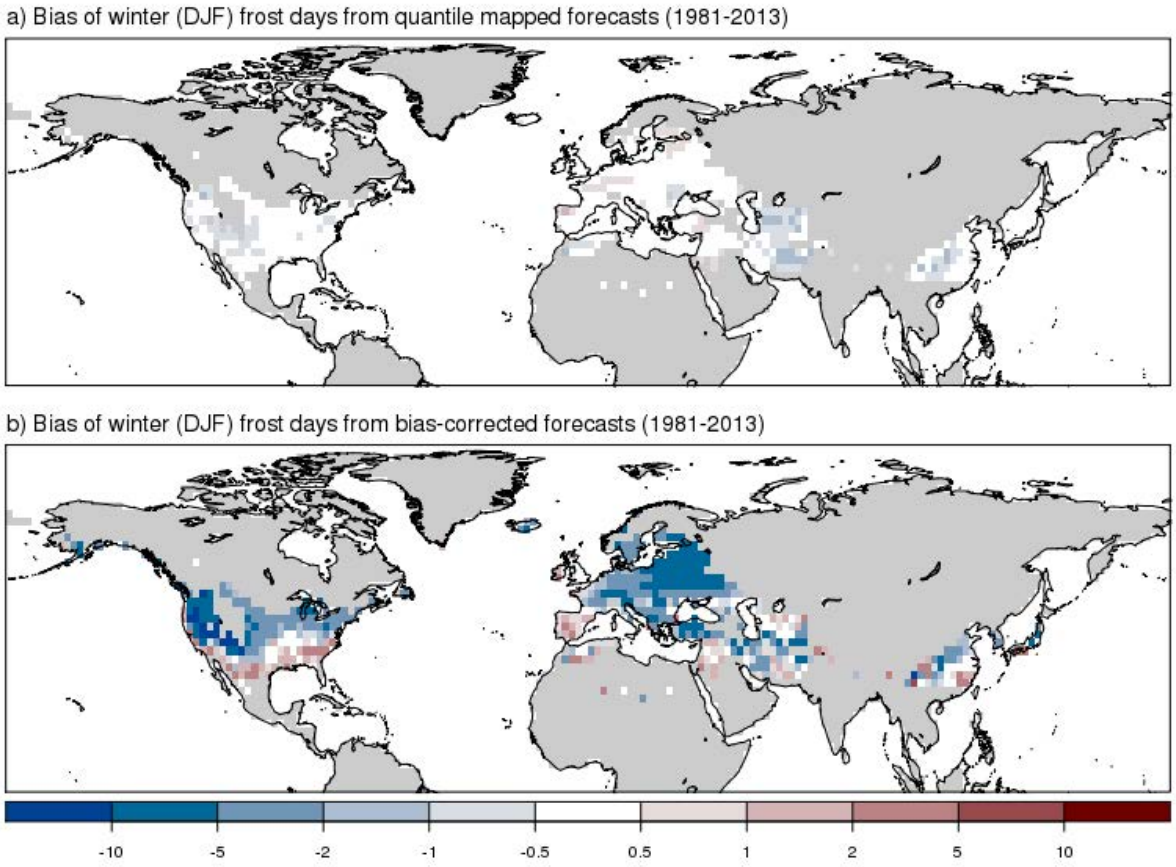

Figure 3.5.1: Bias of winter (DJF) frost days derived from forecasts calibrated using quantile mapping (a) and mean de-biasing (b). Bias is shown in units of days per season. Areas where frost days occur on average on less than 5% or more than 95% of the days per season are masked in grey.

3.5.2. Seasonal Forecast Quality for Frost Days

Correlation of the ensemble mean FD with FD derived from the ERA Interim reanalysis is generally weak (Figure 3.5.2a). Significant positive correlations are found in the Middle East and eastern Asia with correlations generally below 0.6. In north-western Europe positive correlations are found, but these generally do not exceed 0.4.

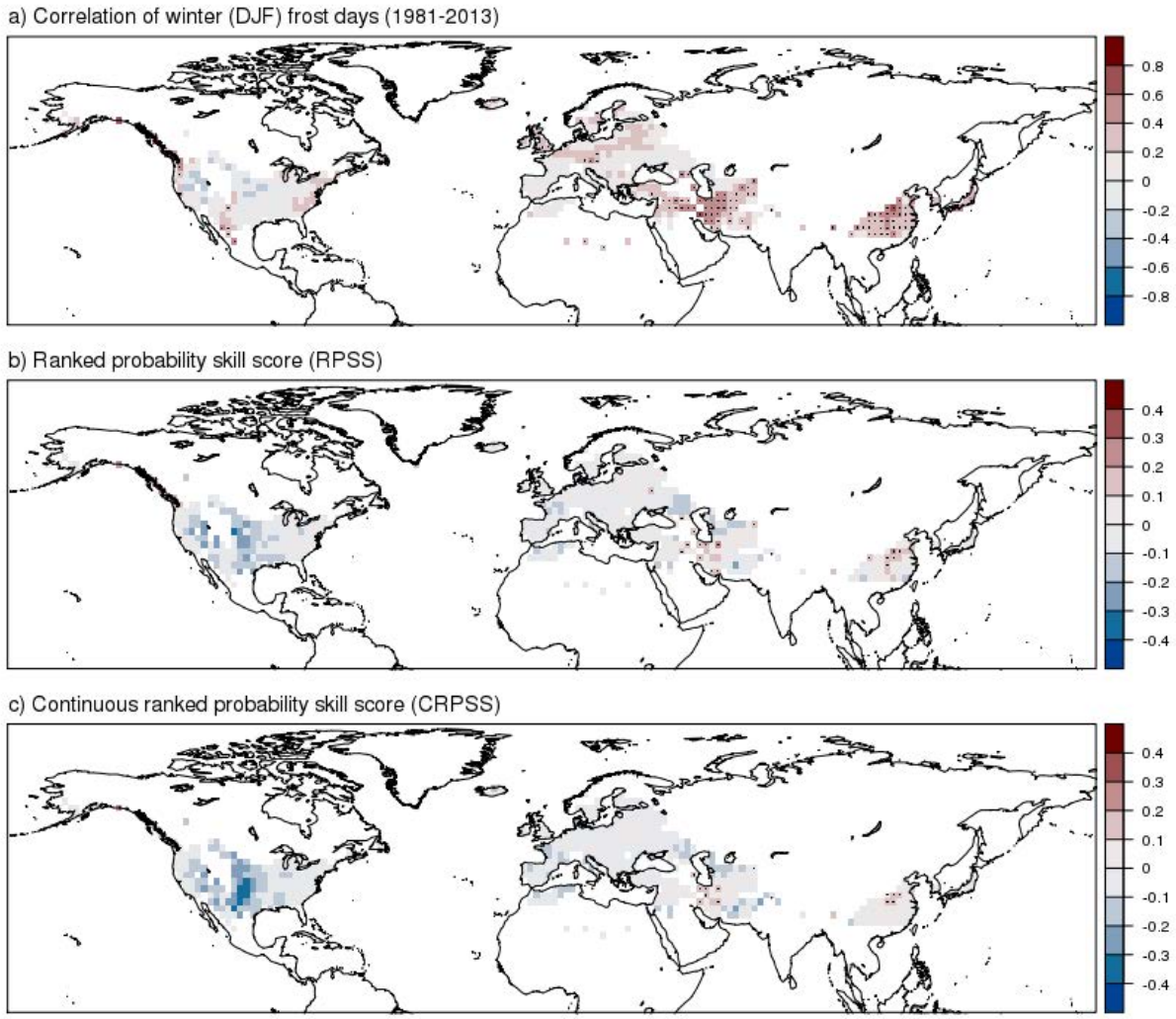

Figure 3.5.2: Correlation (a), RPSS (b), CRPSS (c) and spread to error ratio (d) of forecasts of winter (DJF) frost days. Stippling in a-c indicates correlations and skill scores significantly (at the 5% level) larger than zero.

Similar to the correlation shown in Figure 3.5.2a, we find very limited skill as measured by the ranked probability skill score (RPSS) for tercile forecasts (Figure 3.5.2b) and the continuous ranked probability skill score (CRPSS, Figure 3.5.2c). A few grid points with significantly positive RPSS and CRPSS can be found in the Middle East and eastern China. In Europe, RPSS and CRPSS of winter frost days is marginal, whereas predominantly negative RPSS and CRPSS is found in the western US. The similarity of RPSS and CRPSS illustrates that the lack of skill in CRPSS is not due to residual biases in forecasted frost days (Figure 3.5.1) as these do not affect the ternary category forecasts used to compute the RPSS. The limited skill is also not due to uncertainty introduced by the out-of-sample calibration of daily minimum time series used to derive seasonal frost days. FD forecasts derived from in-sample calibrated daily minimum temperatures exhibit similarly marginal skill (not shown). This leads us to conclude that the ECMWF System4 is not able to skilfully predict frost days in the northern hemisphere.

3.5.3. Conclusion

Forecasts of winter frost days are non-trivial only in limited areas of the mid-latitude. Elsewhere, almost every day is a frost day or frost days occur very rarely. Potential skill in forecasting frost days as measured by correlation is fairly limited with areas of significant positive skill in the Middle East and eastern China. Actual skill as measured by the ranked or continuous ranked probability skill scores is found to be limited or negative. As for indices relevant for wine production, additional calibration of the index may provide a way forward to remove negative skill where there is little potential skill and maybe even enhance skill in areas where there is potential skill.

Key Points: Frost Days

- The choice of bias correction method matters for forecasts of frost days. When using simple mean de-biasing, simulated frost days derived from de-biased forecast time series exhibit substantial biases. These are considerably reduced using more sophisticated daily calibration methods such as quantile mapping.
- At present, ECMWF System 4 seems to be unable to skilfully forecast frost day occurrence in the northern hemisphere.

3.6. Very Heavy Precipitation Days and Total Precipitation

3.6.1. Definition of Indices

The indices presented here represent a selection of indices used in ECA&D and are relevant to water management and agriculture sectors. Under the right conditions the indices can act as indicators of potential extreme river flow. A set of indices based on E-OBS v11 (Haylock *et al.,* 2008) is presented in the following link: http://www.ecad.eu/utils/mapserver/eobs_maps_indices_R.php. The definition of indices is based on Klein Tank *et al.*, 2009:

Very heavy precipitation days (R20mm)

Let RR_{ij} be the daily precipitation amount for day i of period j. Then counted is the maximum number of days where:

$$RR_{ij} \geq 20mm$$

Total precipitation on wet days (> 1mm, PRCPTOT)

Let RR_{wj} be the daily precipitation amount on a wet day w in period j. Then:

$$PRCPTOT = \sum_{j=1}^{n} RR_{wj} \quad \text{where } RR_{wj} \geq 1\,mm$$

3.6.2. Observations, Seasonal Forecast Data and Bias Correction

E-OBS v11 (Haylock *et al.,* 2008) is used as observations, while the ECMWF System-4 (Molteni *et al.,* 2011) is used as the seasonal forecast data. Here, 30-years of re-forecasts of

the model (1981-2010) are used, with a 15-member ensemble and 1-month lead-time. Results are presented for the winter season (3 months statistics, December-January-February) with initialization 1st of November and spring season (3 months statistics: March-April-May) with initialization 1st of February. The area of interest is Central Europe, focusing on the areas of the Rhine and Meuse rivers. This area is chosen in connection with the DWD case study on operational water management for the Rhine basin. Winter and spring are selected based on the occurrence of observed high flows in the Rhine basin (Photiadou *et al.*, 2015). After interpolating the seasonal forecasts to the E-OBS grid, these were bias corrected using quantile mapping.

Maps of biases between the climatology of the indices derived from the bias corrected forecast time series and the observations are presented in Figure 3.6.1 and Figure 3.6.2, for winter and spring respectively. R20mm doesn't show systematic errors but a rather mixed bias for both seasons. PRCPTOT shows in winter a dry bias over most of Central Europe and the countries surrounding the Rhine and Meuse basins. For spring, a wet bias is present for PRCPTOT in the Swiss basin and northern Italy and southern France.

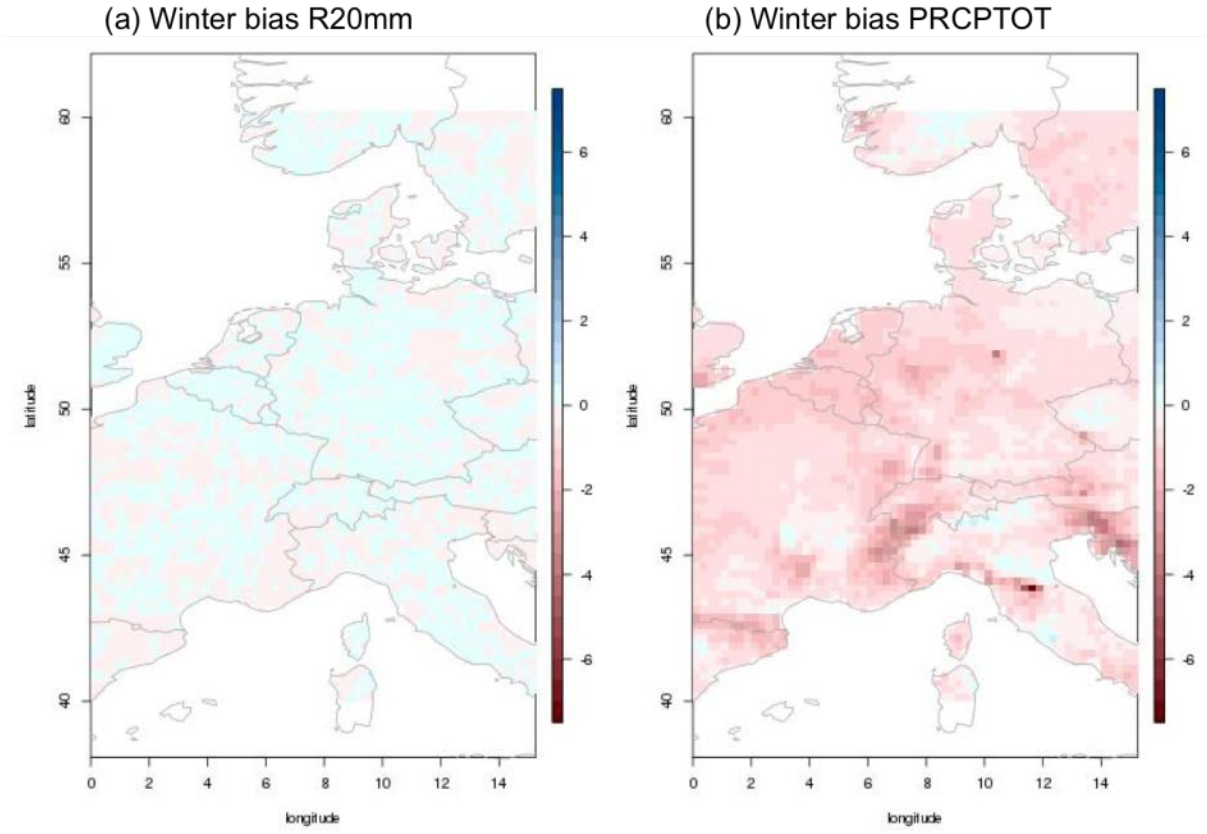

Figure 3.6.1: Winter bias of (a) R20mm and (b) PRCPTOT

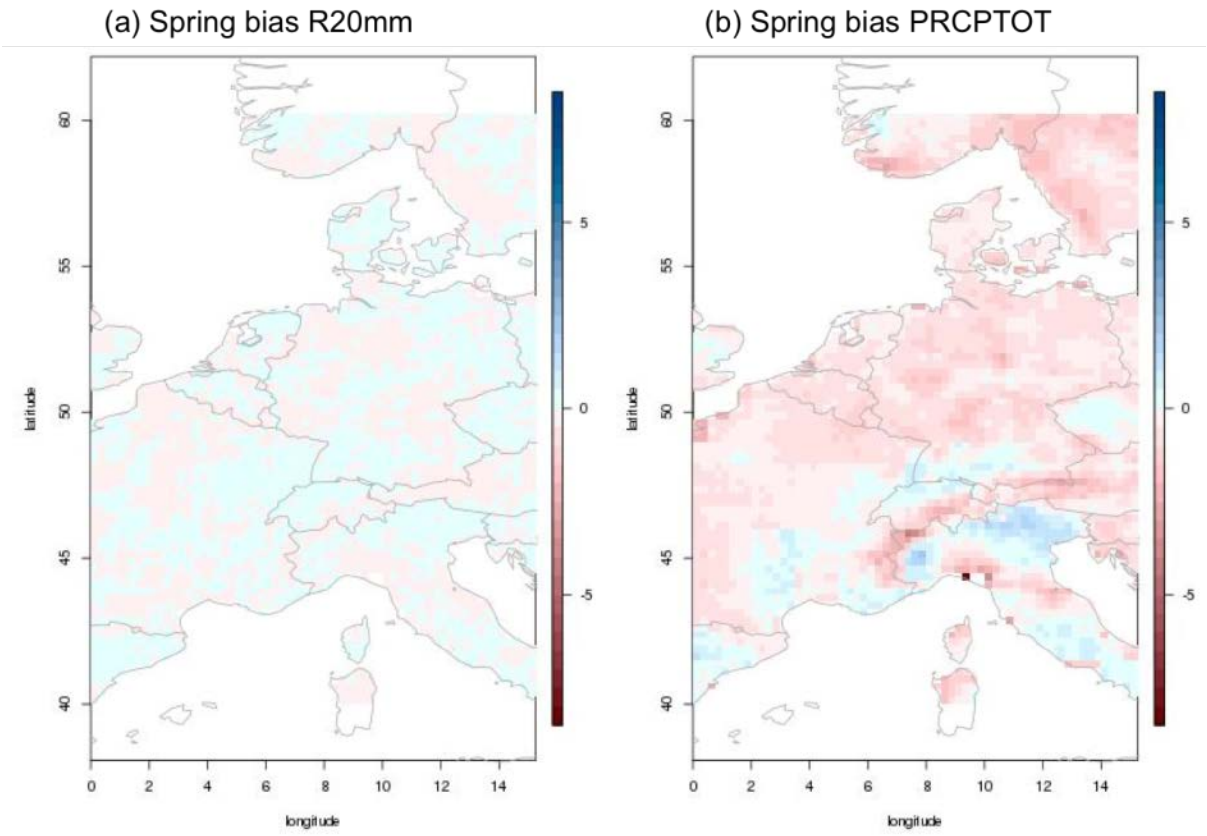

Figure 3.6.2: Spring bias of (a) R20mm and (b) PRCPTOT

3.6.3. Skill Assessment of Seasonal Forecasts

3.6.3.1. Anomaly Correlation

Correlations between the ensemble mean and observed indices are presented for winter and spring in Figure 3.6.3 and Figure 3.6.4 respectively. For both R20mm and PRCPTOT, positive correlations are present in northern Germany and north Italy, while negative correlations are present in most of France and Switzerland. Generally, correlations do not exceed 0.5 suggesting limited predictability of both heavy precipitation days (R20mm) and total precipitation during wet days (PRCPTOT).

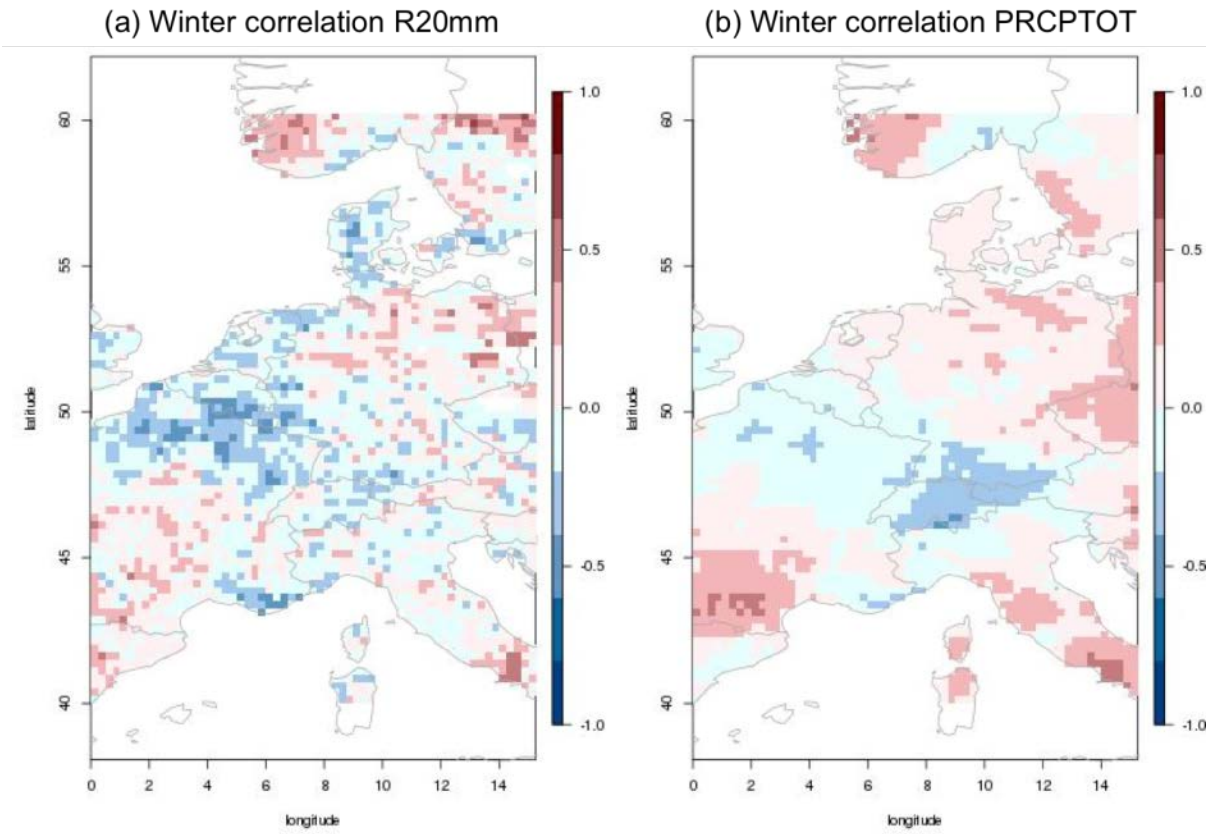

Figure 3.6.3: Correlation of winter (a) R20mm and (b) PRCPTOT

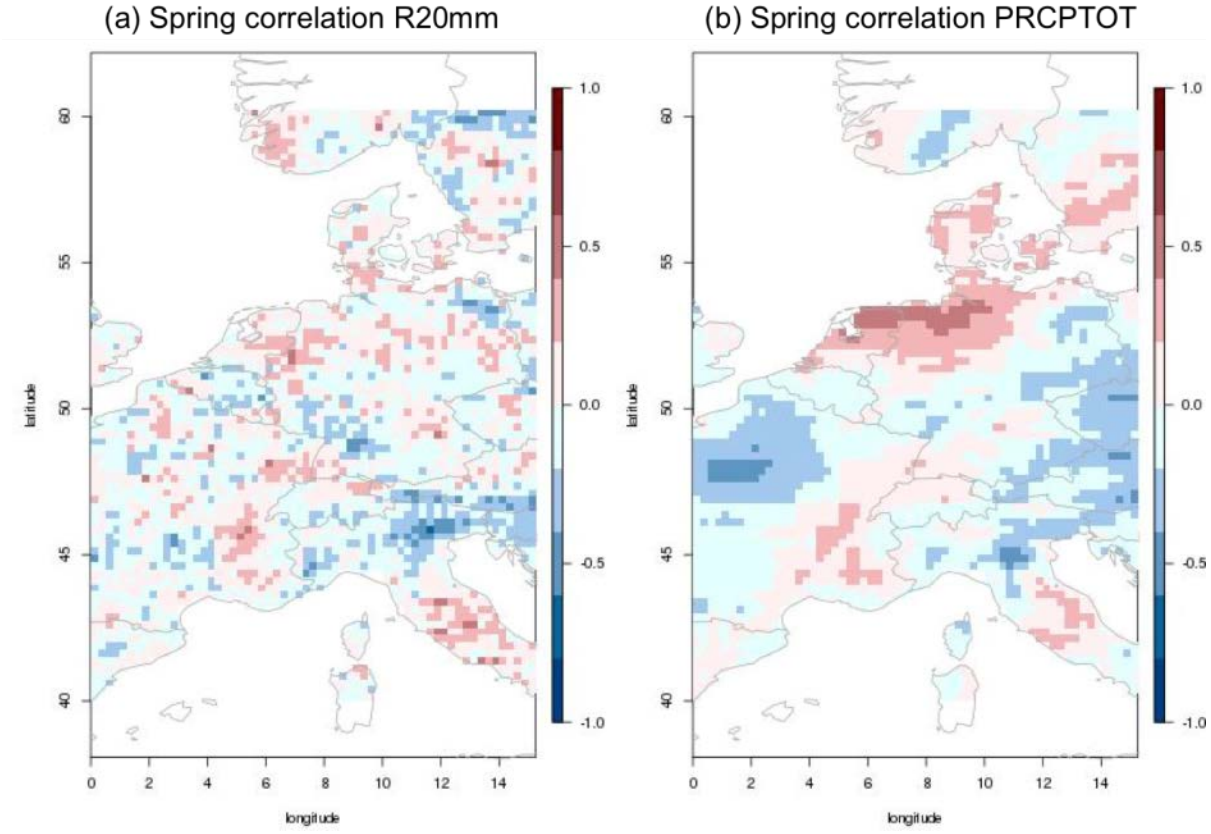

Figure 3.6.4: Spring correlation of (a) R20mm and (b) PRCPTOT

3.6.3.2. Continuous Ranked Probability Skill Score

In Figure 3.6.5 and Figure 3.6.6 we show the continuous ranked probability skill score (CRPSS) for the two indices for winter and spring respectively. For winter the CRPSS is negative for all the countries in the Rhine basin with some small exceptions over Germany. In contrast, CRPSS in spring has larger areas with positive skill, such as Northern Germany and the Netherlands. For both seasons, the CRPSS reflects the negative correlations shown in Figure 3.6.3 and Figure 3.6.4.

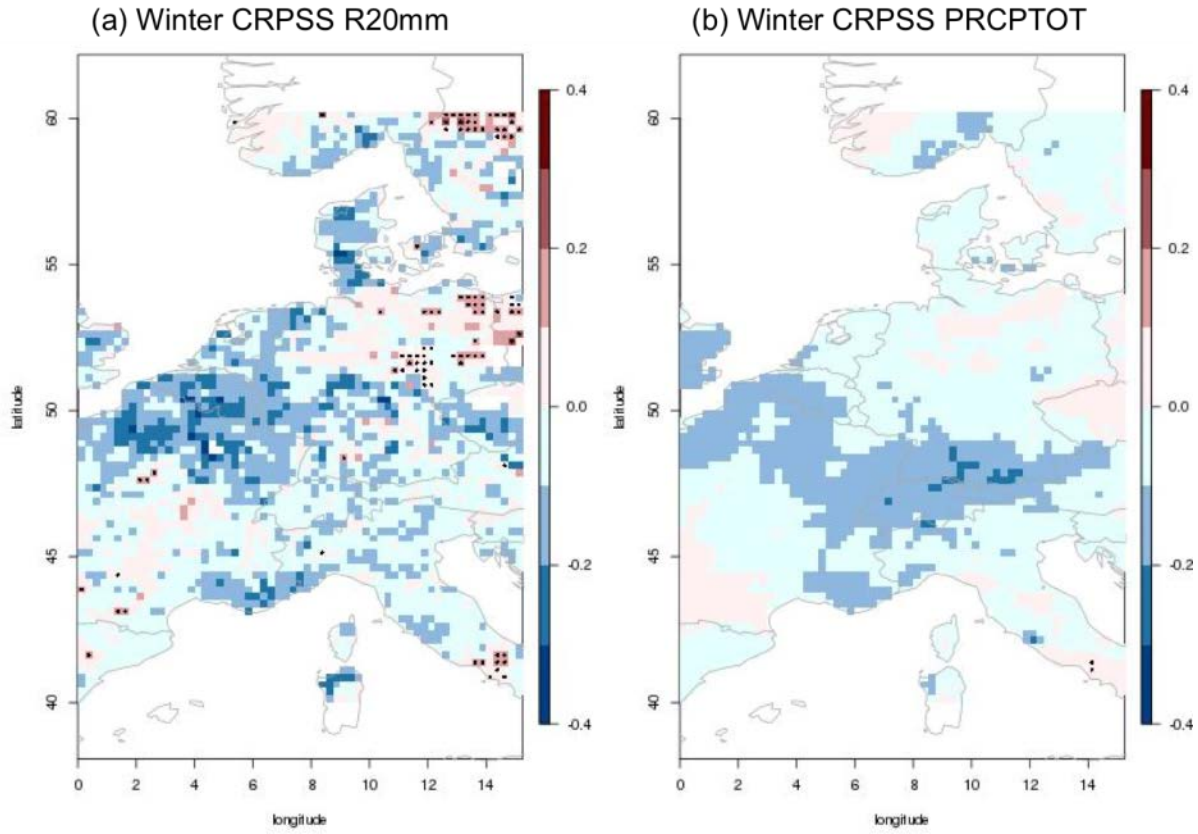

Figure 3.6.5: Continuous ranked probability skill score for winter (a) R20mm and (b) PRCPTOT. Black dots indicate the areas with significant skill at the 95% confidence level.

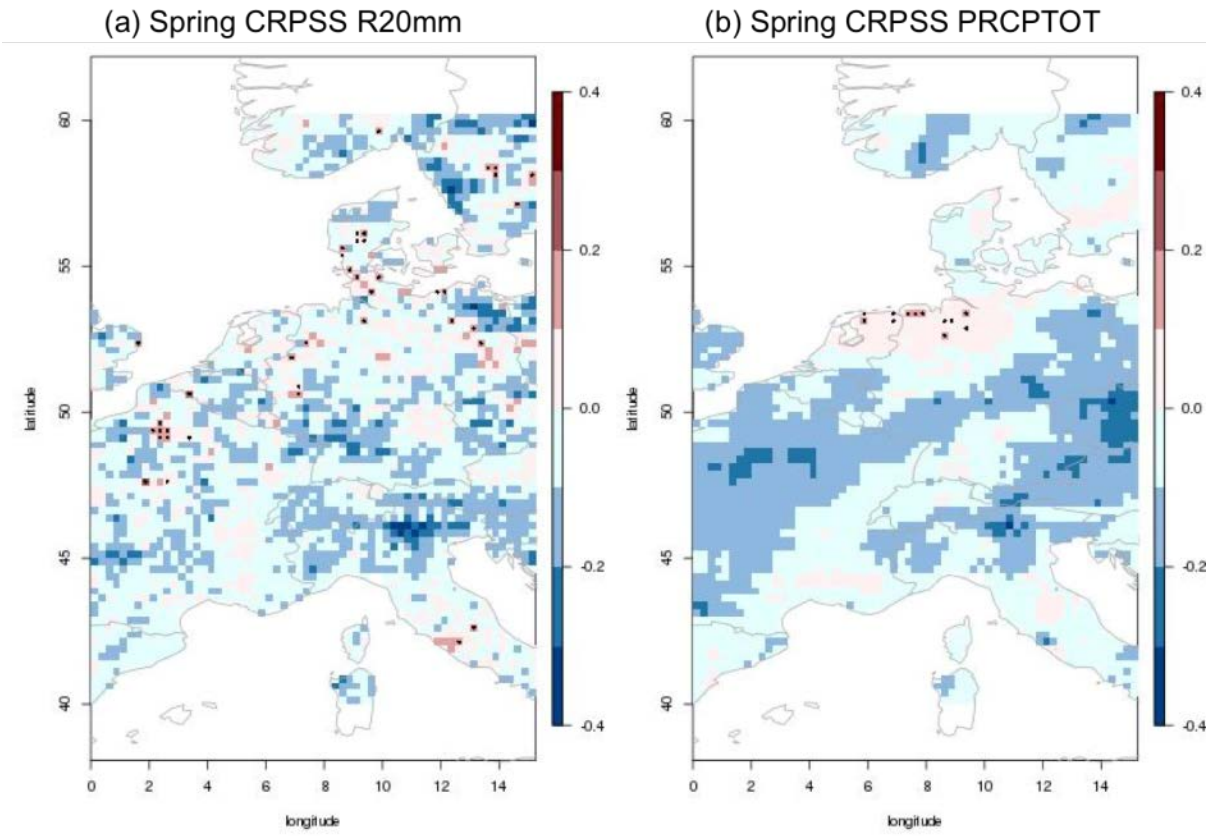

Figure 3.6.6: Continuous ranked probability skill score for spring (a) R20mm and (b) PRCPTOT. Black dots indicate the areas with significant skill at the 95% confidence level.

3.6.3.3. Ranked Probability Skill Score

In Figure 3.6.7 and Figure 3.6.8 we show the Ranked Probability Skill Score (RPSS) for winter and spring respectively for the indices R20mm and PRCPTOT. For winter R20mm, RPSS is characterized by distinct negative values over France, Belgium, Luxembourg and Switzerland, with the exception of small areas over north-eastern Germany. Similar results are found for spring with the exception of small areas over France, Belgium, Netherlands and north-western Germany where positive values indicate a skilful forecast compared to a climatological forecast. For PRCPTOT, negative values in winter are present over northern France, Belgium, Netherlands, Switzerland and southern Germany, while positive values are present in south-western France and northern Germany. In spring, negative values dominate most of Central Europe with the exception of northern Germany and the Netherlands with significant RPSS, and north-eastern Switzerland. This is consistent with the positive correlation found in these areas in spring.

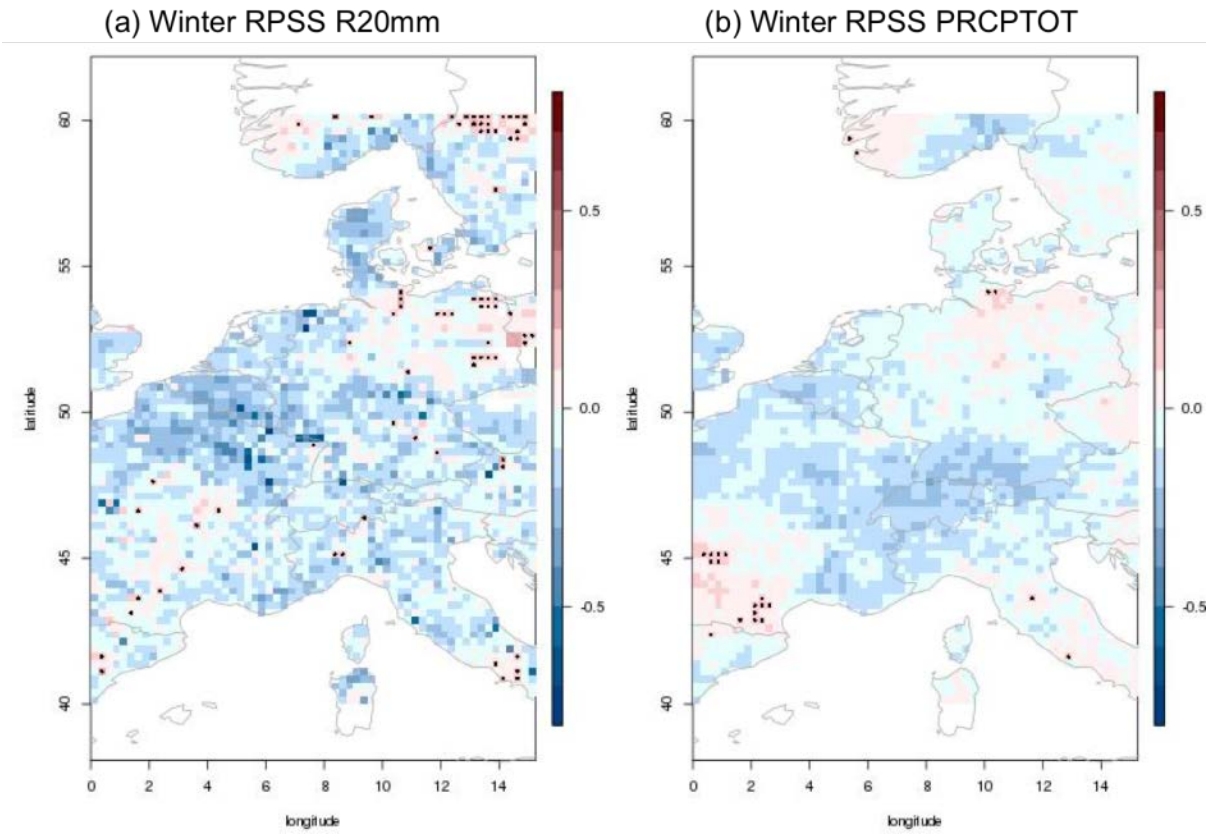

Figure 3.6.7: Ranked probability skill score (terciles) for winter (a) R20mm and (b) PRCPTOT. Black dots indicate the areas with significant skill at the 95% confidence level.

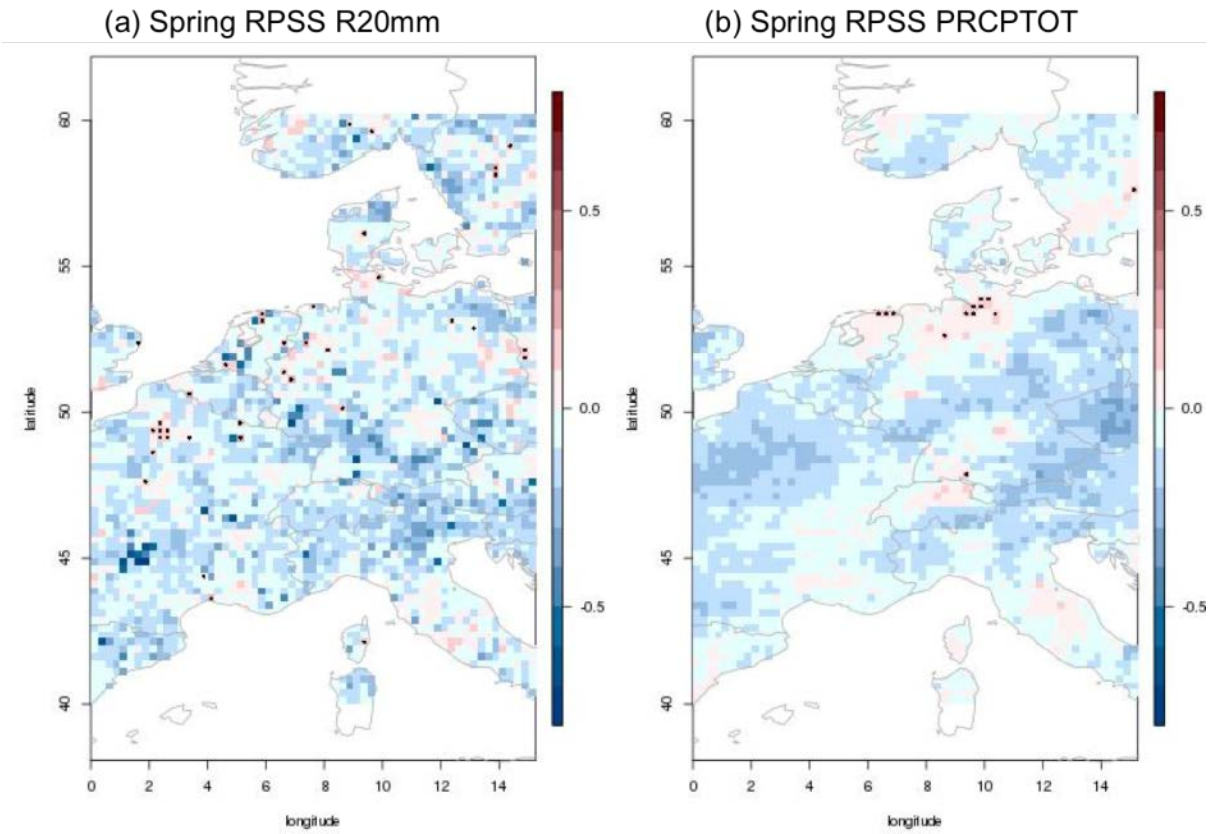

Figure 3.6.8: Ranked probability skill score (terciles) for spring (a) R20mm and (b) PRCPTOT. Black dots indicate the areas with significant skill at the 95% confidence level.

3.6.4. Conclusions

Heavy precipitation days and total precipitation on wet days are indices relevant for forecasting extreme river flow. We identify considerable potential prediction skill (as measured by correlation) for total precipitation in spring in northern Germany and the Netherlands resulting in actual skill (as measured by CRPSS and RPSS). Elsewhere skill is limited. No robust pattern of forecast skill for heavy precipitation days has been identified. This illustrates the difficulty when working with climate indices representative of more extreme events.

Key Points: Heavy Precipitation Days and Total Precipitation

- Skill in predicting winter and spring total precipitation in central western Europe is limited to northern Germany and the Netherlands in spring.
- In contrast, no robust patterns of forecast skill are found for seasonal forecasts of heavy precipitation days. This illustrates the difficulty when working with indices relating to rare events.

3.7. Heavy Precipitation

3.7.1. Definition and Equation

In order to analyse intense precipitation in the seasonal model output we defined an index which looks at the integrated sum of the value of precipitation exceeding the upper quartile of the rainfall distribution during wet days. We call such an index Intense Precipitation Index (IPI). That is:

$$IPI = \sum_{i=1}^{n} \max(P_i - P_{75\%}, 0)$$

Where P_i is the daily mean rainfall at day i, and the summation is over all days of the season. Such a formulation implies that for each season we will only have one number which represents the "intensity" of precipitation during the whole season.

3.7.2. Bias Correction

Being defied in relation to the model climatology the IPI should be less affected by model biases. Nevertheless looking at the histogram of both model output and the observations we realised that biases still exist. These were removed using a simple de-biasing approach based on the mean value of both observations and models over the entire season and hindcast set.

3.7.3. Skill Scores and Significance

3.7.3.1. Anomaly Correlation

First we show the anomaly correlation between the ensemble mean IPI forecasts and IPI derived from WATCH forcing data (Weedon *et al.*, 2014) for the summer (JJA) seasons from 1995-2010 (Figure 3.7.1). We find generally low values of correlation that is generally not significantly different from zero. Three areas of positive correlation exist: the NW part of the Pyrenees and Gascoigne region of France, the southern part of the Balkans, and Tunisia. It is also worth stressing that no correlation between observations and forecasts can be identified over the Alpine ridge which jointly with the area around Poland is one of the regions where IPI reaches the highest climatological values (Figure 3.7.2)

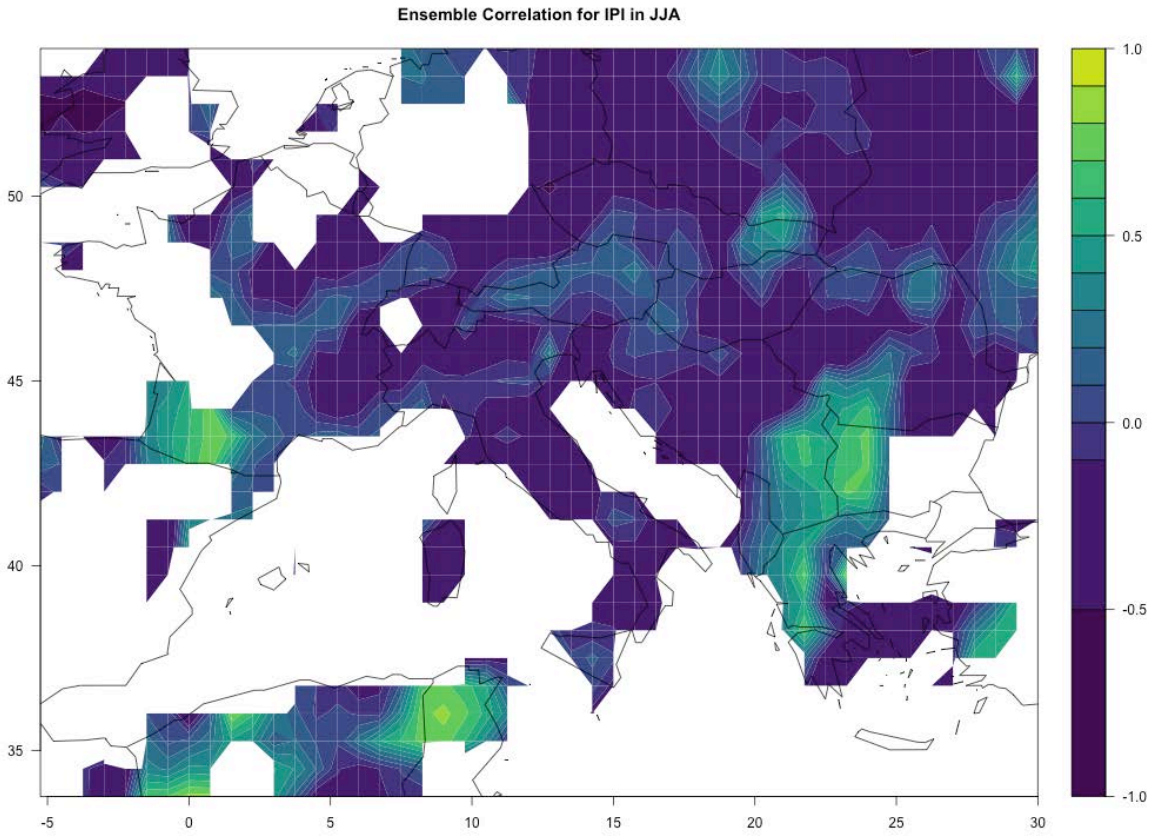

Figure 3.7.1: Correlation between the ensemble mean IPI forecast by ECMWF system 4 for JJA period and the corresponding observations extracted from WATCH forcing data for the period 1995-2010 Forecasts with System 4 have been initialized in May and 25 of the available ensemble members have been used for the calculation.

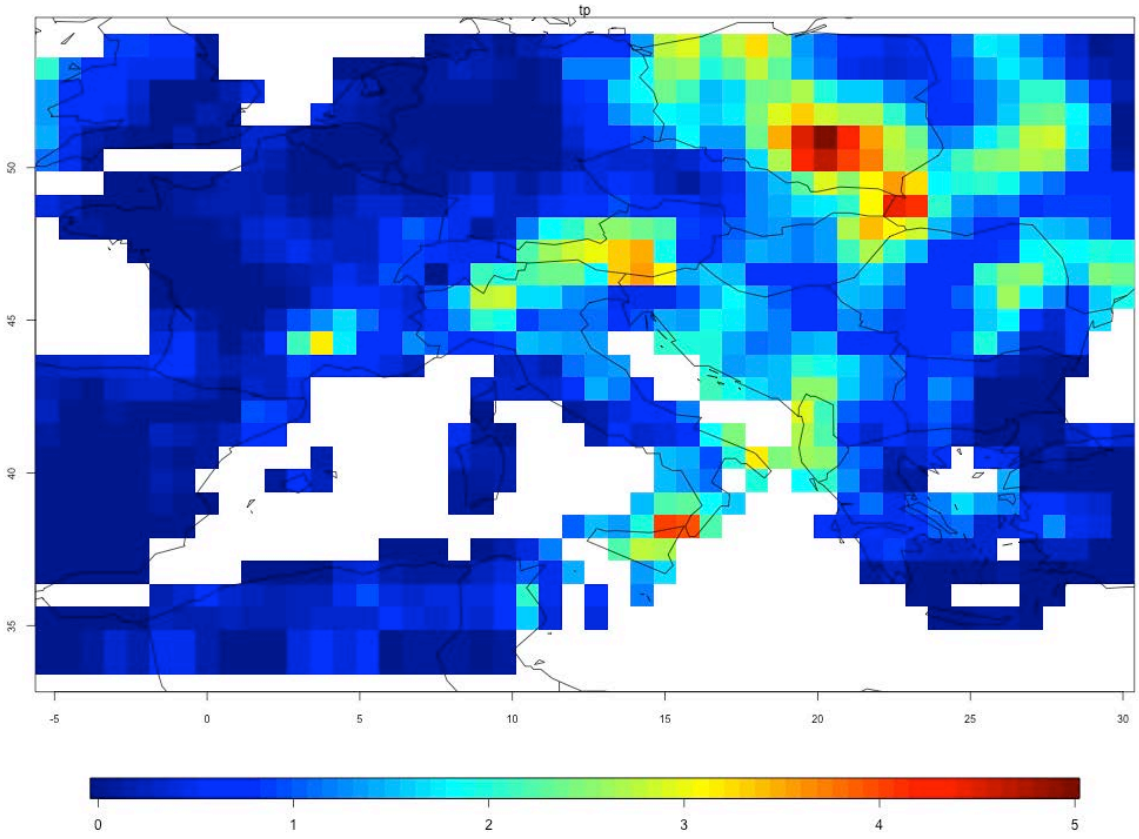

Figure 3.7.2: Observed climatology of intense precipitation.

3.7.3.2. ROC Area

The ROC area skill score (Figure 3.7.3) is used to distinguish forecast skill for different events, here IPI falling in the lower or upper tercile of the distribution corresponding to summers that are characterised by more or less intense precipitation events.

We find some marginal skill (compared with a climatological forecast) in the region around the Mediterranean and more evidently in SW France and the western Balkans. Interestingly the regions where ROC scores reach their highest values correspond to regions of highest correlation between the ensemble means and the observations but also to the region where the climatology of the intense precipitation events is very marginal.

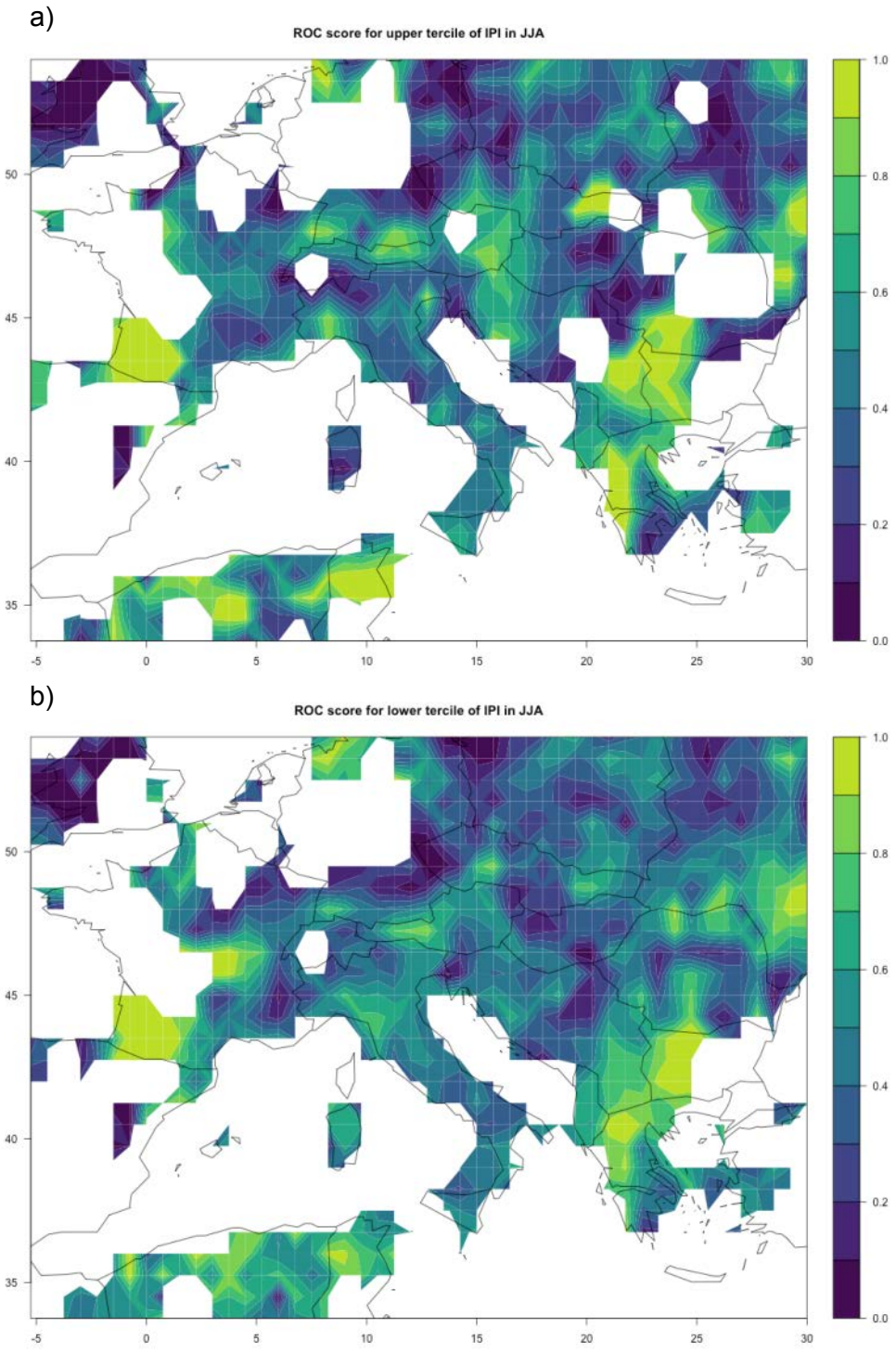

Figure 3.7.3: ROC area skill score for summer (JJA) Intense Precipitation Index falling in the upper tercile (a) and lower tercile (b) for the period from 1995-2010 from forecasts with ECMWF System 4 initialized in May. Seasonal IPI were computed from the raw series of daily rainfall. The ROC area skill score is computed with reference to a climatological forecast.

3.7.3.3. Spread-error Ratio

The spread to error ratio of IPI forecasts from 1995-2010 is generally relatively close to 1 (Figure 3.7.4). Of the regions where anomaly correlation and ROC indicated some potential

only SW France appear to be relatively close to 1 whilst for both Tunisia and the Balkans the spread to error ratio is clearly well above 1 (over-dispersive).

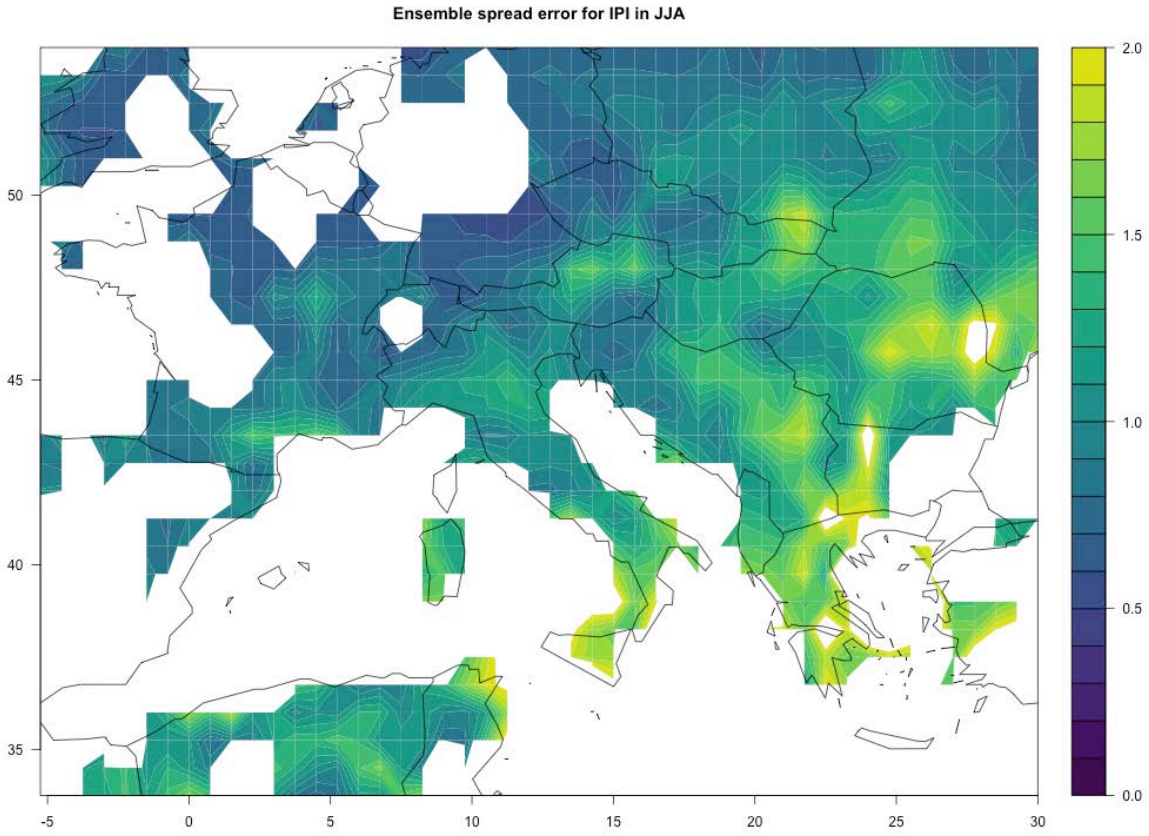

Figure 3.7.4: Spread to error ratio of JJA IPI forecasts with ECMWF System 4 for the period 1995-2010. Spread to error ratios larger than unity indicate forecasts that are over-dispersive, spread to error ratio smaller than unity indicate forecasts that are over-confident.

3.7.3.4. CRPS

The continuous ranked probability skill score for IPI forecasts from 1995-2010 is close to zero everywhere in Europe (Figure 3.7.5). This suggests that forecasts from a dynamical forecasting system are as accurate (in CRPS terms) as a climatological forecast that is the same every year. The comparison with the ROC area skill score presented earlier, suggests that while there may be little skill in absolute IPI forecasts (as shown by the CRPSS), forecasts with reduced detail (e.g. in tercile categories instead of absolute numbers) may nevertheless be skilful for some areas. CRPSS is only marginally affected by the detrending (not shown).

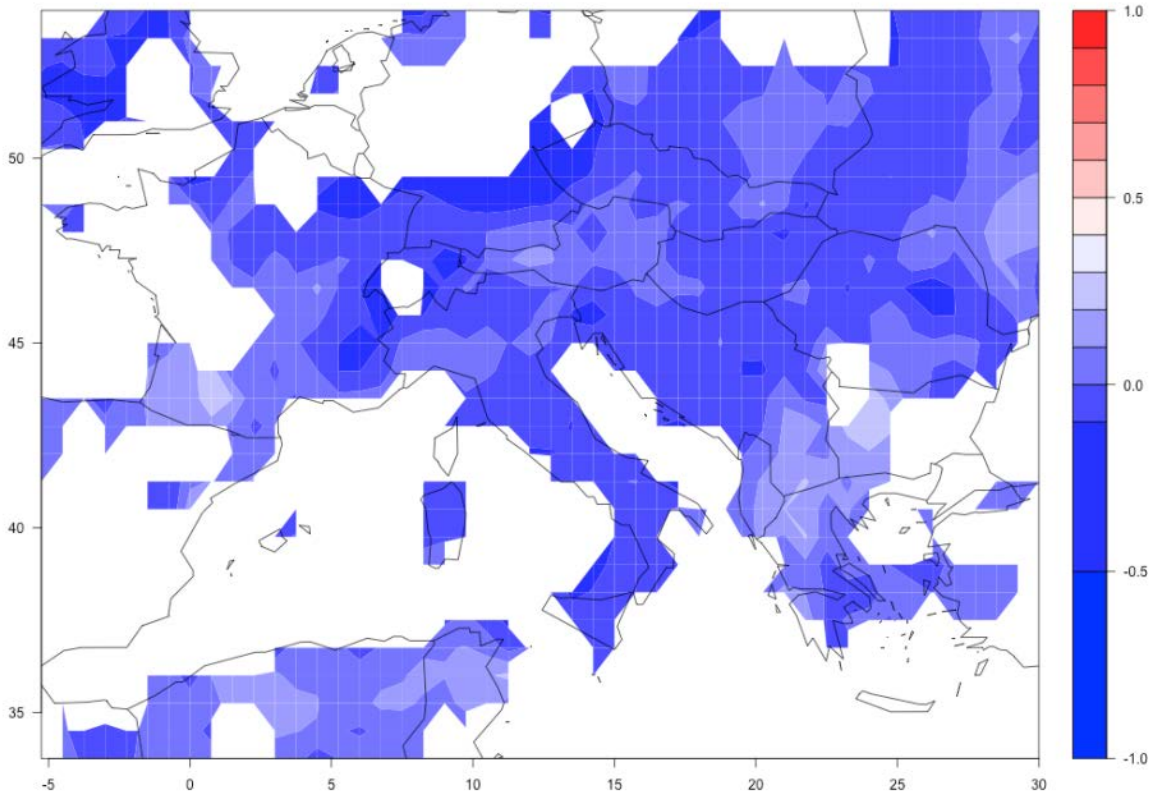

Figure 3.7.5: Continuous ranked probability skill score (CRPSS) of JJA IPI forecasts from ECMWF System 4 for the winters from 1995-2010. Seasonal IPI has been computed using the bias corrected daily total precipitation series from System 4. CRPSS has been computed with respect to a climatological forecast (using all the other years as benchmark). A correction for the effect of the limited ensemble size (both of the ensemble forecast and the climatological forecast) has been applied as proposed in Ferro et al. (2008).

Key Points: Intense Precipitation Index

- There is some skill for forecasts of the heavy precipitation index in summer in south-eastern Europe and south-western France. Elsewhere skill is fairly limited.

3.8. Water Balance and Drought in France

Water available for agriculture during the critical period of summer can be estimated by the seasonal precipitation amount over the June-July-August quarter. Classically, this information is given in both deterministic and probabilistic format. The deterministic format is the mean anomaly: comparison of forecast mean of the year of interest to forecast climatology, the latter calculated from the hindcast period. The probabilistic format is generally a probability to be above/below a threshold, for example a tercile.

These products can be assessed using deterministic (Anomaly correlation, Mean Square Score) and probabilistic scores (ROC areas, Brier scores). Figure 3.8.1 shows an example

of probabilistic score for the French seasonal model ARPEGE System 3, for each 2.5°*2.5° mesh of the model. Scores are calculated for the quarter June-July-August, for an initialization in May (Lead Time 1), at the European scale, and for the upper and lower terciles. Scores are globally low over Europe and in particular over France.

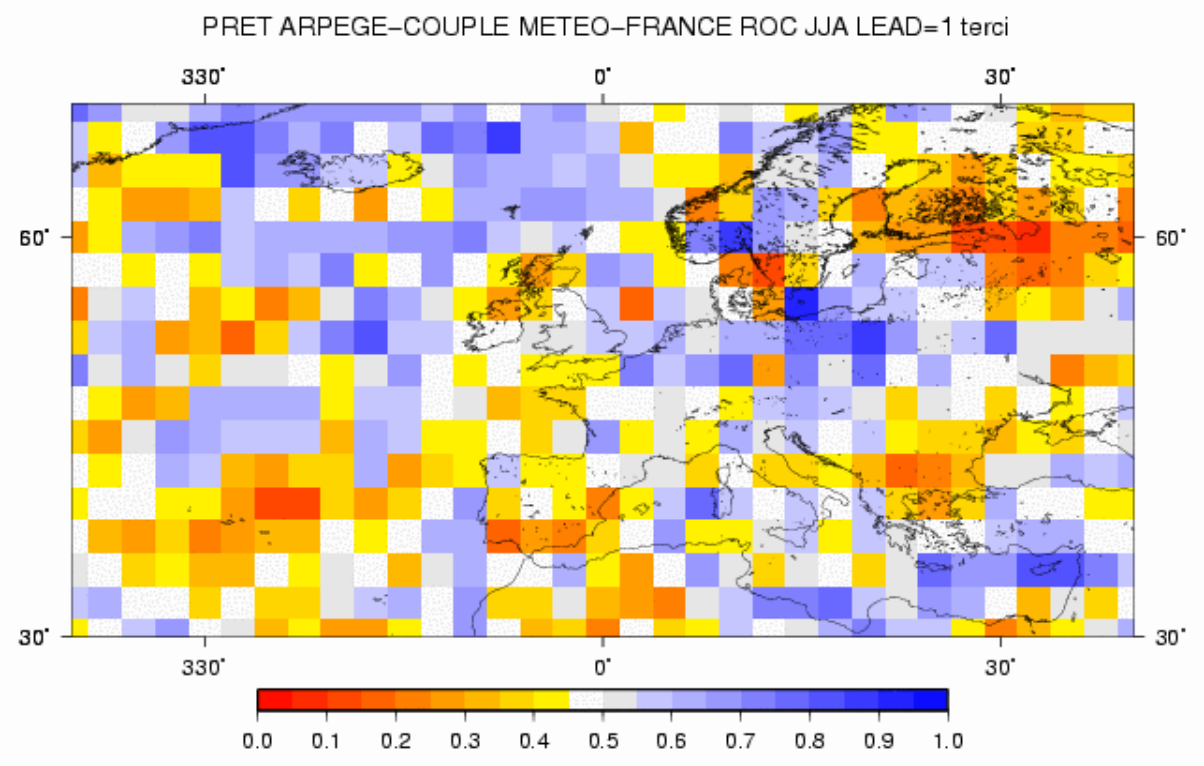

Figure 3.8.1: Map of ROC areas calculated for the lower tercile of precipitation for summer over Europe issued from ARPEGE System 3.

3.8.1. Context and Definition

Water balance and drought (related to agriculture) are well described using river flows and soil moisture. These two CIIs are provided by hydrological models coupled to seasonal forecast models. For example, Météo-France runs the seasonal hydrological forecasting suite called Hydro-SF (Céron *et al.* 2010) that is used in the climate service prototype named RIFF. Basically the hydrological suite (SIM) is forced by seasonal forecasts issued from the ARPEGE System3 model. The SIM suite is composed of two main modules ISBA (a soil-vegetation-atmosphere transfer scheme) and MODCOU (a hydrogeological routing module). River discharges are calculated over around 900 stations and soil moisture is calculated for around 10 000 meshes over France. Soil moisture is described by the Soil Wetness Index (SWI) averaged over the soil depth:

$$SWI = \frac{W - W_{wilt}}{W_{fc} - W_{wilt}}$$

with W the soil water content, W_{fc} the water content at field capacity and W_{wilt} the water content at the wilting point.

Priority has been put on river flows for which scores have been calculated for the 900 stations over France for an initialization in May. For SWIs, the feasibility has been evaluated

in Singla *et al.* 2012 and scores have been calculated for summer but with an initialization in April. SWI scores have not been calculated yet using an initialization in May.

The period of interest is MJJASOND (initialization at 1st of May with forecasts up to the 30th of November i.e. 7 months) in order to prevent drought during summer and the scores have been calculated for the 29 years of the hindcast period (1979-2007). Scores presented in the following paragraphs have been calculated for the June-July-August quarter and for each month of this quarter but here, only the results for JJA are presented.

3.8.2. Scores for Forecasted River Flows

3.8.2.1. Predictability Scores

3.8.2.1.1. Deterministic scores

The first score is the correlation in time of forecasted river flows using Hydro-SF with respect to the river flows issued from the reference SIM reanalysis, at monthly time step from May to November and also for the 3-months period JJA (Figure 3.8.2). The correlation allows checking the association of the mean river flow variations in time with respect to the reference (time-mean over each month, and mean over ensemble forecast).

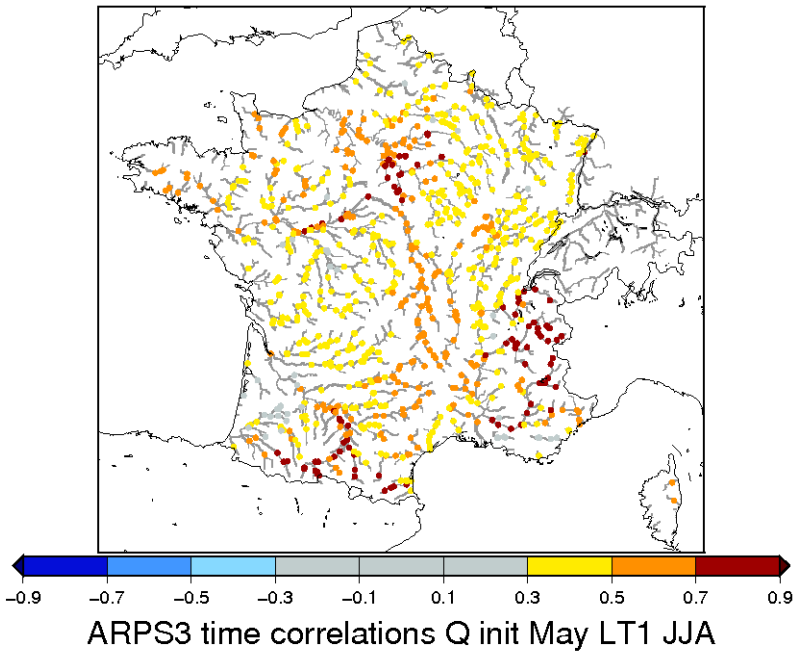

Figure 3.8.2: Correlation map of river flows between Hydro-SF and the SIM reanalysis reference for summer (month of initialization: May). Scores are calculated over the 1979-2007 period.

In a general way, forecasted river flows are quite well correlated in time with the river flow reference and especially over the Seine basin (around Paris) thanks to the presence of an aquifer whose groundwater contributes for around 60-80 % to the river discharge, and also over the Alps and Pyrenees where the snow cover brings predictability.

3.8.2.1.2. Probabilistic scores

The ROC score has been calculated for the lower tercile. It shows the performance of the model as its discrimination threshold is varying. Here it shows that the predictability of low river flows is very high for all the stations for this period (Figure 3.8.3).

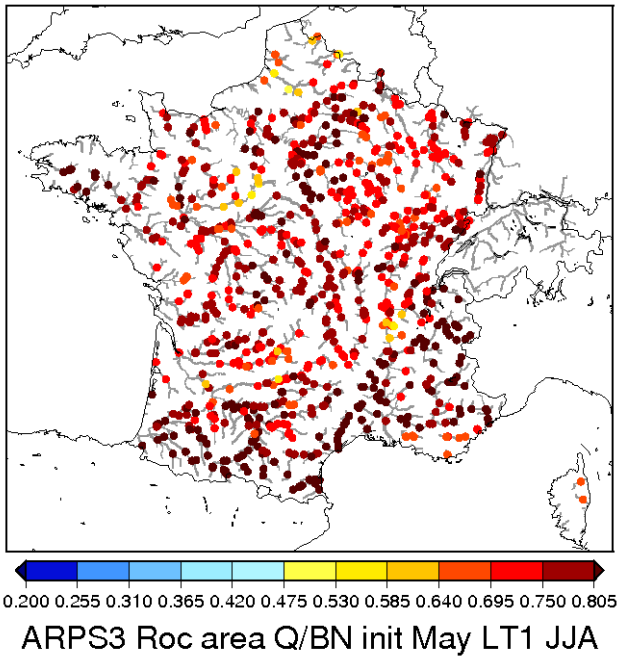

Figure 3.8.3: Map of Roc areas calculated for the lower tercile of river flows for summer (Month of initialization: May). Scores are calculated over the 1979-2007 period and SIM reanalysis is the reference.

Brier scores have been also calculated showing more reserved results (not shown). The Brier score is used to quantify the ability of an ensemble forecast to predict an exceedance (or non-exceedance) of a threshold.

3.8.2.2. Skill Scores

Skill scores are based on the comparison between river flows forecasted by the Hydro-SF suite (SIM is forced by ARPEGE S3 forecasts) and those using quite the same suite, except that random atmospheric forcings replace seasonal forecasts. These atmospheric forcings are randomly taken among the 29 years of the hindcast. Hereafter the first experiment is called Hydro-SF and the second is called RAF for Random Atmospheric Forcings. The RAF experiment is considered as being similar to climatological forecasts. The comparison of the two experiments highlights the skill of the forecasting suite and allows evaluating if there is an added value when using seasonal forecasts or not.

As for the previous scores, only results for the lower tercile are shown.

3.8.2.2.1. Deterministic skill scores

In order to make comparisons between the seasonal hydrological forecasting suite (Hydro-SF) and the random atmospheric forcing experiment (RAF), a bootstrapping method was used with a Student test on the difference of time correlations (Figure 3.8.4). It shows that

river flows issued from the Hydro-SF suite are a little bit better correlated in time than those issued from the RAF experiment (in yellow).

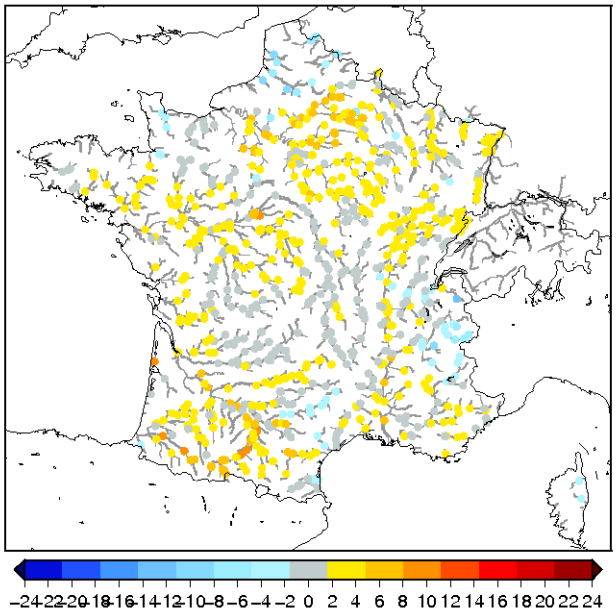

Figure 3.8.4: Map of Student variable of the difference of correlation between Hydro-SF and the RAF experiment for the lower tercile of river flows, for summer.

3.8.2.2.2. Probabilistic skill scores

For the Brier Skill Score (BSS, Figure 3.8.5), the results are not so clearly in favour of the Hydro-SF experiment since it exists some areas and, especially a part of the Seine basin, where the RAF experiment is more skilful (in blue) than the Hydro-SF. It could be partially explained by the contribution of the aquifer which is modelled in both experiments. To investigate the differences between the two experiments for this score, a more in depth investigation should be done in particular by analysing the contribution of the three terms of the Brier score (reliability, resolution, uncertainty).

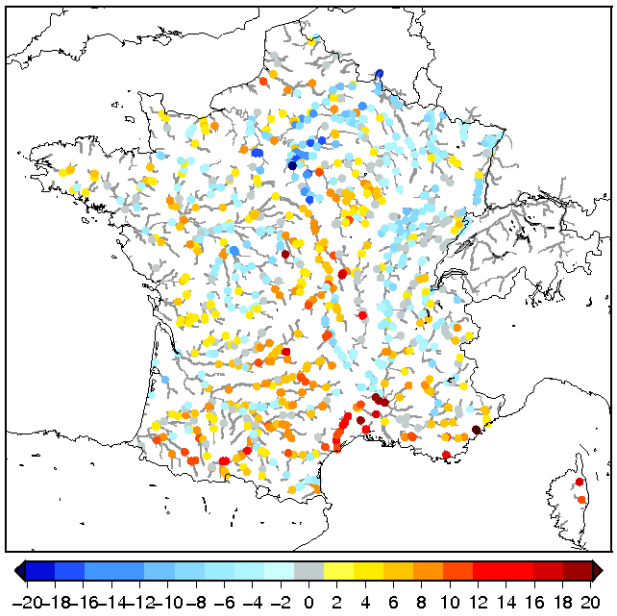

Figure 3.8.5: Map of Student variable of Brier Skill Score for river flows between Hydro-SF and the RAF experiment for the lower tercile, for summer.

For the ROC score (Figure 3.8.6), the method consisting of using a bootstrapping with a Student test is a bit more complex. To assess the skill of the Hydro-SF experiment, we compare results obtained for ROC areas in both cases. Where ROC areas are between 0.475 and 0.640 (yellow) in the RAF experiment, ROC areas are above 0.640 (orange-red) in the Hydro-SF experiment. For these regions, Hydro-SF is more skilful and brings more information than using a climatological forecast. But, for the other regions, the added value of Hydro-SF is not obvious and using seasonal or climatological forecasts seems to be equivalent.

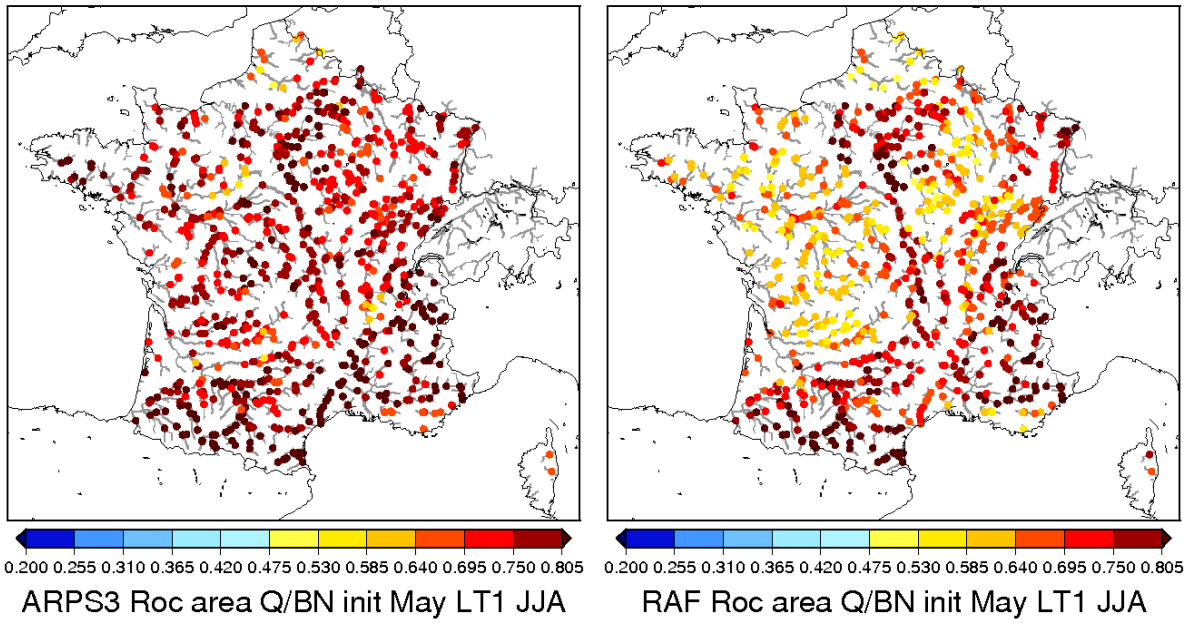

Figure 3.8.6: Map of Roc areas calculated for the lower tercile of river flows, for summer (Month of initialization: May) for the Hydro-SF experiment (left) and for the RAF experiment (right).

Key Points: River Flow

- Temporally and spatially integrated quantities such as river flow tend to exhibit more skill on seasonal time scales.
- Enhanced skill in river flow is due to the predictability in initial conditions of the hydrological model (e.g. soil moisture, snow) rather than predictability of meteorological input quantities such as rainfall.

3.9. Drought in Romania

3.9.1. Introduction

In this study we used as a drought metric the Palmer Drought Severity Index (PDSI) and associated indices such as the soil moisture anomaly index. PDSI measures the cumulated effect of monthly precipitation deficit (surplus) with respect to the optimum precipitation amount, the latter being defined as the precipitation amount needed to maintain the optimum soil moisture in order to support a normal plant growth (i.e. without water stress) in a given region (Palmer, 1965). Summer months (JJA) are important from the standpoint of water availability for crops in Romania so we have focused on predictions for the June to August interval with forecasts initialized in April and May.

3.9.2. Data and Methods

Computation of the PDSI index requires data for precipitation, air temperature, general soil conditions (available water capacity) and solar radiation conditions (i.e. latitude of the specific location). In order to calculate $PDSI_i$ for a certain month i, one has first to determine the index of soil moisture anomaly $ZIND_i$ for that month:

$$ZIND_i = k\,(P - \alpha PE - \beta PR - \gamma PRO + \delta PL) \quad (1)$$

$$PDSI_i = PDSI_{i-1} + ZIND_i/3\ - 0.103\ PDSI_{i-1} \quad (2)$$

where: k is an empirical weighting factor, specific for each region; α, β, γ, δ are coefficients for evapotranspiration, soil water recharge, runoff and water loss from the soil computed to link potential quantities and real ones; P, PE, PR, PRO, PL represent observed precipitation, Thornthwaite potential evapotranspiration (Thornthwaite, 1948), potential recharge, potential runoff and potential water loss from the soil. Potential Recharge is the amount of moisture required to bring the soil to its AWC from the available moisture at the beginning of the month. Runoff is assumed to occur if both surface and sub-surface layers of the Palmer soil model reach their combined moisture capacity, AWC. Potential Loss is the amount of moisture that could be lost from the soil provided the monthly precipitation is zero (Palmer, 1965).

Palmer (1965) built the index based on the components of the hydrological balance in a given area. Here we used the method and software developed by Wells at al. (2004) to compute self-calibrated PDSI values and related indices. The self-calibrated PDSI values numerically match the behaviour of the index at any location by replacing empirical constants of Palmer (1965) with newly calculated constants based on local climate. We computed PDSI and associated indices at 113 stations which cover Romanian territory for the interval 1961-2014. We used AWCs extracted from the European Soil Database (ESDB) for topsoil and subsoil (Hiederer, 2013a; Hiederer, 2013b).

We have investigate drought-related predictability using 2 approaches: (1) hindcasts with a statistical model based on canonical correlation analysis (CCA) of the monthly zonal wind anomalies at 200 hPa over Eurasia in April and the monthly Palmer soil moisture index, temperature and precipitation in June, July and August for the interval 1961-2014; (2) hindcasts based on seasonal predictions of the Palmer soil moisture index, temperature and precipitation from the system 4 of ECMWF for June, July, August starting from April and May (2011-2014) (with 0, 1 and 2 months anticipation). For ECMWF skill evaluations, all summer predicted and observed monthly values are used in order to have a larger sample (i.e. data is not stratified on anticipation interval).

3.9.3. Hindcasts Experiments with CCA Based Models

We identified a predictive signal linking zonal wind anomalies at 200 hPa over Eurasia (from 20 E to 120 E and from 30 N to 70 N) in April and the Palmer soil moisture anomaly index over Romania in June. This signal becomes weaker in July and August. We used the Climate Predictability Tool (http://iri.columbia.edu/our-expertise/climate/tools/cpt/) to build the statistical models, cross-validated their results and estimated the associated skills.

The Romanian stations where the performance indices show highest skill for the Palmer soil moisture index have lower skill for temperature and precipitation (e.g. correlation coefficients between observed and predicted values, ROC scores; see Table 3.9.1 and Figure 3.9.1). This suggests the existence of an added value in predictability of the Palmer index which incorporates temperature, precipitation but also a soil-related constant (the available water capacity - AWC).

Table 3.9.1: Correlation of the statistical model for the prediction of temperature, precipitation and Palmer soil moisture index (ZIND) in June from zonal wind at 200 hPa over Eurasia in April along with the ROC score for ZIND.

Station	Correlation Coefficient			ROC of ZIND	
	Temperature	Precipitation	ZIND	Above	Below
ARAD	-0.08	-0.10	0.26	0.55	0.71
BUCHAREST	-0.22	0.33	-0.20	0.51	0.37
CALARASI	-0.20	0.26	0.07	0.55	0.58
CRAIOVA	-0.20	-0.13	0.14	0.64	0.59
GALATI	-0.23	0.03	0.04	0.44	0.48
GIURGIU	-0.29	0.19	0.18	0.59	0.53
ROSIORI	-0.24	0.13	0.12	0.55	0.62
TIMISOARA	-0.13	-0.43	0.20	0.69	0.61
TURNU MAGURELE	-0.28	-0.15	0.31	0.65	0.76
URZICENI	-0.28	-0.33	-0.07	0.44	0.46

3.9.4. Hindcast Experiments Based on ECMWF Seasonal Predictions (System 4)

We extracted the predicted monthly anomalies of temperature and precipitation and expressed them in percent using as the reference the climate hindcast (1981-2010) of System 4. We interpolated predicted anomalies for summer months at the locations of 9 stations covering low areas in Romania. We added these anomalies to the observed climatologies (1981-2010) and we computed the predicted PDSI and associated indices (e.g. soil anomaly index) for June, July, August starting from the spring months April and May. As skill metrics we used correlation coefficients between predicted and observed values of PDSI and Palmer soil moisture index (ZIND).

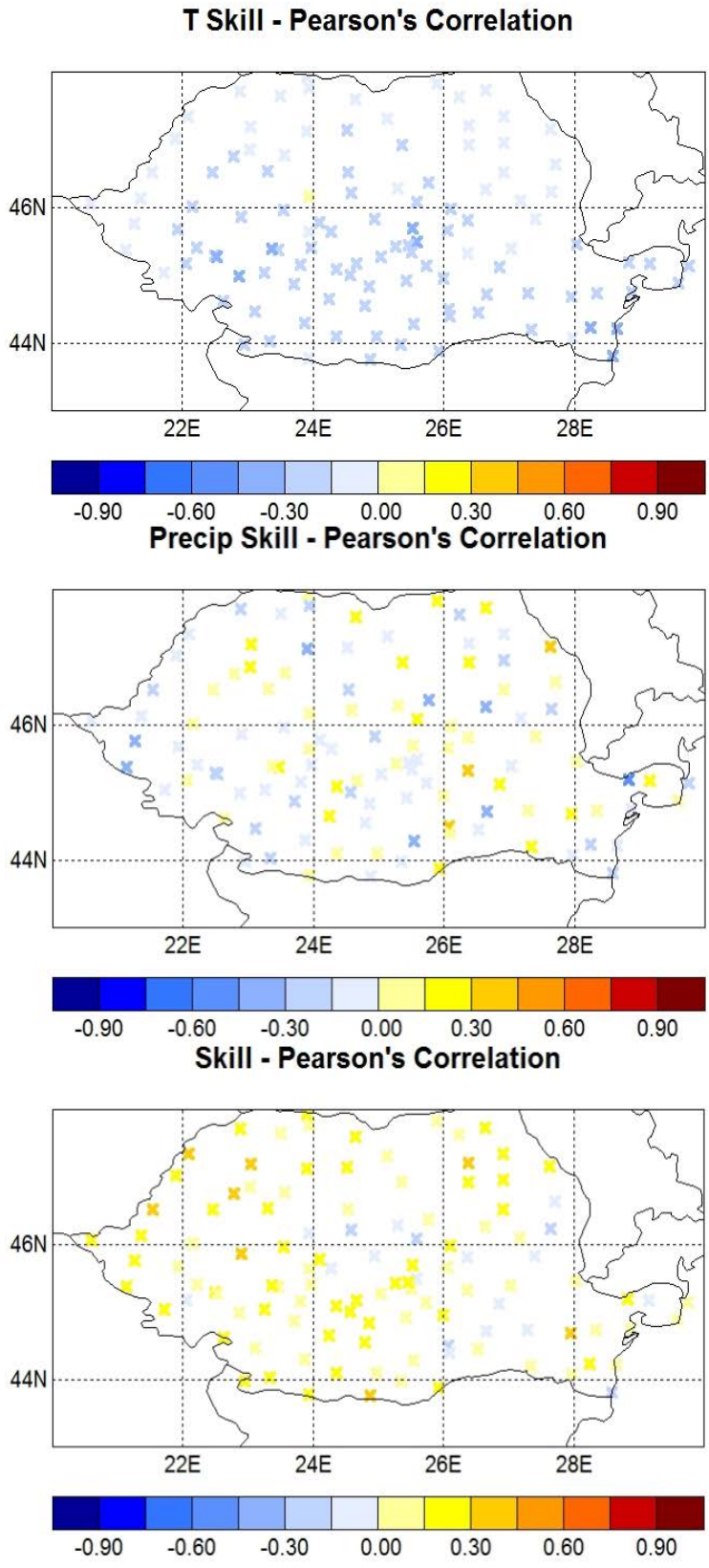

Figure 3.9.1: Correlation coefficients between observed and predicted values of temperature (upper panel), precipitation (middle panel) and soil moisture anomaly index of Palmer (bottom panel).

PDSI indices of summer months (JJA) computed from ECMWF predicted temperature and precipitation starting from April and May have generally better skill than ZIND (see Table 3.9.2). Correlation coefficients between ECMWF-predicted and observation-based PDSI, temperature and precipitation suggest to some extent the existence of an added value in predictability for the Palmer drought severity index which is not the case for ZIND. The higher predictability of PDSI compared with ZIND when using ECMWF predictions could be due to higher persistence of the PDSI and better skill of ECMWF model for June. Please note, however, that the correlation is based on four summers of forecasts only and the resulting coefficients are therefore subject to considerable sampling uncertainty. Figure 3.9.2 and Figure 3.9.3 show best predicted evolutions of summer PDSI at locations in Southern and Western Romanian plains.

Table 3.9.2: Correlation coefficients of the prediction of PDSI and Palmer soil moisture index in summer (JJA) from ECMWF predicted temperature and precipitation starting from April at 9 Romanian stations. Predicted values from System 4 are used.

Station	Correlation coefficient	
	ZIND	PDSI
ARAD	-0.61	0.81
BUCHAREST	0.14	0.41
CRAIOVA	-0.07	0.40
GALATI	-0.19	1.00
GIURGIU	0.03	0.00
ROSIORI	0.34	0.51
TIMISOARA	-0.55	0.80
TURNU MAGURELE	0.33	0.80
URZICENI	0.34	0.73

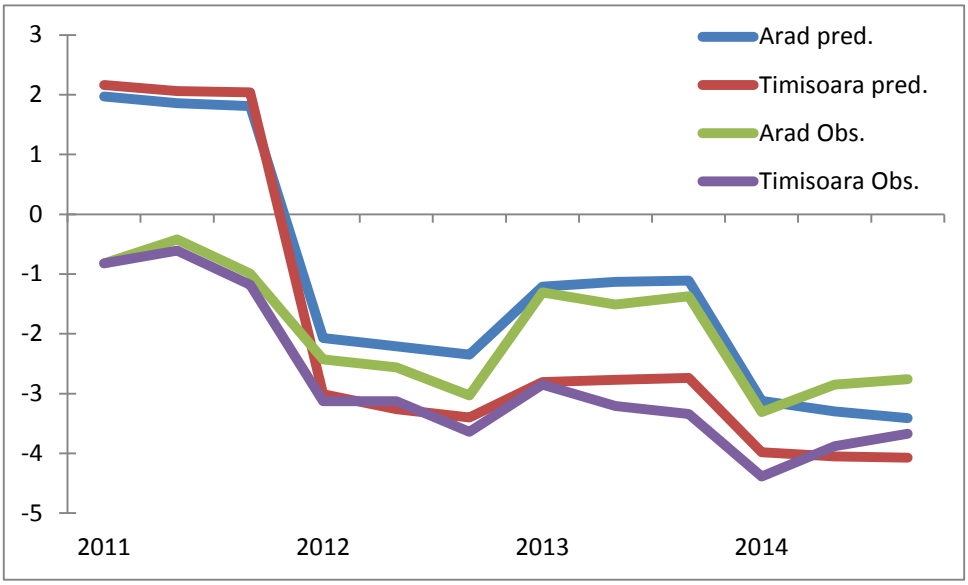

Figure 3.9.2: Temporal evolutions of observed-based and ECMWF-predicted values of PDSI in summer months (JJA) starting from May at 2 stations located in Western Romania.

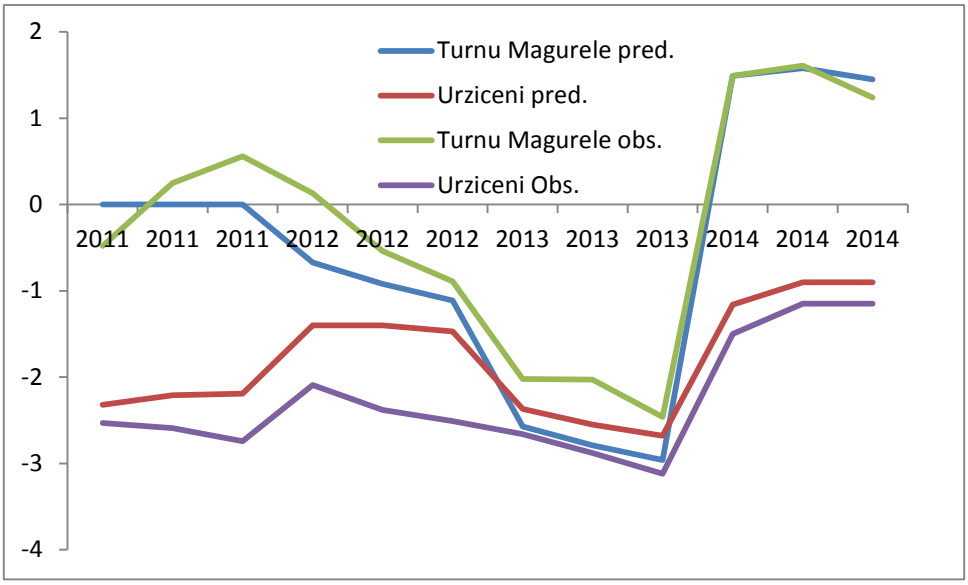

Figure 3.9.3: Temporal evolutions of based and ECMWF-predicted values of PDSI in summer months (JJA) starting from May at 2 stations located in Southern Romania.

3.9.5. Conclusions

We identified a predictive signal linking zonal wind anomalies at 200 hPa over Eurasia (from 20 E to 120 E and from 30 N to 70 N) in April and the Palmer soil moisture anomaly index (ZIND) over Romania in June. The Romanian stations where the performance indices show highest skill for the ZIND have lower skills for temperature and precipitation (e.g. correlation coefficients between observed and predicted values in Table 3.9.1; see also Figure 3.9.1). This fact suggests the existence of an added value in predictability of the Palmer index which incorporates temperature, precipitation but also a soil-related constant (the available water capacity - AWC).

PDSI indices of summer months (JJA) computed from ECMWF predicted temperature and precipitation starting from April and May have generally better skill than ZIND. Correlation coefficients between ECMWF-predicted and observation-based PDSI, temperature and precipitation suggest to some extent the existence of added value in predictability for the Palmer drought severity index which is not the case for ZIND. The higher predictability of PDSI compared with ZIND when using ECMWF predictions could be due to higher persistence of the PDSI. However, these are preliminary conclusions which will be further substantiated for more stations covering Romanian territory.

Key Points: Drought in Romania

- Statistical seasonal forecasts suggest an added value of forecasts of the Palmer drought severity index over forecasts of temperature and precipitation respectively.
- Also, predictability of the Palmer drought severity index seems to be enhanced compared with predictability of the soil moisture anomaly index.

3.10. Tropical Drought

3.10.1. Definition and Equation

The Water Requirement Satisfaction Index (WRSI) measures crop performance based on the balance between water supply and demand during the growing season. Usually, the computation of the water balance is updated with a frequency of ten days. During each ten-day period, the WRSI is computed as the ratio between evapotranspiration and the water requirement of the crop.

By considering AET (Actual Evapotranspiration) – a function of water availability in the soil - and WR (Water Requirement) – a function of atmospheric conditions and plant growth phases, WRSI is determined by the following relation:

$$WRSI_i = 100 \cdot AET_i \,/\, WR_i$$

The underlying conceptual scheme is that of a bucket which is replenished by rainfall and depleted by evapotranspiration. A critical step in the computation of WRSI is in the update of the soil water content. If during a given ten-day period the sum of soil water content plus the cumulated rainfall is less than the plant water requirement, then a water deficit is recorded. In more specific terms, if AET is less than the WR determined by atmospheric conditions and by the plant's growing phase, the plant suffers a determined level of water stress. Conversely, if the sum of soil water content plus the cumulated rainfall exceeds the plant water requirement there is no water deficit

The WR can be calculated by adjusting the "Potential evapotranspiration" by the specific characteristics of the plant at a given growing phase. It is computed as follows:

$$WR_i = PET_i \,/\, Kc_i$$

where i indicates the ten-day period, PET is the Potential evapotranspiration during the considered period and Kc is a crop coefficient, which depends not only on the crop in object but also on the particular growing phase of such crop.

PET (also known as "ETo" in FAO terminology) can be defined as the evapotranspiration rate from a reference surface (a hypothetical grass reference crop with specific characteristics), not short of water, and is a function of local weather parameters, such as solar radiation, air temperature, wind speed, and humidity. As PET depends mainly on solar radiation, which is fundamentally an astronomical parameter, climatological tables for this parameter are usually considered as representative of the actual value.

The WRSI data considered in the present analysis have been computed by using the LEAP software available here http://hoefsloot.com/new/?software=leap-development

The objective of the analysis presented in the following sections is to determine whether the aggregated drought index adopted for this study case, shows similar, better, or worse predictability than the raw physical variable (rainfall) which is adopted as an input to compute the index.

The results are illustrated here by evaluating the skill of the forecasts against a reference WRSI dataset, which is derived by using the ARC2 satellite rainfall estimate as an input to

compute the drought index, which is to date the most widely adopted as an input to the LEAP platform.

3.10.2. Anomaly Correlation

To compare the skill in predicting WRSI with the skill in predicting the corresponding rainfall input, we consider the summer (JJAS) cumulated rainfall as a reference simplified indicator. Figure 3.10.1 shows the anomaly correlation of the ensemble average with the *observed* rainfall and WRSI.

Note that WRSI (right panel) is computed only on a limited subdomain, which covers only the crop area of interest for the main crop season in Ethiopia. The south eastern part of the country is usually too dry to sustain agriculture; therefore WRSI is of no practical relevance. The same applies to the western part of the country, where the annual rainfall largely exceeds the requirements for rain-fed agriculture.

In the left panel, rainfall data are aggregated to the coarser resolution of the global forecasts. Instead, for WRSI (left panel) the native resolution of the computation performed in LEAP is retained.

Figure 3.10.1 shows that a positive correlation skill is concentrated mostly in the north of the country. In particular, WRSI has a high correlation (corr > 0.5) in areas where the cumulated rainfall shows no significant skill. This result is already an interesting indication that the aggregated drought index (WRSI) increases the predictability with respect to the corresponding rainfall values. Indeed, by construction, the WRSI focuses on capturing *negative* deviations from the normal precipitation regimes. Therefore, discrepancies in capturing wet events do not contribute to the overall skill.

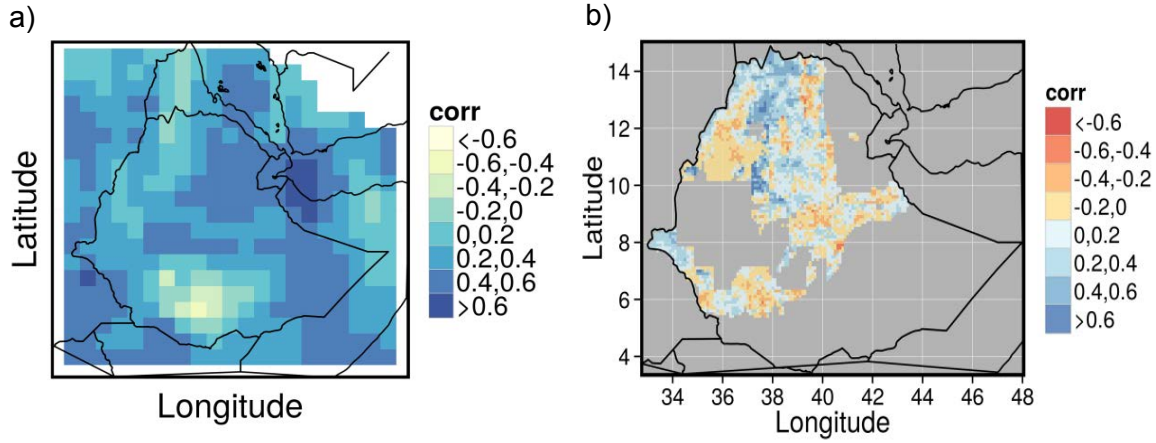

Figure 3.10.1: Ensemble mean, anomaly correlation for (a) JJAS cumulated rainfall and (b) WRSI. The reference observational rainfall data is ARC2.

3.10.3. BSS

In Figure 3.10.2 a similar comparison as in the previous section is presented for the Brier Skill Score. Similarly to the case of anomaly correlation, the BSS for WRSI shows a positive skill in the north of the country in areas where the raw cumulated rainfall has much lower skill.

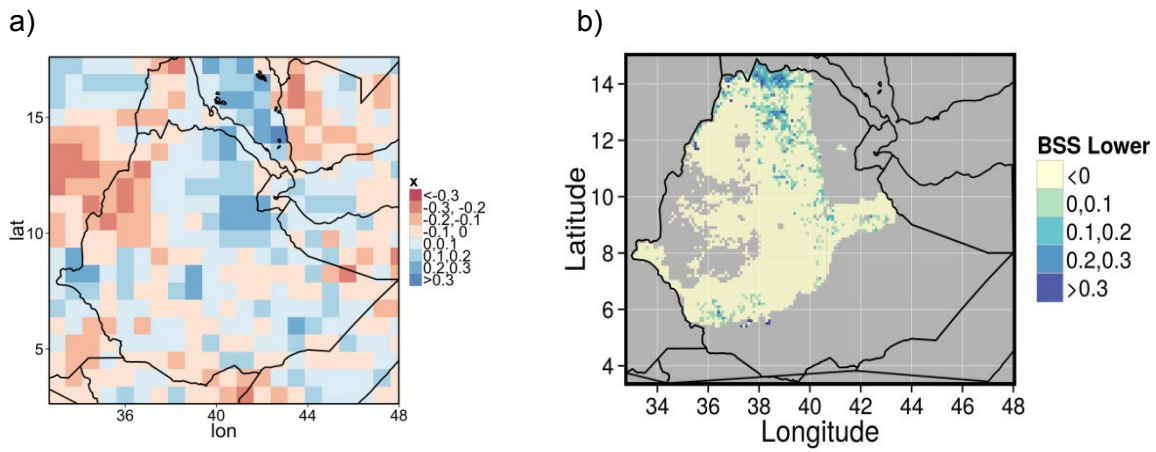

Figure 3.10.2: Brier Skill Score for the lower tercile for (a) JJAS cumulated rainfall and (b) WRSI. The reference observational rainfall data is ARC2.

Key Points: Tropical Drought

- The comparison of skill of forecasts of rainfall and the water requirement satisfaction index provide indication that skill in the index exceeds the skill in the underlying variable in some areas.
- In contrast to mean rainfall, the aggregated drought index is sensitive to episodes of water deficit (i.e. below normal rainfall) only.

3.11. Fire Danger

Wildfires represent the most important natural hazard in the EU-Mediterranean region, where an average of 4,500 km2 of forested and shrubland areas burn every year (San Miguel Ayanz *et al.* 2013), causing economic and environmental damages and loss of life every year. Here, estimating fire risk a few months in advance is therefore an urgent requirement, allowing fire protection agencies a timely reaction and an adequate provision of human and material resources. However, to date studies addressing the seasonal predictability of fire danger are still relatively scarce in the literature, and almost all of them follow an empirical approach to the problem, by statistically linking antecedent climatic variables used as predictors with observed fire activity (e.g. Chu *et al.*, 2002, Preisler and Westerling 2007, Chen *et al.*, 2011). This approach has been successfully used in certain EU-Med environments (Turco *et al.*, 2013, Gudmundsson *et al.,* 2014) with promising results, although the calibration of such models is largely dependent on the domain selected. In addition, a few studies explore the use of GCM outputs to seasonally predict fire danger (Roads *et al.*, 2005) or fire activity directly (number of fires and burned area, Marcos *et al.*, 2015).

In this deliverable, we use hindcast data from the System4 seasonal hindcast of 15 members to produce a forecast of the Canadian Forest Fire Weather Index System (FWI), a fire danger indicator widely used in Europe and worldwide. We analyse several aspects regarding the FWI forecast quality as compared to the reference observations using a number of standard forecast verification metrics. In addition, we address the effect of

Quantile mapping (QM) techniques on the resulting forecast, as well as some methodological issues regarding the application of statistical correction techniques (and in particular QM methods) to ensemble forecast data for the calculation of CIIs integrated by the combination of different input variables, such as FWI. In principle, there are two approaches to bias correct seasonal forecasts for use with climate information indices. On the one hand, bias correction can be performed directly on the CII (QMd hereafter, "d" stands for "direct"). On the other hand, bias correction can be performed on the model output variables before computing the CII (QMc, "c" stands for "components"). The latter may be necessary if the absolute quantity of the underlying variables is important for the computation of the CII, or if consistency across multiple CIIs derived from the same underlying variable(s) is required. This issue is analysed in the perfect-prog downscaling framework by Casanueva *et al.* (2014), but to date an analysis in a bias-correction framework for a seasonal forecasting application is lacking. As a result, in this deliverable we assess the effect of both types of approaches (QMd *vs* QMc) on the resulting FWI predictions.

The objective of the contribution to this deliverable is to better understand the ability of seasonal forecasts of FWI to adequately predict the observed inter-annual variability over the Euro-Mediterranean region (EU-Med). To achieve this objective we have undertaken the following:

- Generation of FWI forecast maps using the ECMWF's System4 seasonal hindcast of 15 members
- Verification of the seasonal forecast skill of the fire danger index used
- Assessment of two different approaches for the correction of FWI:
    - Direct approach: correction of the pre-calculated FWI (QMd hereafter)
    - Component approach: independent correction of the different components prior to FWI calculation (QMc hereafter)
- Verification of the seasonal forecast skill of the underlying variables, in order to assess the possible additional value of the multi-variable index.

3.11.1. Data and Methods

3.11.1.1. Methods

A number of fire danger indices are used globally and for specific regions, many of these derived empirically based on observed fire danger and severe weather incidence in particular locations. One of the most frequently used, and most widely studied indices is the Canadian Fire Weather Index (FWI hereafter), also used extensively across Europe. This deliverable focuses on the skill of seasonal forecast of this particular index; however it should be noted that different fire danger indices may display different levels of skill for different seasons and regions, and further work considering other indices would add significantly to our understanding of the potential utility of such indices in decision making.

The Canadian Fire Weather Index System (van Wagner, 1987) has been found to model fire potential in a broad range of fuel types worldwide (Bedia *et al.*, 2015) and is currently a worldwide reference for the estimation of fire danger (deGroot *et al.*, 2006). It has also been successfully applied for fire activity prediction (e.g. Viegas *et al.,* 1999, Bedia *et al.* 2014b), and also applied in the context of future fire danger estimation in EU-Med (Moriondo *et al.*

2006, Bedia *et al.* 2013, 2014a). As a result, FWI has been proposed as a relevant Climate Impact Indicator (CII) in the context of EUPORIAS WP22, in the category of forest fire danger. A detailed description of the index is provided in D22.1. FWI lacks a compact mathematical definition. However, the full code used for FWI calculation in this deliverable (in R language, R Core Team 2015) is available through this link[1]. The interested reader is also referred to the original document describing FWI calculation and the physical meaning of its different components (van Wagner 1987).

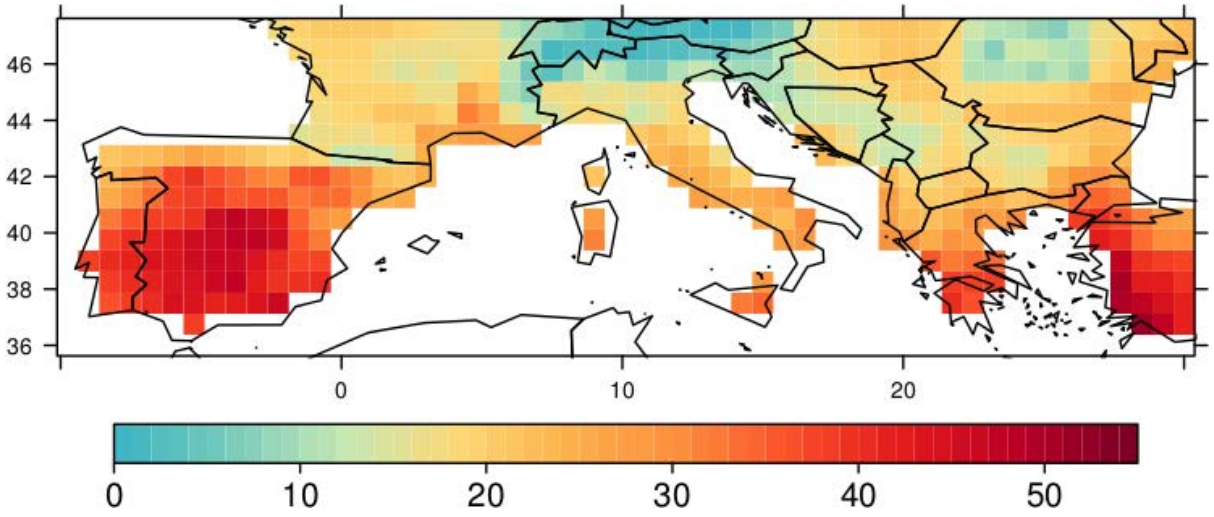

Figure 3.11.1: FWI climatology for the fire season (JJAS) and the 30-year period 1981-2010 according to the reference WFDEI dataset.

3.11.1.2. Observations

Climate data at an appropriate temporal resolution were obtained from the publicly available WATCH Forcing Dataset-ERA Interim (WFDEI, Weedon *et al.,* 2011 and 2014), with a native spatial resolution of 0.5º and 3-hourly time resolution. As a proxy of noon local time, we used the time of maximum solar incoming radiation of the WFDEI dataset for each grid cell and day of the year. More details on this particular dataset and the time resolution of the input variables is included in D22.1.

3.11.1.3. Seasonal Forecast Data

The seasonal forecast data is given by the ECMWF System-4 (Molteni *et al.*, 2011), a state-of-the-art fully-coupled general circulation model that provides operational multi-variable seasonal predictions at 0.75º horizontal resolution. In this study, we consider the 30-year re-forecast of the model (1981-2010) with a 15-member ensemble and 7-month lead-time for predictions. Molteni *et al.* (2011) describe in detail the S4 system and its performance. In particular, in this deliverable we analyse the lead-month 1 FWI predictions, having been earlier prediction horizons analysed in more detail (including a description of model drift effect) in D22.1.

[1] https://github.com/jbedia/fireDanger/blob/master/fwi.R

3.11.2. Bias Correction

To correct for model errors that vary not only as a function of lead time but also across the distribution of model outputs, quantile-quantile mapping is an often used bias correction method. In a multivariate context, QM also allows for multivariable correction in a (reasonably) consistent manner (see e.g. Wilcke et al 2013), as it is the case for the correction of the different input variables involved in FWI calculation. Our implementation of QM fits the hindcast daily empirical cumulative distribution functions (ECDFs) to corresponding observed ECDFs from the reference grids (WFDEI). To avoid potential deleterious effects of QM on the spread of the ensemble, we correct each member according to the joint multimember distribution.

To deal with the dependence of the bias correction on lead time and the seasonal cycle (i.e., model *drift*), we used a moving window to estimate the QM correction. For each day of year (1-365) in the calibration period, ECDFs are constructed using a moving window whose width (in days) depends on the user requirements (a summary of previous applications is shown in Table 1 of an internal EUPORIAS report by Bedia and Bhend, 2015). For instance, a window of 30 days results in 900 values for a 30-year period of calibration. The use of a moving window serves two different purposes:

1. The climatic variability for each particular day of the year is better described. Thus, the correction takes into account varying error characteristics throughout the year. As a result, the window should be wide enough to ensure that climatic variability for each particular day is adequately represented to provide robust estimates.
2. In the particular case of seasonal forecast model data, the use of a moving window can help to minimize the forecast time-dependent bias (model *drift*, analysed in D22.1 for the particular case of FWI). To this aim, the window needs to be sufficiently narrow to encompass periods for which drift can be safely neglected.

In this deliverable we use a moving window of 31 days, as a compromise between the above mentioned points 1 and 2.

3.11.3. Verification

From the set of verification metrics agreed in the frame of WP22 for this deliverable, the continuous ranked probability skill score (CRPSS) has been discarded, as the main interest of forecasting FWI focuses on the upper tail of the distribution (years of high fire danger), while the performance of forecast on lower parts of the distribution is of lower relevance. For this reason, ROCSS has been computed for the upper tercile as a more convenient metric to assess forecast skill in this context. Furthermore, the anomaly correlation coefficient (ACC) has been also considered. All verification metrics have been computed using the easyVerification R package (Bhend 2015). Additional visualization plots have been used for the assessment of the skill over selected sub-areas on an inter-annual basis, in particular tercile plots (see e.g. Diez et al, 2011) and spread plots, the latter providing an overview of observed and forecasted series and the spread of the ensemble. For further details on the tercile plots, see the help of the R function ‘tercileValidation’ in the R package downscaleR (Santander Meteorology Group 2015).

The presence of trends shared by both the observations and the model, particularly for temperature over extensive regions of the study area, warns against the use of raw data for

verification. Thus, we present all the verification metrics based on detrended data. As an example, in Figure 3.11. the effect of trends in the verification step is shown, in particular considering the ROCSS for the upper tercile of temperature. While the effect is not dramatic, local differences in certain regions are noteworthy. In particular, forecast skill as measured by ROCSS is reduced with detrending for example in southern France, Switzerland and northern Italy. This suggests that there is limited predictability of inter-annual variability in these regions. In Hungary, on the other hand, detrending leads to an increase in skill. This in turn may be due to a misrepresentation of trends in the seasonal forecasts.

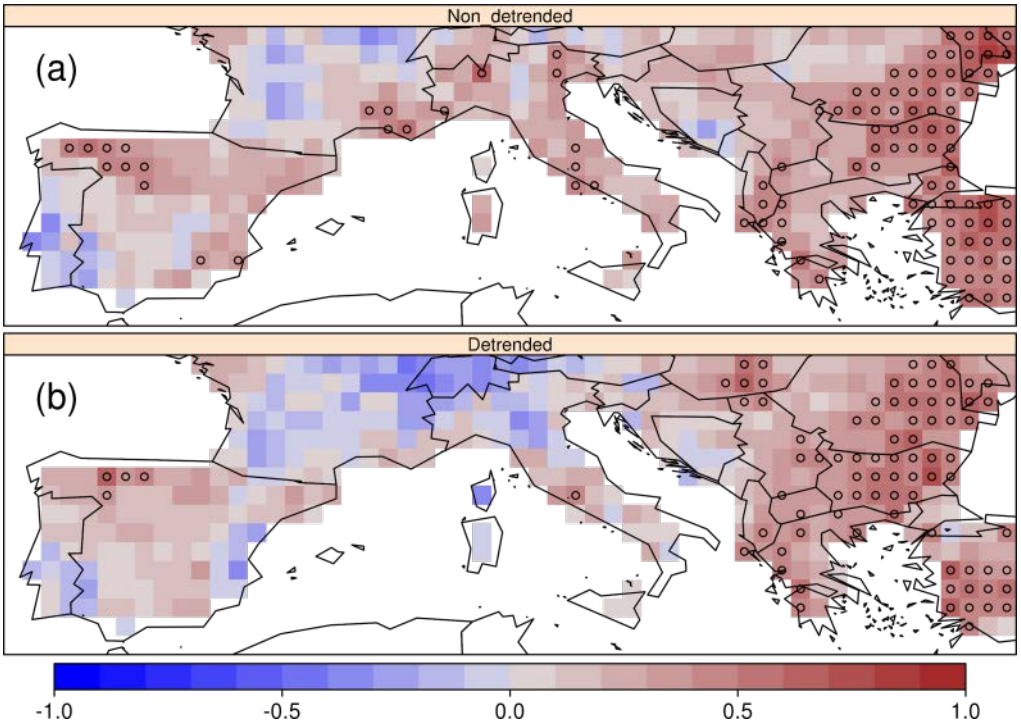

Figure 3.11.2: ROC Skill Score Maps for the upper tercile of temperature after QM correction considering: (a) raw, undetrended data and (b) detrended data. Circles indicate significant ROCSS values (95% c.i.).

Significantly positive ROCSS values were obtained in the eastern and north-eastern corner of the domain (Greece and Bulgaria mainly) for the upper tercile of FWI (Figure 3.11.3). Please note that we show only upper tercile forecast skill (i.e. the above-normal fire danger years). Upper tercile FWI forecasts are the most relevant from the end user standpoint, as these encompass the most dangerous situations in terms of fire ignition and spread, the remaining terciles being less relevant for fire prevention. Equivalently, in the case of relative humidity we will focus on the lower tercile, linked to high FWI situations (see e.g. Fig. 2 in Bedia et al 2012).

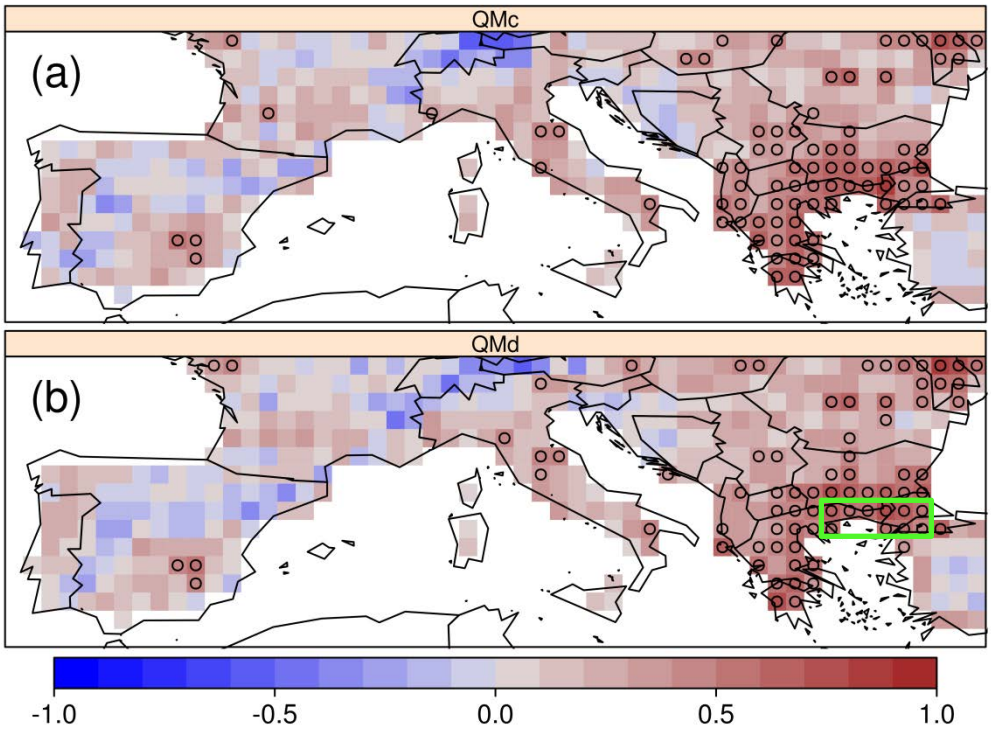

Figure 3.11.3: ROC skill score of the (a) QMc and (b) QMd FWI forecast (detrended), for the upper tercile. The green box depicted in panel (b) is used as a reference area for a more detailed analysis of model skill in section 3.9.3.1. Circles indicate significant ROCSS values (95% c.i.).

The comparison of Figure 3.11.3b (detrended QMd FWI) and Figure 3.11.b (detrended QM temperature) and Figure 3.11.4 (detrended QM relative humidity) suggests that there is not much gain in forecast skill of FWI as compared to that of its input variables, and that most of the skill attained by FWI can be attributable to the additive effect of the skills of relative humidity and temperature. Neither wind, nor precipitation shows a significant skill in the study area in terms of ROCSS[2] (not shown).

While the results of both forms of FWI correction (QMd and QMc) had no significant effect on the validation results (Figure 3.11.3), further analyses may be needed to more thoroughly assess the implications of both approaches in the preservation of trends, and other aspects potentially affecting the verification and the applicability of the forecasted variables. According to our results, QMd may represent a preferred approach given its relative simplicity as compared to QMc, requiring the application of the correction just once. Thus, the results presented hereafter correspond to the QMd approach.

[2] For instance, a preliminary report on the validation of precipitation for System4 seasonal 15 members with further verification metrics plus reliability diagrams is available through this link: http://meteo.unican.es/trac/attachment/wiki/udg/ecoms/dataserver/datasets/validation_report_System4_15members_precip.pdf

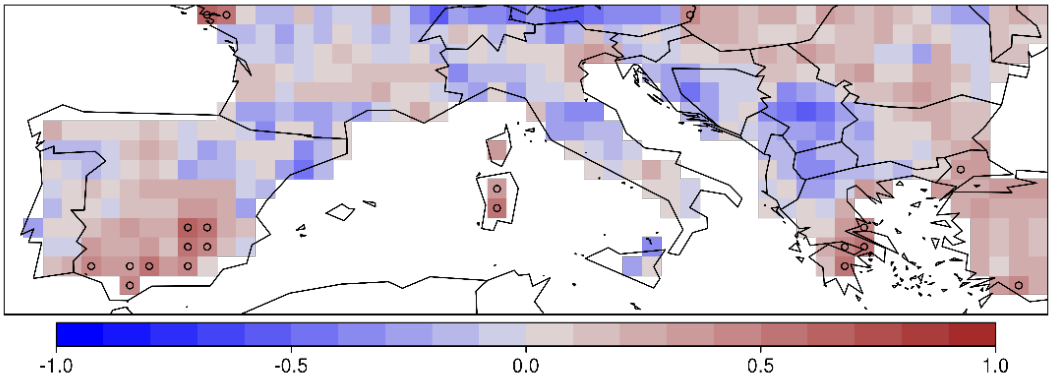

Figure 3.11.4: Same as Figure 3.11.b, but for the lower tercile of relative humidity.

Additional skill measures are presented in the following, which in general corroborate that the only region where seasonal forecast may prove useful for anticipating fire-season FWI one month in advance is the eastern part of the study area. In particular, It has been shown empirically that ACC values around 0.6 correspond to the range up to which there is synoptic skill for the largest scale weather patterns (ECMWF 2015), which are found in certain areas of Greece (Peloponnese), Turkey, Bulgaria and NE corner of the study area (Figure 3.11.5).

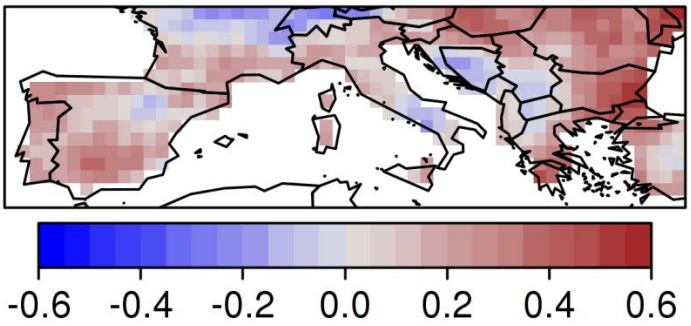

Figure 3.11.5: ACC of QMd detrended forecast FWI against the observed reference.

3.11.4. Regions with Skill

In this subsection we focus on the region within the whole domain that has exhibited some degree of skill (its spatial boundaries are depicted by the green box in Figure 3.11.3b). Even though this is not very high, as said before it may prove useful within an operational framework. Interestingly, the eastern part of the Euro-Mediterranean region has been identified as particularly sensitive to fire-weather interannual variability in terms of burned area (Bedia et al 2015), stressing the potential usefulness of this source of skill. In particular, we show a tercile plot depicting the observed interannual variability of FWI related to the forecast predictions for each particular year (Figure 3.11.6) where the performance of the prediction system can be visually inspected. It can be seen that 8 out of 10 of the years falling in the upper tercile of FWI (indicated by the white circles) were predicted by the model with the highest probability (i.e., the fraction of ensemble members falling in that category was the largest as indicated by the colour bar).

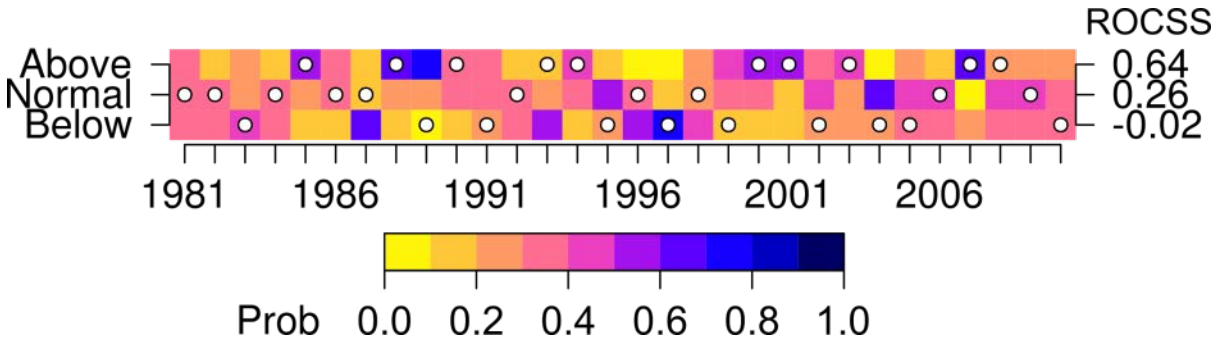

Figure 3.11.6: Tercile validation plot. Data represented correspond to the spatial mean of the box indicated in Figure 3.11.3b (10 grid boxes), an area where some skill has been found.

Correspondingly, Figure 3.11.7 shows the multimodel spread (the IQ range is represented) and the interannual evolution of the observed and predicted mean FWI for the fire season. The tercile boundaries are also indicated by the horizontal dashed lines.

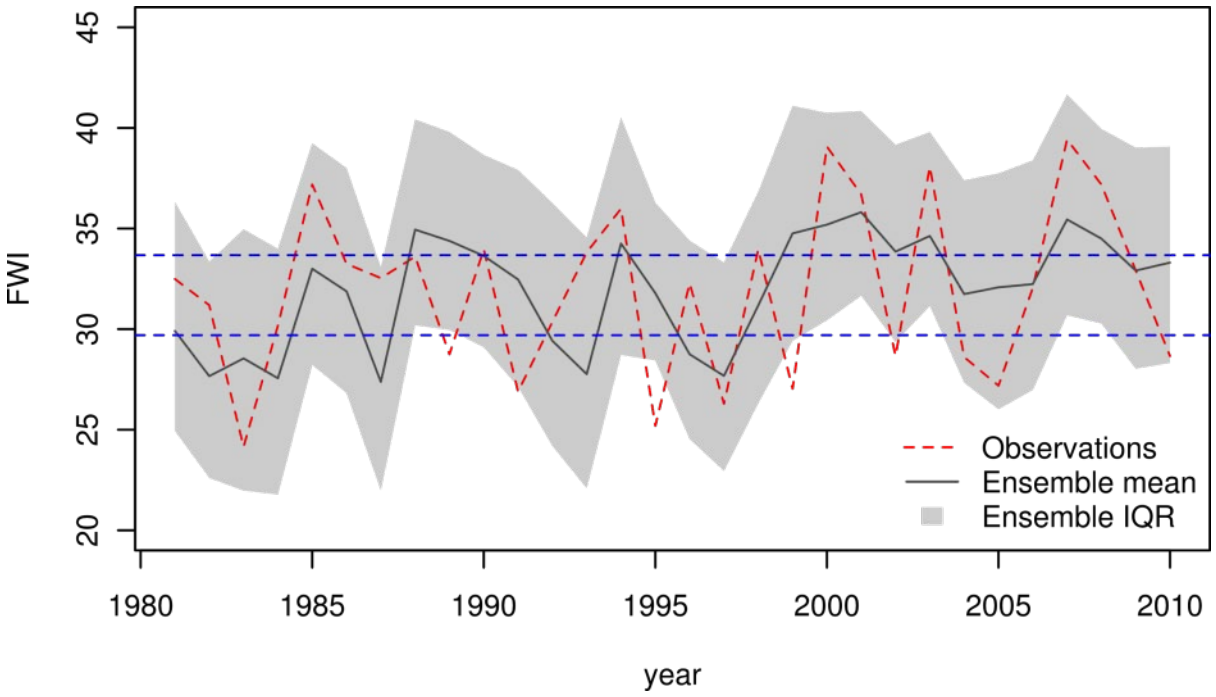

Figure 3.11.7: Observations (red dashed line) and ensemble mean (grey solid line) and spread (interquartile range, grey shadow). Data represented correspond to the spatial mean of the box indicated in Figure 3.11.3b (10 grid boxes), an area where some skill has been found. The blue dashed horizontal lines indicate the FWI terciles.

3.11.5. Conclusions

In light of the results presented, the main conclusions regarding seasonal forecasts of fire weather in southern Europe are as follows:

1. Both long-term trends and inter-annual variability may be predictable to some extent. Here, our aim is to study skill in predicting the latter. Hence we advocate to systematically remove trends present in the data to obtain more reliable estimates of skill in predicting inter-annual variability.

2. No significant effect of both QM approaches was found for FWI forecast skill. As a result, we opted for the use of QMd as a more convenient and computationally less demanding approach. However, changes in the predicted trends may be altered by QMc (not shown), which will be subject of further research.
3. The forecast skill for above average FWI across the region as a whole was not improved in comparison to the forecast skill of the underlying variables (this was tested on temperature and relative humidity only, by looking at their respective ROCSS). However further analysis is needed to comment on the skill in individual locations and with other input variables that may prove skilful in other world areas (e.g. precipitation).
4. In general, the forecast skill was poor in the domain of analysis. However, the eastern and south-eastern areas of analysis exhibited some degree of skill, suggesting the potential usefulness of System4 forecasts of FWI in this region for early warning of particularly dangerous fire seasons. These forecasts may be improved through the application of statistical downscaling techniques where local historical weather records of sufficient quality are available.

Key Points: Fire Weather

- There is significant skill in predicting summer fire weather in south-eastern Europe, elsewhere forecast skill for fire weather is poor.
- Skill in forecasting fire weather seems to be insensitive to the choice of approach for bias correction applied (i.e. correcting daily inputs vs. correcting the index).

3.12. Temperature Related Mortality

For the health case study, the performance of a climate-driven mortality model to provide probabilistic mortality predictions for heat wave and cold spell scenarios was assessed. The skill of the mortality forecasts using (a) ensemble forecasts of daily apparent temperature and (b) reanalysis daily apparent temperature, as inputs to the mortality model, was compared.

3.12.1. Definition and Equation

Daily mortality data corresponding to 187 NUTS2 regions across 16 countries in Europe were obtained from 1998–2003. Data were aggregated to 54 larger regions in Europe, defined according to similarities in population structure and climate. Location-specific average mortality rates, at given temperature intervals over the time period, were modelled to account for the increased mortality observed during both high and low temperature extremes and differing comfort temperatures between regions. The temperature–mortality dependency for each aggregation was estimated as follows: The range of observed temperatures was divided in equally spaced intervals. Days belonging to each interval were grouped and daily temperature and mortality data within each interval were averaged. Interval mean mortality was smoothed by means of a centred 31-term filter, corresponding to nearly 3°C intervals. The lowest value defines the interval of comfort temperature. This threshold divides the range of temperatures into 'warm' and 'cold' tails (see Ballester *et al.*,

2011 for further details). The model used to fit the temperature-mortality curves was formulated as follows:

$$y_{ik} \sim N\left(\alpha_j + \beta_{1j} x_{ik} + \beta_{2j} x^2{}_{ik} + \beta_{3j} x^3{}_{ik}, \sigma_j^2\right),$$

where y_{ik} is the logarithm of the average mortality rate (per million population) at region, i and temperature interval, k. Then, for each region i, the log mortality rate was formulated as a non-linear function of temperature, x_{ik}, (a third order polynomial), with location specific intercept, αj. Note that parameters are fitted separately for the warm tail ($j = w$) and cold tail ($j = c$), depending on whether the temperature is greater than ($x_{ik} \geq x_{im}$) or less than ($x_{ik} < x_{im}$) the comfort temperature (i.e., the temperature of minimum mortality), x_{im}. The models are fitted in a Bayesian probabilistic framework (Lowe *et al.*, 2015). Figure 3.12.1 shows the mean and 95% credible intervals (Bayesian equivalent of confidence intervals) obtained from the probabilistic simulations of the Bayesian probabilistic model, for each of the 54 regions. The region-specific comfort temperature, i.e. the temperature at minimum mortality, is indicated by a purple dot.

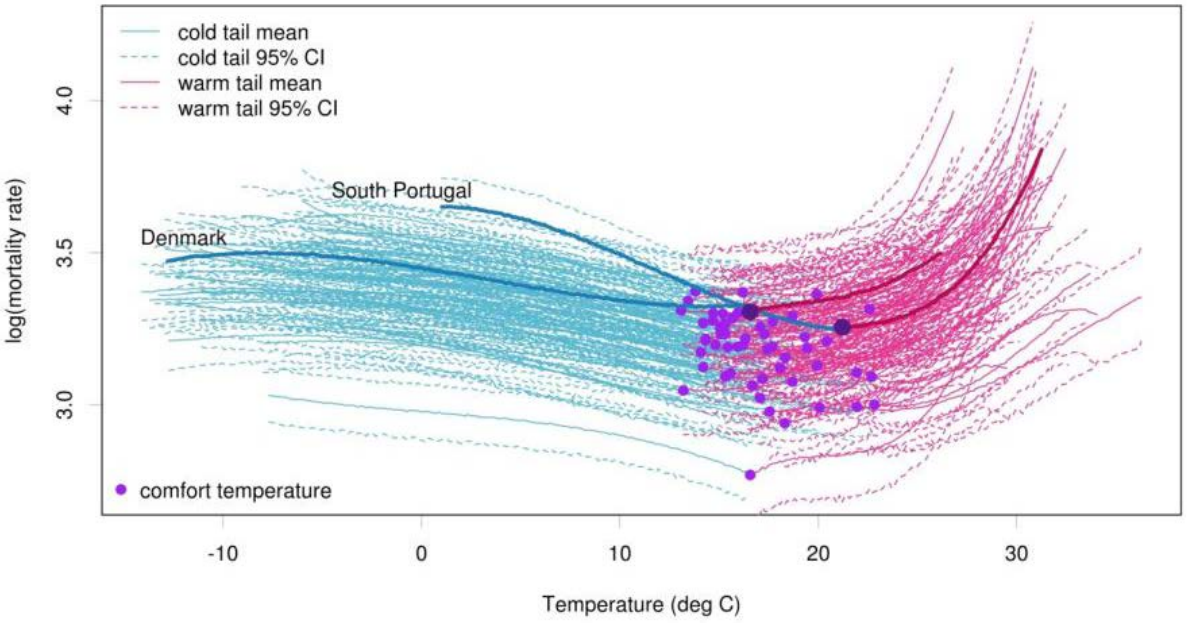

Figure 3.12.1: Posterior predictive distributions (mean and 95% credible intervals) for cold tail (blue) and warm tail (pink) estimations for all 54 regions across Europe. The comfort temperature threshold for each region is marked with a purple dot. The mean mortality curves for two contrasting regions (South Portugal and Denmark) are magnified.

In order to simulate mortality predictions for heat wave or cold spell scenarios, spatio-temporal apparent temperature data, $x_{it,}$ where t is the time step (daily), for heat wave and cold spell scenarios were combined with 1000 samples of the parameters estimated from the warm tail model ($j = w$)

$$y_{it} \sim N(\alpha_w + \beta_{1w} x_{it} + \beta_{2w} x^2{}_{it} + \beta_{3w} x^3{}_{it}, \sigma_w^2),$$ if $x_{it} \geq x_{im}$

and cold tail model ($j = c$)

$$y_{it} \sim N(\alpha_c + \beta_{1c} x_{it} + \beta_{2c} x^2{}_{it} + \beta_{3c} x^3{}_{it}, \sigma_c^2),$$ if $x_{it} < x_{im}$, respectively.

Apparent temperature, defined in the following equation, is the climatological input to the mortality model.

$$T_{app} = -2.653 + 0.994\,T_{air} + 0.0153\,T_{dewpt}^{2}$$

Where T_{app} is the apparent temperature and T_{air} and T_{dewpt} the air and dew point temperatures at 2m, all in degrees Celsius.

ECMWF System4 ensemble forecasts (hindcasts), with 15 ensemble members, were used. Forecasts with one month lead time for each season (DJF, MAM, JJA, and SON) were obtained from The ECOMS User Data Gateway. For this case study, the climate data needs to be combined with regionally aggregated mortality data. Thus, the System4 data has to be aggregated to the 54 regions, described above. For each region, the climate model data at grid points found inside the region are identified and averaged for each time step. In case the region is smaller than the grid squares, the value of the nearest neighbour grid to the centroid of the region is used. Therefore, 54 time series with a daily resolution, for each of the ensemble members and forecasts, are computed. Both series for 2m temperature and 2m dew point temperature are generated, and then combined, following the above equation, to produce regional seasonal forecasts of apparent temperature.

3.12.2. Bias Correction

The regional apparent temperature forecasts were first bias-corrected before using them as an input to the mortality model. The bias correction methodology that we used is as follows. First, the apparent temperature forecasts are aligned in the forecast time and averaged over the run times and ensemble members, in order to compute a forecast-time dependent bias. Along with that, a daily climatology of the reference dataset (ERA-Interim) is computed and smoothed with a LOWESS filter following Mahlstein *et al.* (2015), using alpha=0.4. We inspected the performance of this correction by using quantile-quantile plots, and the results were very good, so it was decided that a more complex correction (such a quantile mapping), was not necessary in this case. Note that using regionally aggregated data reduces the noise and makes the correction more robust.

Therefore, using the forecast apparent temperature data to run the mortality model, 1000 samples of daily mortality rates were generated for each region, each day and each ensemble member (15,000 samples). For comparison, reanalysis apparent temperature (from ERA-Interim) was used to run the mortality model, producing 1000 samples of daily mortality rates for each region and each day. This allows an assessment of the improvement in skill when replacing forecast with observed climate data to run the mortality model. As a case study, daily mortality rate samples were averaged for two climatological events of interest: the heat wave period, 1-15 August 2003 and the cold spell period, 1-15 January 2003, using both forecast and observed climate data as inputs to the mortality model.

3.12.3. Skill Scores and Significance

3.12.3.1. Probability Forecast Maps

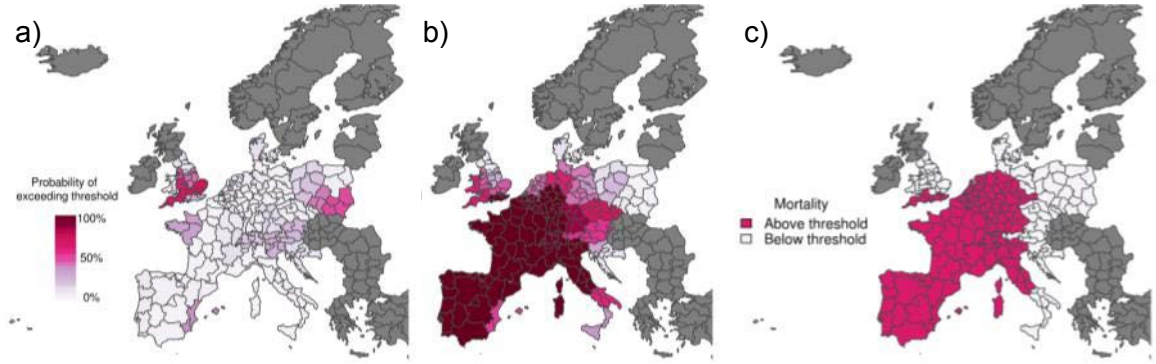

Figure 3.12.2: Probabilistic map of exceeding emergency daily mortality threshold (75th percentile of daily mortality distribution in the warm tail) using (a) ensemble forecast (ECMWF System4) and (b) reanalysis (ERA-Interim) apparent temperature as input to the mortality model. (c) Corresponding observations during a heat wave scenario (1–15 Aug 2003). The graduated colour bar represents the probability of exceeding the mortality threshold (ranging from 0%, pale colours, to 100%, deep colours).

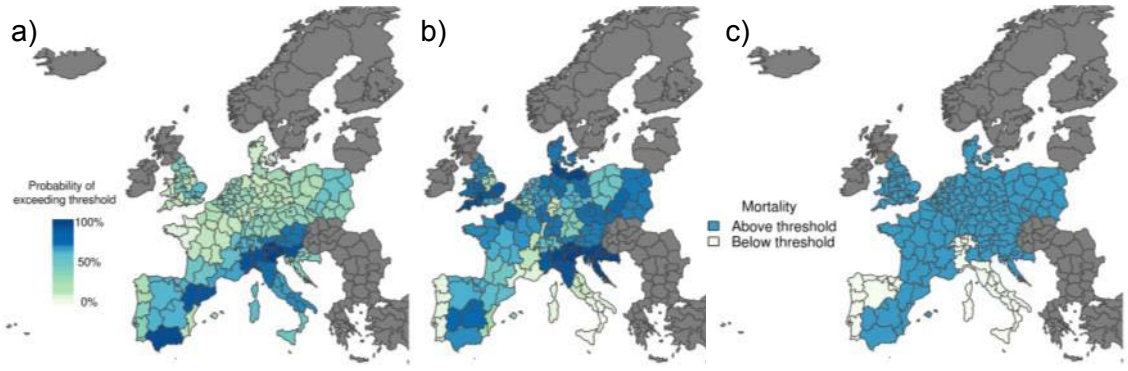

Figure 3.12.3: Probabilistic map of exceeding emergency daily mortality threshold (75th percentile of daily mortality distribution in the cold tail) using (a) ensemble forecast (ECMWF System4) and (b) reanalysis (ERA-Interim) apparent temperature as input to the mortality model. (c) Corresponding observations during a cold wave scenario (1–15 Jan 2003). The graduated colour bar represents the probability of exceeding the mortality threshold (ranging from 0%, pale colours, to 100%, deep colours).

Figure 3.12.2 shows the predicted probability of mortality rates exceeding the 75th percentile of the mortality distribution, given that temperatures are greater than the comfort temperature (i.e. the warm tail distribution) for the heat wave period 1–15 August 2003 using (a) forecast climate (forecast issues in May 2003) and (b) observed (reanalysis) climate. The latter is indicative of a mortality prediction given we had a perfect meteorological forecast (i.e. the observed apparent temperature).The corresponding observations (i.e. whether the mortality rate exceeded the threshold or not) are also displayed (Figure 3.12.2c). Figure 3.12.3 shows the predicted probability of mortality rates exceeding the 75th percentile of the mortality distribution, given that temperatures are colder than the comfort temperature (i.e. the cold tail distribution) for the winter period 1–15 January 2003 using (a) forecast climate

(forecast issues in November 2002) and (b) observed (reanalysis) climate. The corresponding observations (i.e. whether the mortality rate exceeded the threshold or not) are also displayed (Figure 3.12.3c).

3.12.3.2. ROC Score, Hit Rates and False Alarm Rates

Table 3.12.1 shows an evaluation of the probabilistic predictions for the heat wave and cold spell scenarios, using (A) forecast and (B) observed climate as inputs to the mortality model. Probability decision thresholds of 70% and 30% were selected a priori. By using a probability decision "cut off", the efficacy of the model for these specific scenarios was evaluated in a binary framework. To assess the correspondence between mortality predictions and observations, the proportion of correct predictions, and conditional probabilities, i.e. the hit rate and the false alarm rate were calculated. For this exercise, the 'proportion correct' is defined as the proportion of the 54 regions for which the model correctly anticipated that mortality rates would or would not exceed the emergency threshold. The 'hit rate' (sensitivity) is the proportion of regions that correctly predicted that the emergency threshold would be exceeded. Conversely, the false alarm rate (1-specificity) is the proportion of regions for which mortality rates was predicted to exceed the emergency threshold, but did not. The ROC score (equivalent to the area under the modelled ROC curve) characterises the quality of a forecast system by describing the system's ability to anticipate correctly the occurrence or non-occurrence of pre-defined events. A ROC score value of 50% indicates zero skill while a value of 100% represents perfect skill. The results show that the mortality prediction using forecast climate data show no overall skill, with a ROC score for both scenarios less than 50%. However, when we use observed (reanalysis) apparent temperature data to drive the mortality model, the model shows considerable skill, with a ROC score of 97% for the heat wave scenario and 78% for the cold spell scenario. Please note that the latter represents an upper bound to forecast skill of pre-defined events given we had a perfect meteorological forecast (i.e. the observed apparent temperature).

The results indicate that daily disaggregated seasonal climate forecasts, with lead times up to three months, are not sufficient to predict increased mortality for the heat wave and cold spell scenarios examined. Further work will be conducted to test the seasonal climate forecasts in the mortality model at coarser time resolutions (e.g. month, season) and using forecasts with shorter lead times. Due to the short time period of the mortality data (6 years), the skill assessment was performed across space for specific scenarios. The mortality database is currently being updated until 2010. Once this data is available (expected by next year) we will have an increased temporal coverage (15 years) to perform temporal skill analyses for specific seasons across Europe.

Table 3.12.1: Evaluation of heat waves and cold spell scenarios given pre-defined emergency and probability decision thresholds. Mortality model driven by (A) ensemble forecasts (ECMWF System4) of apparent temperature, with a three month lead for the heat wave scenario and a two month lead for the cold spell scenario, and (B) reanalysis (ERA-Interim) apparent temperature.

A

Scenario	**Emergency threshold** Defined for each region using daily data 1998–2003	ROC score	Probability decision threshold	Hit Rate	False Alarm Rate	Proportion correct
Heat wave 1–15 Aug. 2003	75th percentile of mortality distribution given that temperature is warmer than the comfort temperature.	27%	70%	3%	5%	37%
			30%	12%	30%	33%
Cold spell 1–15 Jan. 2003	75th percentile of mortality distribution given that temperature is colder than the comfort temperature.	25%	70%	14%	30%	24%
			30%	80%	100%	65%

B

Scenario	**Emergency threshold** Defined for each region using daily data 1998–2003	ROC score	Probability decision threshold	Hit Rate	False Alarm Rate	Proportion correct
Heat wave 1–15 Aug. 2003	75th percentile of mortality distribution given that temperature is warmer than the comfort temperature.	97%	70%	85%	5%	89%
			30%	100%	55%	80%
Cold spell 1–15 Jan. 2003	75th percentile of mortality distribution given that temperature is colder than the comfort temperature.	78%	70%	66%	20%	69%
			30%	93%	40%	87%

3.12.4. User Requirements

Forecast data and projections for heat are primarily used by health systems within the decision-making framework of heat–health actions plans (HHAP). HHAPs rely on early-warning systems for timely activation and to allow for longer-term resource planning.

Heat–health action plans can be evaluated based on inclusion of nine core elements (Bittner *et al.*, 2014, McGregor *et al.*, 2015, Matthies *et al.*, 2008).

1. Agreement on a lead body and clear definition of actors' responsibilities
2. Accurate and timely alert systems, heat–health watch-warning systems
3. Health information plan
4. Reduction in indoor heat exposure
5. Particular care for vulnerable groups
6. Preparedness of the health/social care system
7. Long-term urban planning
8. Real-time surveillance
9. Monitoring and evaluation

Of specific interest from the perspective of providing a climate service for heat is core element 2, followed by preparedness of the health/social care system (core element 6) and long-term urban planning (core element 7). These three core element components would have different relevant lead-times for forecasting: timely alert systems to trigger the activation of HHAPs are typically days in advance; for countries that have not experienced a severe heat or cold event for several years, an early warning a few months before the event could allow time to update action plans. Longer-term health system preparedness and better resource management would benefit from longer monthly-to-seasonal forecasts to allow for increase of capacity of health services, heat reduction in healthcare facilities and improving health-care networks. Urban planning for heat occurs on seasonal-to-decadal or longer timescales, and includes cross-sectoral issues such as increasing green and blue spaces, changes in building design, changes in land-use decisions, energy consumption reduction, and individual and public transport policies.

Of the 18 countries of the WHO European Region with known heat–health action plans, with varying degree of geographical coverage and measures included, 16 had a clearly defined alert system and a health system preparedness component, all included an information plan, but only four included long-term urban planning within the HHAP itself.

This illustrates a potential for longer forecast lead-times to be incorporated into already existing HHAP in the European Region, whilst the remaining countries could benefit from incorporating such available information at the design stage of new HHAP development. However, these preliminary results indicate that a compromise will have to be reached between user needs and the capabilities of seasonal climate forecasts over Europe, to provide skilful mortality predictions at different temporal resolutions.

Key Points: Temperature-related Mortality

- While episodes of heat and cold related mortality during heat waves and cold spells are predictable given an accurate meteorological forecast, forecast skill with lead times of up to three months is not sufficient to predict increased mortality for the two events studied.
- When addressing impact models for health it may be relevant to consider more than one target forecast (namely, peak timing of excess mortality, maximum incidence at the peak, etc.…) to provide a more comprehensive picture.

Table 3.12.2: Assessment of 18 heat–health action plans in the WHO European Region.

Countries / Indicators	Year	Lead body	Alert system	Information plan	Indoor heat reduction	Vulnerable groups	Health care preparednes	Urban planning	Real-time surveillance	Evaluation
Austria (regional)	2011							**		
Belgium	2005							**		
Croatia	2012									
France	2012							**		
Germany (regional)	2004–2008							**		
Hungary	2007									
Italy	2008									
Luxembourg	2006				**			**		
Moldova	2010									
Monaco	2012							**		
Netherlands	2007							**		
Portugal	2010							**		
Romania	2008									
Serbia	2012							**		
Spain	2012				**			**		
Switzerland (regional)	2007							**		
the FYR Macedonia	2010–2011									
UK (regional)	2012									

** covered elsewhere other than in HHAP. Source: adapted from Bittner et al (2014).

4. Lessons Learnt

Two distinct types of CIIs are identified. CIIs defined with respect to absolute thresholds such as frost days or heating degree days and CIIs defined with respect to relative thresholds such as the percentage of time with wind above the 90th percentile. CIIs based on relative thresholds are beneficial in that they are less prone to systematic model biases. Often, however, user communities have traditionally used CIIs based on absolute thresholds reflective of identified system thresholds in the real world. To derive forecasts of CIIs of the latter type, systematic errors of the seasonal forecasting systems need to be corrected before computing such indices. While such bias correction is important, skill of CII forecasts generally seems to be independent of the choice of bias correction method suggesting that simple de-biasing methods are sufficient. In some cases (e.g. frost days), however, enhanced skill with more sophisticated bias correction methods (e.g. quantile mapping) is found.

The forecast time horizon of long-range forecasts is too long to forecast daily or weekly events deterministically. To account for the inherent uncertainty of such long-range forecasts, these are therefore framed probabilistically. This implies that climate indices (and impact models) have to be computed using all available ensemble members rather than the ensemble mean to be able to represent this uncertainty. The probabilistic framing also implies that calibration methods should be applied in a manner that allows the uncertainty information contained in the ensemble of forecasts to be retained in the bias corrected forecasts. Future research may also focus on incorporating medium-range forecast information to exploit the reduced uncertainty and enhanced forecast skill at shorter lead times.

Seasonal forecast skill varies by climate index or variable, region, season, forecast lead time, and spatio-temporal aggregation. In general, skill of seasonal forecasts is limited in Europe with higher skill during summer than during winter and higher skill for indices related to temperature than for indices related to precipitation. For CIIs derived from a single meteorological quantity (e.g. frost days), forecasts of CIIs are generally at most as skilful as the forecasts of the underlying variable. Even if skill in forecasts of CIIs is not enhanced compared with the underlying meteorological variables, CII forecasts are potentially more relevant and useful to the users as these allow framing the seasonal forecast in a more application-specific way.

Enhanced skill is found for impact variables such as the water requirement satisfaction index or river discharge. Enhanced skill of the latter is due to a combination of the spatio-temporal integration and the additional predictability from initial conditions with long-term memory (e.g. soil moisture and snow in the catchment area). Temporal aggregation has been found to be beneficial in that scores for 3-month periods are generally better than scores for individual months. For some applications, however, information at the monthly time scale may be more useful than seasonally integrated forecasts.

To better understand predictability on seasonal time scales in Europe, the sources of skill should be investigated. Trends due to external forcings are such a source of skill. When assessing a forecast, it is therefore important to distinguish if skill is due to trends or based on the ability of the forecast system to predict interannual variability. Preliminary analysis suggests that long-term trends are an important source of predictability in winter, whereas

predictability in summer seems to be mostly independent of long-term trends. Also, some bias correction methods (e.g. quantile mapping) can affect the representation of trends and thereby forecast skill. The effect of the choice of calibration method on trends and thereby predictability, however, needs further investigation and will be a focus of deliverable D22.3.

5. Links Built

5.1. Links to Other Deliverables, WPs, and Prototypes in EUPORIAS

Wind speed assessments are highly relevant to the RESILIENCE prototype development in WP42, and to the feedback received by end users from WP12. There is also a close link to Vortex, the wind energy partner involved in this WP. The collaboration with Vortex resulted in the organization of a workshop with energy sector stakeholders interested in S2D predictions or potential candidates to use this information in their current activities. The main output of the workshop was the analysis provided by the participants on the S2D climate predictions and its potential impact in their daily activities. It was highlighted that there is a low visibility and demand for probabilistic climate forecasts but all participants agreed in their interest in this type of products. The main recommendation was to facilitate the experimental usage by advanced users in the sector and make publicly available test cases to provide examples of real world usage.

As part of EUPORIAS, MeteoFrance has collaborated with Laurent Dubus (EDF, EUPORIAS partner) concerning the comparison of two hydrological models. Also, MeteoFrance has collaborated with WP23 on impact models for impact predictions for the water sector and with WP33 on how to visualize forecasts of CIIs.

The technical findings from work on the Fire Weather Index within this work package have been developed further into a case study considering in a more holistic sense the value of the CII to a range of decision makers in Europe. Drawing on findings from surveys and interviews conducted as part of WP12, and following the methods developed there through our own online user survey, we gained a much deeper understanding of the potential use of fire indices by decision makers in the forestry, land management, insurance and civil protection sectors. We also considered the current barriers to the use of CIIs and where the value of developing new seasonal forecasts of the fire index would be greatest. Combined with the technical evaluation carried out in this work package and the remaining work under this work package we hope to build a more robust understanding (for both providers and users of the information) of where we can add value to the decision making process through use of CIIs.

The Climate Dynamics and Impacts Unit (UDIC) of IC3 have been liaising with WP33 to help define skill score thresholds, to translate levels of skill into user-friendly categories for stakeholders.

In WP21 the added value of downscaling for provision of CII forecasts will be assessed, a task that will build on the expertise gained in this WP and will use methods and tools developed as part of its deliverables.

5.2. Technical Collaborations

In collaboration with WP21 and building on software developed as part of the FP7 project SPECS, a set of software tools for verification of seasonal forecasts has been developed (https://www.github.com/meteoswiss/easyverification). Access of seasonal forecast data using tools developed as part of WP21 and its verification using the above software has been documented (Bedia and Bhend, 2015). The software is freely available and it is used in a number of analyses performed in the FP7 projects EUPORIAS and SPECS.

In addition, KNMI, together with MeteoSwiss, is working on the extension of the R package climdex.pcic to include in the calculation routines a wide selection of indices. A technical report on the newly added indices is expected by December 2015. The extended version will be used as a validation procedure in the CLIPC project (WP6) to validate the python library ICCLIM, developed within CLIPC for the calculation of indices.

5.3. External Links

The Climate Forecasting Unit (CFU) of IC3 is involved in a number of international projects and a national project called RESILIENCE aligned with the objectives of the EUPORIAS project either because their aim is to improve seasonal-to-decadal (S2D) predictions or because they are centred in the energy sector needs on climate services. The team is a partner in the PREFACE European project and in two European projects within the Horizon 2020 programme: PRIMAVERA and IMPREX. PREFACE aims to improve the understanding and capabilities to predict tropical Atlantic climate and its impacts; PRIMAVERA aims to deliver novel, advanced and well-evaluated high-resolution global climate models tailored to and actionable for sector-specific end-users such as the energy sector; and IMPREX aims to improve the prediction and management of meteorological and hydrological extremes which might have an impact on energy facilities, energy production and management. IC3 also participates in an ERA-NET project called NEWA for the preparation of the New European Wind Atlas that will include seasonal and sub-seasonal wind predictions.

IC3-CFU has also collaborated in the CLIM4ENERGY proposal that aims at responding to the Copernicus Climate Change Service (C3S) objectives, by demonstrating the added value of tailored climate information for the transitioning European energy sector. IC3-CFU (under its new name as Barcelona Supercomputing Centre) together with MeteoSwiss collaborated in the Copernicus C3S proposal QA4Seas on the evaluation and quality control for seasonal forecasting systems available through the Copernicus C3S.

Lessons learned from this work package have fed into design and post-processing of sub-seasonal to seasonal forecasts at MeteoSwiss, the Swiss national weather service. Also, findings from WP22 feed into the Project HEPS4Power (Extended-range Hydrometeorological Ensemble Predictions for Improved Hydropower Operations and Revenues), funded by the Swiss Science Foundation through the Swiss Competence Centre for Energy Research. As part of this project, stakeholder interaction with exponents from the energy sector will be intensified. In turn, insights on presentation formats of long-range forecasts gained through stakeholder interaction and public dissemination of such forecasts have been made available in EUPORIAS.

Also, MeteoSwiss will collaborate with FU Berlin and MPI Hamburg as part of MiKlip II - Module E “Evaluation of the MiKlip Decadal Prediction System”. Thereby insights on the calibration and recalibration of seasonal forecasting systems developed as part of this work package will feed into the development of the German decadal prediction system MiKlip.

The Climate Dynamics and Impacts Unit of IC3 (IC3-UDIC) is involved in a number of international projects, consortium, outreach, and user-engagement activities. In collaboration with the FP7 projects DENFREE and SPECS, IC3-UDIC led the development of a prototype climate service for health by issuing a dengue early warning forecast, driven by seasonal climate forecasts, three months ahead of the 2014 World Cup in Brazil. Dengue early warning predictions were disseminated widely, for example as part of the European Centre for Disease Control (ECDC) health risk assessment, reported by the UK National Health Service (NHS) and published by more than 18 international press outlets, including the BBC. The operational use of seasonal climate forecasts in routine dengue early warnings is now a priority for the Brazilian Climate and Health Observatory, in collaboration with the Brazilian Space Agency (INPE).

IC3-UDIC is collaborating with the developers of the UrbClim model in the framework of the NACLIM project to represent the Urban Heat Island effect on health impacts. This collaboration has led to the preparation of a Copernicus Climate Change Services (CS3) tender proposal regarding a Multi-Sectoral Information System for Urban Areas (UrbanSIS). IC3 has also collaborated in Copernicus Climate Change Services (CS3) tender proposal (SECTEUR) that aims to identify user requirements and gaps related to the use of climate information to support decision making.

In addition, IC3 has organized a two-week training activity entitled “Modelling tools and capacity building for climate and public health” at the Oswaldo Cruz Foundation Itaboraí Palace in Petrópolis, Rio de Janeiro, Brazil. Also contacts exist to the Public Health Agency of Barcelona and the Ministry of Health in Catalonia to implement an alert system for mortality-derived casualties associated to heatwaves.

6. References

Ballester, J., Robine, J. M., Herrmann, F. R., & Rodó, X. (2011). Long-term projections and acclimatization scenarios of temperature-related mortality in Europe. Nature Communications, 2, 358.

Barnston A., Tippet MK., van den Dool HM, Unger DA (2015): Toward an Improved Multimodel ENSO Prediction. *J. Appl Meteo. Clim.*, 54, 1579-1595

Bedia, J., Bhend, J., 2015. A bias correction framework in EUPORIAS WP22: description and examples. Internal EUPORIAS Tech. Rep., v1. URL: http://meteo.unican.es/work/downscaler/wiki/linked_docs/20150722_biasCorrection/examples WP22_bhj.pdf

Bedia, J., Herrera, S., Camia, A., Moreno, J.M., Gutierrez, J.M., 2014a. Forest Fire Danger Projections in the Mediterranean using ENSEMBLES Regional Climate Change Scenarios. Clim Chang 122, 185–199. doi:10.1007/s10584-013-1005-z

Bedia, J., Herrera, S., Guti'errez, J.M., Benali, A., Brands, S., Mota, B., Moreno, J.M., 2015. Global patterns in the sensitivity of burned area to fire-weather: implications for climate change. *Agr. Forest Meteorol.* 214–215:369–379. doi:10.1016/j.agrformet.2015.09.002

Bedia, J., Herrera, S., Gutiérrez, J.M., 2014b. Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain. *Nat Hazards Earth Syst Sci* 14, 53–66. doi:10.5194/nhess-14-53-2014

Bedia, J., Herrera, S., Gutiérrez, J.M., Zavala, G., Urbieta, I.R., Moreno, J.M., 2012. Sensitivity of fire weather index to different reanalysis products in the Iberian Peninsula. *Nat. Hazards Earth Syst. Sci.* 12, 699–708. doi:10.5194/nhess-12-699-2012

Bedia, J., Herrera, S., San-Martín, D., Koutsias, N., Gutiérrez, J.M., 2013. Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. *Clim. Change* 120, 229–247. doi:10.1007/s10584-013-0787-3

Bhend, J., 2015. easyVerification: Forecast verification metrics for large datasets. R package version 0.1.5.3

Bittner MI *et al.* Are European countries prepared for the next big heat-wave? *European Journal of Public Health*, 2014, 24(4):615–619.

Blanco-Ward D, Queijeiro JMG, Jones GV (2007) Spatial climate variability and viticulture in the Miño River Valley of Spain. *Vitis* 46: 63−70

Branas J, Bernon G, Levadoux L (1946) *Eléments de viticulture générale*. Imp. Dehan, Montpellier

Casanueva, A., Frías, M.D., Herrera, S., San-Martín, D., Zaninovic, K., Gutiérrez, J.M., 2014. Statistical downscaling of climate impact indices: testing the direct approach. *Climatic Change* 127, 547–560. doi:10.1007/s10584-014-1270-5

Céron, J. P., Tanguy, G., Franchistéguy, L., Martin, E., Regimbeau, F., & Vidal, J. P. (2010). Hydrological seasonal forecast over France: feasibility and prospects. *Atmospheric Science Letters*, 11(2), 78-82.

Chen, Y., Randerson, J.T., Morton, D.C., DeFries, R.S., Collatz, G.J., Kasibhatla, P.S., Giglio, L., Jin, Y., Marlier, M.E., 2011. Forecasting Fire Season Severity in South America Using Sea Surface Temperature Anomalies. *Science* 334, 787–791. doi:10.1126/science.1209472

Chu, P., Yan, W., Fugioka, F., 2002. Fire-climate relationships and long-lead seasonal wildfire prediction for Hawaii. *Int. J. Wildland Fire* 11, 25–31.

de Groot, W.J., Goldammer, J.G., Keenan, T., Brady, M.A., Lynham, T.J., Justice, C.O., Csiszar, I.A., O'Loughlin, K., 2006. Developing a global early warning system for wildland fire. *Forest Ecology and Management* 234, S10.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hóllm EV, Isaksen L, Kallberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteor. Soc.*137(656):553–597, doi: 10.1002/qj.82.

Diez, E., Orfila, B., Frias, M.D., Fernandez, J., Cofino, A.S., Gutierrez, J.M., 2011. Downscaling ECMWF seasonal precipitation forecasts in Europe using the RCA model. *Tellus A* 63, 757-762. doi:10.1111/j.1600-0870.2011.00523.x

Ferro, C. A. T. (2014). Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1917-1923.

Ferro, C. A. T., Richardson, D. S., & Weigel, A. P. (2008). On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15(1), 19–24. http://doi.org/10.1002/met.45

Gangstø, R., A. P. Weigel, M. A. Liniger, and C. Appenzeller (2013). Methodological aspects of the validation of decadal predictions. *Climatic Research*, 55(3), 181-200.

Gudmundsson, L., Rego, F.C., Rocha, M., Seneviratne, S.I., 2014. Predicting above normal wildfire activity in southern Europe as a function of meteorological drought. *Envir. Res. Lett.* 9, 084008. doi:10.1088/1748-9326/9/8/084008

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

Herrera, S., Bedia, J., Gutierrez, J.M., Fernandez, J., Moreno, J.M., 2013. On the projection of future fire danger conditions with various instantaneous/mean-daily data sources. *Clim. Change* 118, 827–840. doi:10.1007/s10584-012-0667-2

Hiederer, R., 2013a: *Mapping Soil Properties for Europe – Spatial Representation of Soil Database Attributes.* EUR26082EN Scientific and Technical Research series. Publications Office of the European Union, Luxembourg. DOI:10.2788/94128

Hiederer, R., 2013b: *Mapping Soil Typologies – Spatial Decision Support Applied to European Soil Database.* EUR25932EN Scientific and Technical Research series. Publications Office of the European Union, Luxembourg. 127 pp. DOI:543 10.2788/87286

Hopson, T. M. (2014). Assessing the ensemble spread-error relationship. *Monthly Weather Review*, 142(3), 1125-1142.

Huglin P. (1978) Nouveau mode d'évaluation des possibilities héliothermiques d'un milieu viticole. *CR Acad Agr* 64: 1117–1126

Jackson D (2001) *Climate: monographs in cool climate viticulture–2*. Daphne Brasell Associates, Wellington

Jones GV, Davis RE (2000) Using a synoptic climatological approach to understand climate-viticulture relationships. *Int J Clim* 20 (8):813-837

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, March, 1996.

Kirtman, B., Pirani, A. (2009). The state of the art of seasonal prediction: Outcomes and recommendations from the First World Climate Research Program Workshop on Seasonal Prediction. *Bulletin of the American Meteorological Society*, 90(4), 455-458.

Klein Tank, A.M.G, Zwiers, F.W., Zhang, X. (2009). Guidelines on analysis of analysis of extremes in a changing climate in support of informed decisions for adaptation. *Climate data and Monitoring*. WCDMP-No72

Kumar, A. (2009). Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, 137(8), 2622-2631.

Lienert, F., Fyfe, J. C., & Merryfield, W. J. (2011). Do Climate Models Capture the Tropical Influences on North Pacific Sea Surface Temperature Variability? *Journal of Climate*, 24(23), 6203-6209.

Lowe, R., Ballester, J., Creswick, J., Robine, J. M., Herrmann, F. R., & Rodó, X. (2015). Evaluating the Performance of a Climate-Driven Mortality Model during Heat Waves and Cold Spells in Europe. *International Journal of Environmental Research and Public Health*, 12(2), 1279-1294.

Magalhães, N. (2008) - *Tratado de Viticultura* - A Videira, A Vinha e o "Terroir". 1ª. edição. Espanha, Chaves Ferreira – Publicações, S.A..

Mahlstein, I., C. Spirig, M. A. Liniger, and C. Appenzeller (2015). Estimating daily climatologies for climate indices derived from climate model data and observations. *Journal of Geophysical Research: Atmospheres*, 120(7), 2808-2818.

Malheiro AC, Santos JA, Fraga H, Pinto JG (2010): Climate change scenarios applied to viticultural zoning in Europe. *Clim Res* 43: 163−177

Marcos, R., Turco, M., Bedia, J., Llasat, M.C., Provenzale, A., 2015. Seasonal predictability of summer fires in a Mediterranean environment. *International Journal of Wildland Fire*. doi:10.1071/WF15079

Matthies F *et al. WHO Europe, Heat-health action plans* (Matthies F *et al.*, eds.). Copenhagen, Denmark, WHO Regional Office for Europe, 2008 (http://www.euro.who.int/__data/assets/pdf_file/0006/95919/E91347.pdf).

McGregor GR *et al.*eds. *Heatwaves and Health: Guidance on Warning-System Development*. Geneva, Switzerland, World Meteorological Organization, 2015.
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.
Moriondo, M., Good, P., Durao, R., Bindi, M., Giannakopoulos, C., Corte-Real, J., 2006. Potential impact of climate change on fire risk in the Mediterranean area. *Climate Research* 31, 85–95.
Palmer, W.C., 1965. *Meteorological drought.* Research Paper No. 45. U.S. Weather Bureau. NOAA Library and Information Services Division, Washington, D.C. 20852.
Panofsky, H. A., and G. W. Brier (1968). *Some Applications of Statistics to Meteorology.* The Pennsylvania State University, University Park, PA, USA.
Pepler, A. S., L. B. Díaz, C. Prodhomme, F. J. Doblas-Reyes, and A. Kumar (2015). The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. *Weather and Climate Extremes*, 9, 68-77.
Photiadou, C.S., Van den Hurk, B., Van Delden, A., Weerts, A. (2015). Incorporating circulation statistics in bias correction of GCM ensembles: hydrological application for the Rhine basin. *Climate Dynamics*. doi:10.1007/s00382-015-2578-1
Preisler, H.K., Westerling, A.L., 2007. Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States. *Journal of Applied Meteorology and Climatology* 46, 1020–1030. doi:10.1175/JAM2513.1
Prodhomme, C., Doblas-Reyes, F. J., Bellprat, O., Dutra, E. (2015) Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Climate Dynamics.*
R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.
Roads, J., Fugioka, F., Chen, S., Burgan, R., 2005. Seasonal fire danger forecasts for the USA. *Int. J. Wildland Fire* 14, 1–18.
San-Miguel-Ayanz, J., Moreno, J.M., Camia, A., 2013. Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management* 294, 11–22. doi:10.1016/j.foreco.2012.10.050
Santander Meteorology Group 2015. *ecomsUDG.Raccess: R interface to the ECOMS User Data Gateway.* R package version 4.0-0. https://meteo.unican.es/trac/wiki/udg/ecoms
Santander Meteorology Group, 2015b. *downscaleR: Climate data manipulation and statistical downscaling.* R package version 0.8-3. https://github.com/SantanderMetGroup/downscaleR/wiki
Santos, JA; Malheiro AC; Pinto JG; Jones GA (2012): Macroclimate and viticultural zoning in Europe: observed trends and atmospheric forcing. *Clim Res* 51: 89–103, doi: 10.3354/cr01056
Siegert, S., 2014. *SpecsVerification: Forecast verification routines for the SPECS FP7 project.* R package version 0.3-0. http://CRAN.R-project.org/package=SpecsVerification
Singla, S., Céron, J. P., Martin, E., Regimbeau, F., Déqué, M., Habets, F., & Vidal, J. P. (2012). Predictability of soil moisture and river flows over France for the spring season. *Hydrology and Earth System Sciences*, 16(1), 201-216.
Stocks, B.J., Lawson, B.D., Alexander, M.E., Van Wagner, C.E., McAlpine, R.S., Lynham, T.J., Dube, D.E., 1989. The Canadian Forest Fire Danger Rating System - An Overview. *For. Chron.* 65, 450–457.
Themeßl, M., A. Gobiet, and G. Heinrich (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2), 449-468.
Tonietto J, Carbonneau A (2004) A multicriteria climatic classification system for grape-growing regions worldwide. *Agric Meteorol* 124:81–97
Turco, M., Llasat, M.C., von Hardenberg, J., Provenzale, A., 2013. Impact of climate variability on summer fires in a Mediterranean environment (northeastern Iberian Peninsula). *Clim. Change* 116, 665–678. doi:10.1007/s10584-012-0505-6
van Wagner, C.E., 1987. *Development and structure of the Canadian Forest Fire Weather Index* (Forestry Tech. Rep. No. 35). Canadian Forestry Service, Ottawa, Canada.
Venäläinen, A., Korhonen, N., Hyvärinen, O., Koutsias, N., Xystrakis, F., Urbieta, I.R., Moreno, J.M., 2014. Temporal variations and change in forest fire danger in Europe for 1960–2012. *Natural Hazards and Earth System Science* 14, 1477–1490. doi:10.5194/nhess-14-1477-2014
Viegas, D.X., Bovio, G., Ferreira, A., Nosenzo, A., Sol, B., 1999. Comparative study of various methods of fire danger evaluation in southern Europe. *Int. J. Wildland Fire* 9, 235–246. doi:10.1071/WF00015

Weedon, G.P., Gomes, S., Viterbo, P., Shuttleworth, W.J., Blyth, E., Österle, H., Adam, J.C., Bellouin, N., Boucher, O., Best, M., 2011. Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. *Journal of Hydrometeorology,* 12, 823–848. doi:10.1175/2011JHM1369.1

Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J. and Viterbo, P., 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50, doi:10.1002/2014WR015638

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2009). Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? *Monthly Weather Review*, 137(4), 1460-1479.

Weigel, A. P. (2012), Ensemble forecasts, in *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed., chap. 8, edited by D. B. Stephenson and I. T. Jolliffe, pp. 141–166, Wiley-Blackwell, Oxford, U. K.

Wells, N., Goddard, S., Hayes, M., 2004. A Self-Calibrating Palmer Drought Severity Index, *J. Clim.*, 17, 2335-2351. DOI: 10.1175/1520-0442(2004)017<2335:ASPDSI>2.0.CO;2

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

Winkler AJ, Cook JA, Kliwer WM, Lider LA (1974) *General viticulture*. University of California Press, Berkeley, CA

Wotton, B.M., 2009. Interpreting and using outputs from the Canadian Forest Fire Danger Rating System in research applications. Environmental and Ecological Statistics 16, 107–131. doi:10.1007/s10651-007-0084-2