# Skill Assessment of Seasonal Temperature and Precipitation Forecast over Europe

In partial fulfillment of the requirements for the degree of
Master in Data Science (2015-16)
at Barcelona Graduate School of Economics (BGSE)

and in joint collaboration with
the Earth Sciences Department
at Barcelona Supercomputing Center (BSC)

By Niti Mishra

This research leading to these results has received funding from the Horizon 2020 EU program under grant agreements 641811- IMproving PRedictions and management of hydrological EXtremes (IMPREX)


www.imprex.eu

*I am specially grateful to my supervisor, Chloè Prodhomme, who supervised me throughout the three months and until the very end, encouraging me to continue to improve. My thanks also to Eleftheria Exarchou and Christian Brownless for reviewing my thesis and provding insightful comments, which has allowed me to complete this master's thesis.*

# Contents

# 1 Introduction

## 1.1 Need for Seasonal Climate Forecasting

Forecast or prediction is an estimate of future event. Forecasts have to be verified by comparing them with observations. There are forecasts for economic indicators such as unemployment and inflation as well as forecasts for weather such as daily temperature. There are also forecasts for climate, which are made on seasonal, annual or decadal time scale. Forecasts are useful for effective decision making as they help minimize risks associated with uncertainty. For example; extreme climate events, such as floods and heatwaves can be destructive to humans, animals and environment. Hence, climate forecasting is an essential component of managing risks associated with extreme climate events.

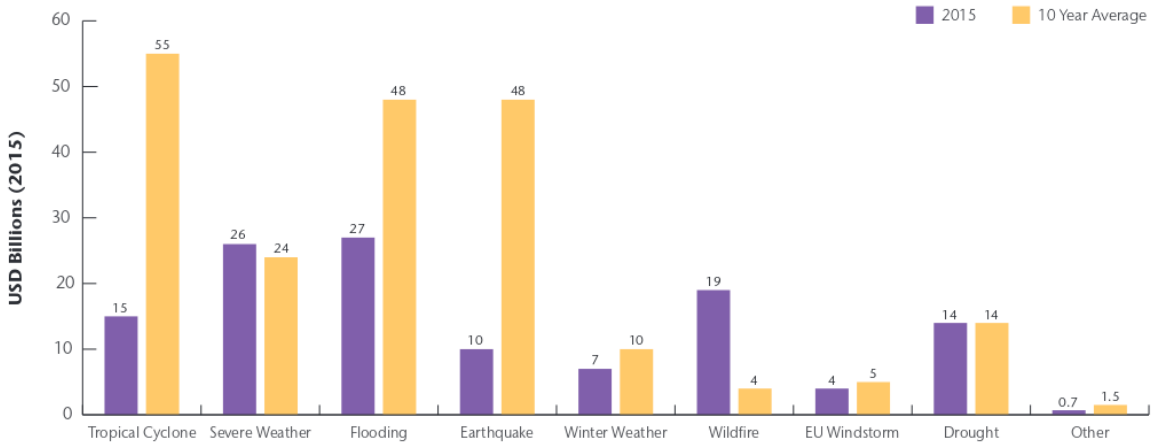


Figure 1: Global Economic Losses by Peril, adapted from http://thoughtleadership.aonbenfield.com

Understanding the costs of climate variability provides a context in which the use of climate forecast can be valued. In 2015, the global economic loss by perils exceeded EUR 109 billion (Fig 1). Recent hydrological extreme events demonstrate the vulnerability of European society to water-related natural hazards and there is strong evidence that climate change will worsen these events [1]. The IMproving PRedictions and management of hydrological EXtremes (IMPREX) project is designed to support the reduction of Europe's vulnerability to extreme hydrological events through improved understanding of the intensity and frequency of future disrupting features. IMPREX invests in improving current state-of-the-art forecasting systems and the development of new forecasting tools. It also focuses on customizing climate information to stakeholders' needs. This thesis takes place in the current framework of IMPREX project.

## 1.2 Seasonal Climate Forecast

Seasonal climate forecasts (SCFs) are forecasts of climatic events at timescales of few weeks up to a few months. It is different from weather forecasting that are made a few days into the future. It also differs from climate change which focuses on predicting changes in climate in very long time scale such as century. Figure 2 illustrates various weather and climate scales and the sources of predictability for each timescale. This thesis is focused on the predictability of climatic event on a seasonal scale which falls in the category of climate variability. What is the probability of having warmer or colder temperature during a season at a certain place given past observed temperatures? SCF aims to estimate such probabilities. It provides the range of values which is most likely to occur during the next season. It is important to bear in mind that it is not possible to predict the daily weather variations at a specific location months in advance due to the complex and stochastic nature of the atmospheric circulation. However, it is possible to predict anomalies in long-term climatic conditions to a certain extent.

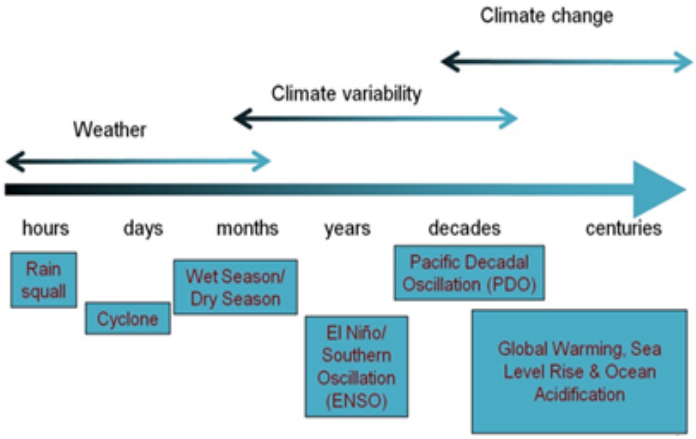


Figure 2: Weather and Climate Scales, adapted from: http://www.pacificclimatefutures.net

SCFs are currently under utilized for prevention, adaptation and prediction by the public and economic sectors. In Europe, the use of SCF is limited to particular sectors such as energy, water, insurance, and transport [2]. This is partly because the existing skill and reliability of SCFs in Europe is low and varies considerably depending on the geographical area, the time of the year and the climate variable [2]. This is also because climate forecasts in general are poorly understood due to the lack of communication between providers and end-users of SCFs [3].

General Circulation Models (GCMs) are the dynamical models that employ mathematical equations and simulate the Earth's atmosphere, oceans or both (known as coupled GCMs). Dynamical models have seasonal forecast skills in regions with strong connection to ENSO [4]. However, the predictions from GCMs are penalized due to model biases and uncertainties. Uncertainties arise due to the limited information

regarding the state of the ocean, sea-ice, snow, land and the physical processes of their interactions [5] and also due to the inability to perfectly model the dynamic climate system [6]. Uncertainties in the initial state have impact on model skill and the growth of forecast errors. For example, an initial value problem is predicting an ENSO state that affects the predictions of later states. Figure 3 illustrates ensemble generation and modelling that incorporates some of these uncertainties, which has made longer time scale prediction more plausible [7]. In this thesis, I use ensemble forecasts. I provide further detail on ensemble in following sections.

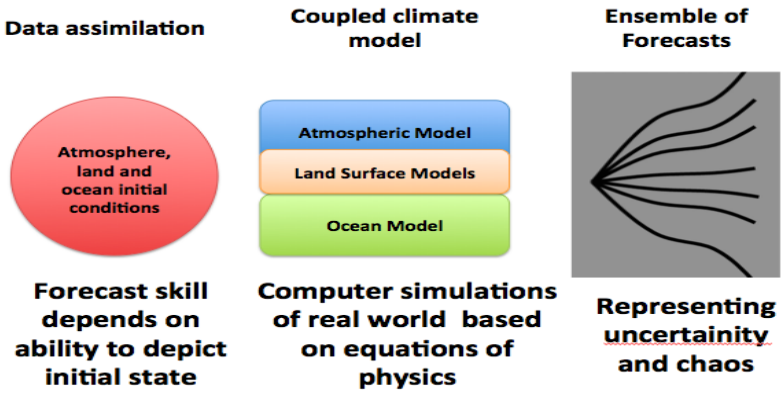


Figure 3: Process of ensemble generation from General Climate Model (GCM)

## 1.3 Values of Seasonal Climate Forecast

Advanced information in the form of SCFs can improve decision-making, increase preparedness and lead to better socio-economic and environmental outcomes. In some parts of the world such as Europe, SCFs may have economic values. In some other parts of the world, such as areas extremely vulnerable to climate hazards, it may be a matter of life and death. Inclusion of SCFs in the detection, monitoring and early warning of climate hazards can increase the warning time and information to effectively minimize associated risks [8].

The benefits of SCFs are particularly high in areas with high climate variability [9] [8]. For example; in the northern part of the Australian grain belt, wheat is grown in an extremely variable climate [9]. Decision Support Systems (DSS) can quantify risks associated with various decision options prior to planting by integrating knowledge of long-term climate records, development pattern of plantations and response to various fertilizers. The payoff to using fertilizer depends on the chance of getting a high-yielding season [9]. However, the range of likely yield is often wide as it depends on the extent of climate variability. Hence, the eventual outcome is strongly dependent on the uncertain nature of that season and its interaction with the decisions made at planting time. This provides further compelling reason to understand, monitor and

predict climate variability.

Research programs such as Climate Change, Agriculture and Food Security (CCAFS) are designed to work with farming communities to build *climate-smart villages* that aim to achieve food security and broader development goals under a changing climate and increasing food demand. In Yatenga region of North Burkina, farmers have to cope with erratic rainfall patterns, varying greatly every year. Overall, majority of farmers exposed to climate information showed willingness to pay for seasonal forecasts indicating benefits of SCFs [10]. It led to higher yield on average for the production of cowpea while also saving in seed and pesticides costs. However, similar economic benefits were not realized for sesame production. This indicates the importance of assessing the capacity to understand and interpret forecasts for specific contexts in order to optimize the potential of good years and minimize the losses during poorer years.

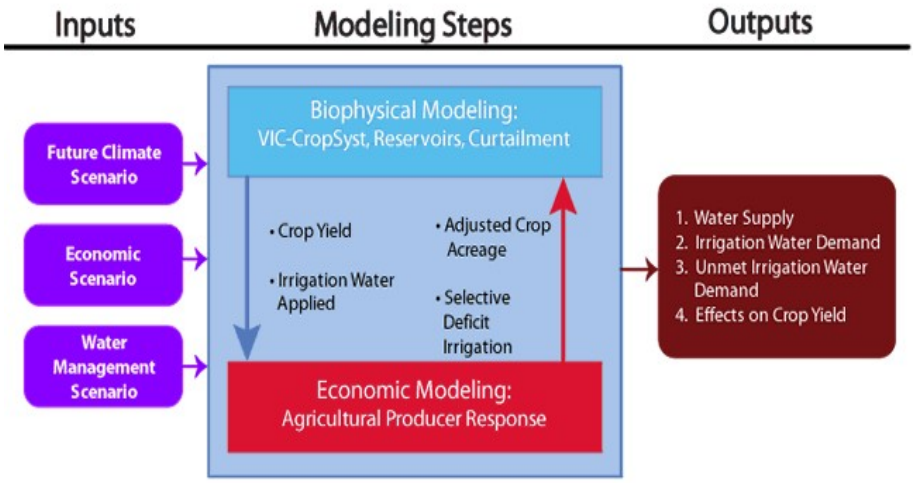


Figure 4: An example of hydroeconomic-model that integrates biophysical modeling with economic modeling, adapted from: http://www.ecy.wa.gov

Appropriate integration of seasonal precipitation and temperature forecasts can provide an opportunity for developing a proactive approach towards water management. Singla et al. [11] studied the importance of wet land conditions for spring predictability of the hydrological system over France and found that the predictive skill varied among regions based on seasonal climate and elevation. For example; significant improvement of river flow was observed in the north-east of France but not in Mediterranean area. Similarly, Kahil et al. [12] presented a hydro-economic model for sustainable water management in an arid and semiarid basin in Southeastern Spain. Figure 4 illustrates their model, which estimates the distribution of available water among users under each anticipated climatic scenarios. Their model takes into account interaction between supply nodes such as rivers, reservoirs, aquifers and demand nodes such as irrigation districts, households and aquatic ecosystems. It is then incorporated into an economic model that searches for optimal behavior of water-use under a set of technical and resource constraints. They concluded that

having a water policy that accounts for detailed regional-economic component is optimal for the society[12].

Furthermore, increasing climate variability has increased the demand of weather-related insurance, bringing the financial instrument of weather derivatives to the fore [13]. The weather derivative market ties contract values to location-specific weather outcomes in order to hedge potential temperature risks[14]. Stern presented an approach to the pricing of weather derivatives that employed a combination of empirical data including forecast verification data, regional synoptic classification data and data associated with climate indices on a global scale such as the Southern Oscillation Index[14]. Carriquiry and Osgood [15] also emphasized the value of Index-based weather insurance that encourage farmers to take advantage of more profitable options when climate risks are lower and to take more protective measures when the risks are higher.

All these examples above illustrate the values of SCFs across a range of application sectors to manage risks and prepare better for the future. Recent developments in climate research has led to increased interest in SCFs within academia, research institutes and meteorological forecast services. Much of these developments are related to the identification of sources of predictability of SCFs, which is discussed in the next section.

## 1.4 Sources of Predictability

The enormous value of seasonal forecasts calls for a comprehensive understanding of the sources of predictability for both seasonal temperature and precipitation. For example, the anthropocentric emissions of greenhouse gases (GHGs) are leading to a global climatic warming, illustrated in Figure 5. Coupled dynamic models are able to capture this evolution. Therefore, this strong warming trend is a source of predictability on a seasonal timescale, especially over Europe [16].

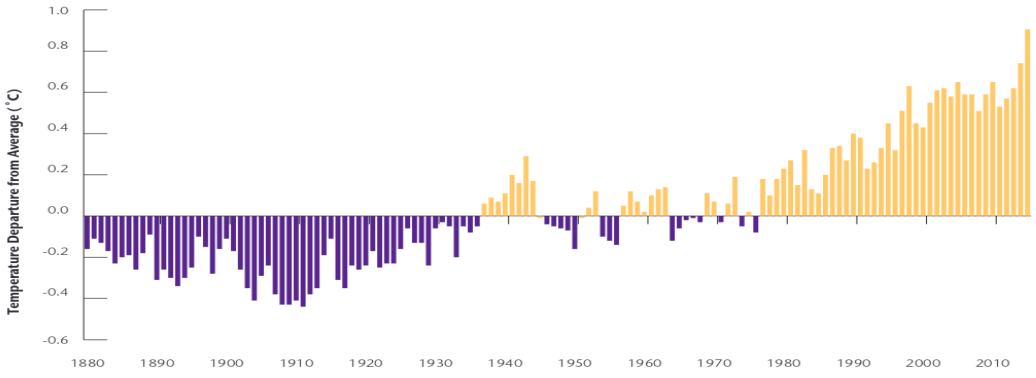


Figure 5: Global Land and Ocean Temperature Anamolies, adapted from: http://thoughtleadership.aonbenfield.com

El Niño Southern Oscillation (ENSO) is the most important mode of climatic variability. ENSO refers

to the periodical variation in wind and SST in the tropical eastern Pacific Ocean that causes large spatial shifts in global atmospheric and precipitation patterns [17]. The positive phase of ENSO, known as El Niño, is associated with the warm SST anomalies that cause an increase in heat flux from ocean to atmosphere. This brings about changes in the structure of rainfall and in the release of latent heat in the troposphere. The negative phase of ENSO, known as La Niña, is characterized by the opposite extreme of the ENSO cycle. ENSO events last approximately 1 year and the time between the two phases varies between 2 and 7 years [4].

Accurate prediction of ENSO is linked to improved prediction of climate pattern over tropics and subtropics [18]. Variations associated with ENSO have also led to increased skill in predicting extreme temperatures [8]. Other tropical ocean basins also have major impact on climate variability of surrounding regions. For instance, SST anomalies over the tropical Atlantic affect the precipitation pattern over north-eastern Brazil and western Africa [19] [17]. Interestingly, SST variability of the Atlantic and Indian basins are linked to that of the tropical Pacific suggesting interdependence between ocean basins and resulting SST anomalies [19] [20]. Thus, SCF is mainly possible due to the predictability of the oceanic circulation and by the fact that variability in tropical sea surface temperatures (SST) have a significant global impact on the atmospheric circulation [4]. Large SST anomalies can alter the atmospheric boundary layer which can change the structure of rainfall and the intensity of latent heat in the troposphere [8]. This further impacts the atmospheric circulation leading to climatic anomalies.

Besides variations in SST, atmospheric pressure, snow cover [21], soil moisture [22], and sea-ice [23] are also proven to be effective sources of predictability. North Atlantic Oscillation (NAO) is an index of normalized pressure difference between Iceland and the Azores. NAO exhibits large scale interannual variability which are considered to cause shifts in seasonal climate in the nearby continents [24]. It is of particular interest to the western Europe and eastern United States as it has been linked to the winter climate variability around the Atlantic basin [25] [22]. Recently, Scaife et al. (2014) have suggested that NAO could be predicted by the UK MetOffice's GLosea5 system forecast [25].

The insulating and reflecting properties of the snow along with its role in the hydrological cycle has also been useful in climate prediction [4]. Recent observational and model studies have noted local effects of snow-cover on surface air temperatures and on large-scale circulation patterns. For example; the colder surface over Eurasia during the extensive snow periods was found to influence the planetary-scale wave and the subsequent climate in the Euroasian landmass [26].

Similarly, the preconditioning of extreme summer temperatures by preceding precipitation suggests soil-moisture information to be essential for correctly predicting summer temperatures over land [27]. For example, dry condition in the soil moisture preceding the heat wave allowed more energy for sensible heating and thus created suitable condition for increase of near surface temperature, leading to the occurrence of

2010 heat wave over Russia (Fig 6) [22]. Although, initial soil moisture data has led to better temperature forecasts in some regions, the impacts of land-surface initialization on predicting variability of seasonal climate over the European region is considered to be relatively low [22].

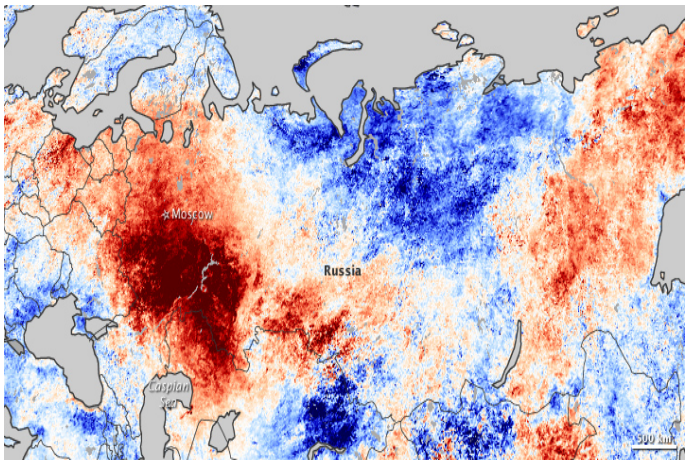


Figure 6: Temperature anomalies based on land surface temperatures observed on NASA's Terra satellite for the Russian Federation from July 20–27, 2010, compared to temperatures for the same dates from 2000 to 2008. Areas with above-average temperatures (red), areas with below-average temperatures (blue) and oceans and lakes (gray) are shown, adapted from: http://earthobservatory.nasa.gov

Researches have also shown improvements in forecast skill over land and extra-tropical oceans due to a better representation of anthropogenic green house gases (GHGs) forcing and land-use changes [28] [29]. Large scale deforestation during the industrial era in Euroasia and North America for agricultural cropping and grazing lands has been associated with climate variations [30]. The largest effect of deforestation is estimated to be at high latitudes because the snow-covered open grounds reflect more sunlight than snow on trees [31].

Thus, complex interdependent nature of the climate system makes seasonal climate prediction inherently difficult. However, it is possible. As discussed above, various sources of predictability have been identified to improve climate prediction on seasonal time scale, leading also to an increase in its demand. As we rely more on SCFs to minimize risks, assessing the quality of these forecasts becomes a topic of utmost importance. This process, known as forecast verification, is essential to any scientific forecasting system. If we do not verify SCFs then we have no way of learning whether these forecasts are indeed better than the best guess we make or the best model we are using. Hence, forecast verification is essential because it provides users with best available information needed for improved decision-making, increased preparedness, leading to better socio-economic and environmental outcomes. In the next section, I briefly discuss the context of forecast verification within climate science along with further description on ensemble forecast, which is the type of forecast I use in this thesis.

## 1.5 Forecast Verification in the context of Climate Prediction

In 1884 J.P. Finley published a paper in *American Meteorological Journal*, where he reported "percentage of verification" exceeding 95% for an experimental tornado forecasting program (Fig 7) [32]. The index of

performance was defined as percent of correct tornado over no-tornado forecasts, $2708/2803 = 96.6\%$. This index was criticized by many as inappropriate measure of performance in this context because it is possible to do even better by always forecasting "No Tornado" and get $2752/2803 = 98.2\%$ forecast success rate [33].

| | Observations | | |
|---|---|---|---|
| Forecasts | Tornado | No tornado | |
| Tornado | 28 | 72 | 100 |
| No tornado | 23 | 2680 | 2703 |
| | 51 | 2752 | 2803 |

Figure 7: Finley's Tornado Forecast in terms of 2 x 2 contingency table, adapted from Murphy, 1996

Allan Murphy, who is considered a pioneer in the field of forecast verification, distinguished three types of "goodness" of a forecast: (1) Consistency (2) Quality and (3) Value [34]. As seen in the case of Finley's tornado forecasts, Murphy argued that reducing vast amount of information from a set of forecasts and observations into a single verification measure can misinterpret verification results [35]. He also noted that value of forecast is often more important than skill alone emphasizing the need for user-oriented approach in forecast verification [34]. He suggested instead a diagnostic verification approach based on the joint distribution of the forecasts and observations [33]. Given a sufficiently large data set, the joint distribution can be interpreted as an empirical relative frequency distribution. Diagnostic verification on joint distribution gives information about the nature of forecast errors as well as clues as to the sources of the errors [7], which makes this approach particularly appealing. In this report, I also take distribution oriented diagnostic verification approach for the assessment of SCF skill over Europe.

In the last decade, research and development of new verification strategies and reassessment of traditional forecast verification methods have received a great deal of attention from the scientific community [7]. Verification practices vary between different national services. The World Meteorological Organization (WMO) provides a Standard Verification System for Long-Range Forecasts that is intended to facilitate the exchange of comparable verification scores between different centres [33]. Besides there is also a constant need to adapt practices as forecasts, data and users continue to change [33]. For example, the advances in spatial forecasting has complicated the verification process as access to a wider range of satellite imagery has led to redefinition of cyclones [33]. Hence, apparent trends in cyclone frequency could be due to changes of definition rather than to genuine climatic trends.

Another example is the use of ensembles, illustrated in Figure 8, which were infeasible 30 years ago but are now widespread [7]. The seasonal temperature and precipitation forecasts used in this report also consists of ensemble forecasts. Ensemble is a collection of several independent forecasts generated by first, sampling (usually via Monte Carlo techniques) slightly different initial conditions using observations and then, calculating the evolution of these states using dynamical models. Rather than integrating single best guess of the initial condition, this method is more consistent with the true stochastic nature of the atmospheric flow [33]. Wide variation in ensemble forecasts suggest a lot of uncertainty while low ensemble variance provide more confidence in predicting a particular event [7]. *Good* ensemble behave as representatives of

random draws from the "true model state" and provide an estimate of the forecast uncertainty [33]. Thus, if the initial condition from the ensemble truly characterize the "true" initial condition uncertainty, and if the atmospheric dynamics are well represented by the model, then the forecast ensemble can accurately represent the true atmospheric state. However, in practice, neither are the models perfect nor is it easily possible to construct the true representative state of initial conditions. Thus, specific issues continue to arise and need for further research and development in forecast verification remains.

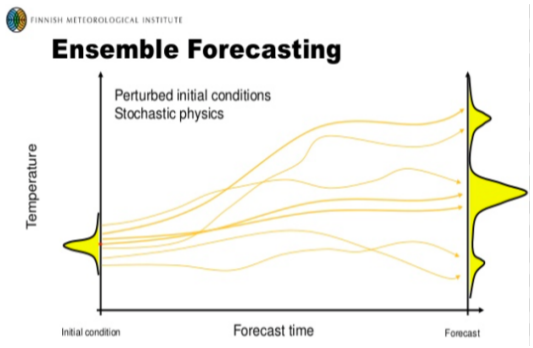


Figure 8: Ensemble generation and modelling, adapted from: http://en.ilmatieteenlaitos.fi

Much of the discussion made thus far is aimed towards drawing the reader's attention to (1) the need of SCFs, (2) the sources of predictability and the associated complexity of climate prediction and (3) the subsequent difficulty to verify SCFs in a comprehensive manner. In the following sections, I will focus on the skill assessment of EUROSIP's ensemble forecasts for seasonal precipitation and temperature over Europe. In particular, I assess the forecast skill for each of the four forecasting models provided by EUROSIP and compare how they perform throughout the European region. I perform this assessment for winter and summer season for both temperature and precipitation. Then, I proceed to assess the forecast skill of multi-model ensemble system, which is forecasts based on average of the four individual models. Multi-models often outperform the best single forecast system [36]. Therefore, I aim to assess whether forecast skill can be gained in regions with low skill through the multi-model approach. In section 2, I provide a detailed description of the ensemble forecast data and the methodologies used for forecast verification. In Section 3, I present the subsequent results of each forecast verification methods followed by a final conclusion in Section 4.

# 2 Methodology

## 2.1 Data

### Observation

The observation dataset for temperature are obtained from ERA-Interim (ERAint) database. ERAint consists of a reanalysis of global atmosphere since 1979 and continues in real time [37]. Initial condition for ERAint atmospheric model is obtained through the 2006 release of cycle 31r2 of ECMWF's Integrated Forecasting System (IFS). Additional information on data assimilation system and modelling can be found in the IFS Cy31r1 documentation [38]. ERAint system includes a 4-dimensional variational analysis (4D-Var) with a 12-hour analysis window. The spatial resolution of the data set is $\approx$ 80 km (T255 spectral) on a reduced Gaussian grid on 60 vertical levels from the surface up to 0.1 hPa [38].

The observation dataset for precipitation is obtained from Global Precipitation Climatology Project (GPCP). GPCP provides a combined observation-only dataset, which is a gridded analysis based on gauge measurements and satellite estimates of precipitation. GPCP has promoted the development of an analysis procedure for blending various estimates together to produce the necessary global gridded precipitation fields [39]. Further details on their operational procedure is documented in GPCP Version 2.2 Combined Precipitation Data Set covering the period January 1979 through the present [39]. The values are represented in a grid of 2.5x2.5 latitude–longitude (Cylindrical Equal Distance) global array of points.

### Forecasts

| | ***Ensemble Forecast System*** | ***No. of Members*** |
|---|---|---|
| 1 | Global seasonal forecasting system (Glosea5, Met Office) | 24 |
| 2 | European Centre for Medium Range Weather Forecasts (System4, ECMWF) | 51 |
| 3 | National Centers for Environmental Prediction (System2, NCEP) | 24 |
| 4 | Meteo France (System4, MF) | 15 |

Table 1: List of Coupled Seasonal Forecasting System integrated in EUROSIP's database

Table 1 lists the forecast systems from which the data was obtained for the analysis of forecast skill assessment. These forecast systems include a comprehensive set of seasonal forecasts of temperature and precipitation over Europe. They are provided by the EUROSIP database that consists of these four independent coupled seasonal forecasting systems integrated into a common framework [37].

### Location and Time Period

The data set covers temperature and precipitation values for location specified as $20^o$ W to $70^o$ E and $25^o$ N to $75^o$ N, which covers the European region of our interest. The final grid is of dimension 71x128. The forecast grids from four models were interpolated to matched the observation grid. Verification at a single grid represents an evaluation of forecasts across time at that grid location. The data are loaded from the storage, interpolated and the sea point are masked using the *Load* function of the *s2dverification* (version 2.6.0.) https://cran.r-project.org/web/packages/s2dverification/ package in R.

The data is obtained on a monthly time scale i.e. the values on each grid point refers to the monthly averages of daily means. Seasonal data for winter and summer is obtained for the years between 1992 to 2012. Winter data consists of months June, July and August (JJA) with $1^{st}$ of May as initialization date. Summer data consists of months December, January and February (DJF) with $1^{st}$ of November as the initialization date. The initialization dates are also known as starting dates from which the model starts the forecasts. The lead time, which refers to the number of months the experiments are set to be run, is 3.

### Tools

The final format of the data obtained is a list of 4 R objects; (1) array of observations (1x1x21x3x71x128), (2) array of ensemble forecasts (4x51x21x3x71x128), (3) vector of longitudes (128x1) and (4) vector of latitudes (1x71). The names of the dimensions of the arrays are listed in Table 2 below:

| | ***Dimension Name*** | ***Dimension Size*** |
|---|---|---|
| 1 | Experimental/observation datasets | 4 / 1 |
| 2 | Ensemble members | max 51 (*varies per model*) |
| 3 | Start dates | 21 |
| 4 | Lead times | 3 |
| 5 | Longitudes | 128 |
| 6 | Latitudes | 71 |

Table 2: List of the names of 6-dimensional array of the data obtained for analysis in this report

Both forecast and observation data are imported from Barcelona Supercomputing Centre (BSC) system database using in-house R package, *s2dverification* (version 2.6.0.). The package is developed in collaboration between BSC and Institut Català de Ciències del Clima (IC3). All of the verification assessments are done using functions available in *s2dverification*, *easyVerification* (version 0.6.0) https://cran.r-project.org/web/packages/easyVerification/index.html, and *specsVerification* (version 0.4.1) https://cran.r-project.org/web/packages/SpecsVerification/index.html in R. These packages provide a set of tools

to efficiently perform forecast verification analyses on data stored as an array in R. Readers are referred to the github repository: https://github.com/nitimkc/Seasonal_Forecast_Verification for more information on the specific use of these packages along with relevant code scripts.

## 2.2 Methods of Forecast Verification

There are numerous ways of assessing forecast quality and different types of forecasts call for different verification methods. Therefore, it is important to identify verification goal in order to choose correct verification statistics, measures and graphics that match the type of forecast and the attribute of interest. Table 3 below distinguishes forecasts along with verification methods that are appropriate for that type of forecast. Items in red circle represent the type of forecasts used in this report for verification analyses.

| **Nature of Forecast** | **Examples** | **Verification Methods** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | *Visual* | *Dichotomous* | *Multi-category* | *Continuous* | *Probabilisitc* | *Ensemble* | *Spatial* |
| Deterministic (Non-probabilistic) | Quantitative precipitation forecast | X | X | X | X | | | X |
| Probabilistic | Probability of precipitation, ensemble forecast | X | | | | X | X | |
| Qualitative | 5-day outlook | X | X | X | | | | |
| **Space-Time Domain** | | | | | | | | |
| Time Series | Daily maximum temperature for a city | X | X | X | X | X | | |
| Spatial Distribution | Map of geopotential height, rainfall chart | X | X | X | X | X | X | X |
| Pooled Space and Time | Monthly average global temperature anomaly | | X | X | X | X | X | |
| **Specificity of Forecast** | | | | | | | | |
| Dishotomous (yes/no) | Occurrence of fog | X | X | | | X | X | X |
| Multi-category | Cold, normal, or warm conditions | X | | X | | X | X | X |
| Continous | Maximum temperature | X | | | X | X | X | X |
| Object- or Event-oriented | Tropical cyclone motion and intensity | X | X | X | X | X | | X |

Table 3: Appropriate verification methods for different forecast types. Items in circle represent the type of forecasts used in this report, adapted from: http://www.cawcr.gov.au/projects/verification/

As the goal of this report is to assess the skill of ensemble forecasts for seasonal temperature and precipitation that are continuous in nature, I use corresponding verification methods for their skill assessment. In particular, I take a diagnostic verification approach in my analysis and look at the joint relationship between the forecasts of each model and the observation. Diagnostic verification approach is favorable for skill assessment because it takes into account the full distribution of both forecasts and observations. This is useful as various aspects of forecast quality can be explored by factoring the joint probability of observation and forecast into conditional and marginal distributions [33].

Based on the joint relationship between forecast and observation, a scoring rule can assign numerical scores to the performance of a forecasting model in numerous ways. Depending on its construction, scoring rule can be positively or negatively oriented, meaning either high or low score may be preferred. All distribution oriented scoring rules can be decomposed to explore not only the nature of forecast errors but also

gain clues as to the sources of the errors [7], which helps improve forecast models. A scoring rule. $s(p, x)$, is said to be *proper* if given a forecaster issues the predictive distribution $p$ and observation $x$ is realized from distribution $q$, the expected score is maximized when $p = q$. The scoring rule is *strictly proper*, if the expected score is uniquely maximized [40].

Another appealing aspect of scoring rule is that it allows a meaningful comparison between two forecast models. Indeed, forecasts are often judged in relative terms meaning they are compared to forecasts from another model to determine their *skill score*. Skill score is a statistical approach of evaluating the accuracy of a forecast over a reference forecast. It is defined as:

$$Skill\ \ score = \frac{Score - S_{ref}}{S_{perf} - S_{ref}} \tag{1}$$

where $S_{ref}$ is the score of reference forecast and $S_{pref}$ is the score of perfect forecast. It is an index that takes the value 1 for perfect forecast skill and 0 for same skill as the reference forecast. A negative value implies worse skill than the reference model [33]. There are usually a whole range of possible skill scores that measures the relative quality of different forecasts. Therefore, it is necessary to define a baseline against which a forecast can be judged [33]. Two common baseline forecasts used in climate science are (1) Persistence and (2) Climatology. Persistence is forecasting whatever is observed at the present time as the forecast to persist into the next period. This strategy is often successful for short-range forecasts. Climatology refers to the average conditions over some recent reference period. Climatology is defined as:

$$\widehat{X}_{CLIM} = E(X) = \frac{1}{N}\sum_{t=1}^{n} x_t$$

where $N$ is the total number of observations over time and $x_t$ is value of the observations at a given time. In this report, climatology is used as the reference forecast for skill comparison as it is the standard reference for long-range forecasts.

For the first part of the forecast verification analysis, Pearson's correlation coefficient measure is used to obtain temporal correlation between forecasts and observations. Then, the ensemble forecasts are converted into probabilistic forecasts and the Continuous Probability Ranked Skill Score (CRPSS) is computed for each of the models. However, CRPSS has a few drawbacks which can be partly corrected by Fair Continuous Probability Ranked Skill Score (FCRPSS), which is the final skill score used for analysis in this report. The same methodology is repeated for multi-model forecasts that is obtained by averaging the four independent forecast systems. In this report, correlation and CRPSS are considered because these assessment methods are also used within the hydrologists' community of IMPREX to asses the usefulness of seasonal forecasts used in their hydrological model. Such hydrological models require complete distribution of the seasonal forecasts.

Hence, instead of converting the forecasts into rain or no-rain type of binary or categorical forecasts, we use the whole ensemble and its probability distribution for the skill assessment.

### 2.2.1 Correlation

Correlation coefficient is the *association* attribute of forecast quality. It measures the linear association between two variables. It is a measure that is invariant to shifts in the mean and is not affected by data transformation. Due to this invariance property, correlation is widely used in weather and climate forecasting [33].

I use the Pearson's correlation coefficient measure to assess the linear temporal consistency between the forecasts and the observations for the 21 years of 1992-2012. The correlation coefficient is calculated for each grid by using the formula below:

$$\rho = cor(X, \hat{X}) = \frac{cov(X, \hat{X})}{\sqrt{var(X), var(\hat{X})}} \tag{2}$$

where

$$cov(X, \hat{X}) = \frac{1}{N} \sum_{t=1}^{n} (x_t - \bar{x_t})(\hat{x}_t - \bar{\hat{x}_t})$$

is the covariance between the observations and forecasts which is estimated from past samples, $t = 1, ..., N$. $var(X)$ and $var(\widehat{X})$ refer to variances of observations and forecasts, respectively.

A parametric approach is taken to test the significance of correlation for each grid point. The test is to reject the null hypothesis of no-skill based on assumption that the forecasts and observations are independent and normally distributed. Under this null hypothesis, the sample test statistic is distributed as one-sided Student t-distribution with $n-2$ degrees of freedom. Given this test statistics, prediction interval for no-skill correlation is obtained at 95% confidence level i.e., at 5% level of significance.

While correlation is one way to measure forecast skill of a prediction system. It is sensitive to outliers and does not take into consideration the full distribution of the forecast. Averaging the ensemble members to obtain one-point forecast leads to loss in data, which can otherwise provide useful insight of the prediction quality. Hence, to account for the full distribution of the ensemble forecasts, CRPSS and FCRPSS quality measures are also computed. These methodologies are detailed in the following sections.

### 2.2.2 Continuous Ranked Probability Skill Score (CRPSS)

Since the temperature and precipitation forecasts take values on a continuous scale and come in a form of ensemble forecast, Continuous Ranked Probability Score (CRPS) is an appropriate skill metric for the assessment of these forecasts [41]. To compute CRPS, ensemble forecasts are first converted into probabilistic forecasts through a probability distribution. A *probability distribution* is a rule or a function that assigns a specific probability to all the possible values of an ensemble forecast.

CRPS takes the full distribution obtained from the ensemble members and compares it with that of the observation. For a specific value observed, the corresponding cumulative distribution function (CDF) is a step-wise function (also known as *Heaviside*), where the transition from step 0 to 1 occurs at the step corresponding to the observed value. Each ensemble member forecasts lie on a continuous range of the steps. The CRPS is the total area between the CDF of the forecast and the CDF of the observation (Fig 9). The general expression of CRPS of a forecast is given by:

$$CRPS = \frac{1}{N}\int_{-\infty}^{\infty}\left(F_t^f(x) - F_t^o(x)\right)^2 dx$$

where $F_t^f(x)$ and $F_t^o(x)$ are CDFs for $t^{th}$ forecast and observation, respectively.

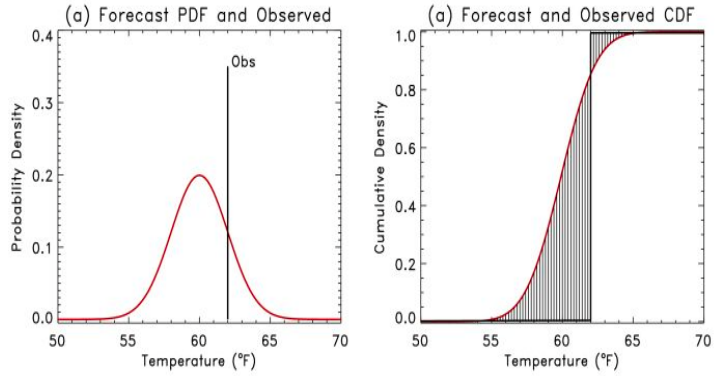


Figure 9: (a) The probability distribution function (pdf) of ensemble member forecasts (red) and of observation (black). (b) The cumulative distribution function (CDF) of ensemble member forecasts (red) and of observation (black). CRPS is the total area between the two CDFs, adapted from: http://www.eumetcal.org/

To calculate CRPS, I take the set of $m = 51$ ensemble members $\{1, 2, ...M\}$, each with $n = 21$ ensemble forecasts $\{1, 2, ...N\}$ from 1992 to 2012. Let $\widehat{x}_{t,(i)}$ denote the $i^{th}$ ensemble member of the $t^{th}$ forecast with $m$ ensemble members being sorted in ascending order (bracket around subscript index $i$ indicate order statistics). Further, let $\widehat{x}_{t,(0)}$ = -$\infty$ and $\widehat{x}_{t,(m+1)} = \infty$ (for easier interpretation of continuous scale). Then, given equal probability assigned to each of the ensemble members, the CRPS for each grid point is computed as:

$$CRPS = \frac{1}{n}\sum_{t=1}^{n}\left[\sum_{i=1}^{m}\alpha_{t,i}\left(\frac{i}{m}\right)^2 + \sum_{i=0}^{m-1}\beta_{t,i}\left(1-\frac{i}{m}\right)^2\right] \quad (3)$$

where steps are defined as;

$$\alpha_{t,i} = \begin{cases} 0 & \text{if} \quad x_t \leqslant \widehat{x}_{t,(i)} \\ x_t - \widehat{x}_{t,(i)} & \text{if} \quad \widehat{x}_{t,(i)} \leqslant x_t \leqslant \widehat{x}_{t,(i+1)} \\ \widehat{x}_{t,(i+1)} - \widehat{x}_{t,(i)} & \text{if} \quad \widehat{x}_{t,(i+1)} \leqslant x_t \end{cases}$$

and

$$\beta_{t,i} = \begin{cases} \widehat{x}_{t,(i+1)} - \widehat{x}_{t,(i)} & \text{if} \quad x_t \leqslant \widehat{x}_{t,(i)} \\ \widehat{x}_{t,(i+1)} - x_t & \text{if} \quad \widehat{x}_{t,(i)} \leqslant x_t \leqslant \widehat{x}_{t,(i+1)} \\ 0 & \text{if} \quad \widehat{x}_{t,(i+1)} \leqslant x_t \end{cases}$$

Although CRPS does not require reduction of the ensemble forecasts to discrete probabilities of categorical events [41], it is computed discretely because observations and forecast distributions are recorded on discrete intervals. Thus, in this case CRPS is equal to the Mean Absolute Error (MAE) and has a clear interpretation. CRPS has a negative orientation as it rewards concentration of probability around the step function located at the observed value i.e. the intersection point of the two CDFs (Fig 9). This is because such probability distribution of forecasts will have smaller integrated squared distance from the observation. Finally, once the CRPS is computed, the skill score of CRPSS is computed using the standard skill score formula (Equation 1):

$$CRPSS = \frac{CRPS_f - CRPS_{clim}}{CRPS_{perf} - CRPS_{clim}}$$

where $CRPS_f$, $CRPS_{clim}$ and $CRPS_{perf}$ stand for CRPS of forecast of interest, CRPS of reference climatology and CRPS of perfect forecast, respectively. The range is -$\infty$ to 1. Note that perfect CRPS score is 0 due to its negative orientation.

Despite its advantage of applicability on continuous ensemble forecasts, CRPSS also has drawbacks. While it measures the reliability attribute of a set of ensembles, it does not award individual ensemble. Recall the description of ensemble forecasts from Section 1.3. Ensembles are intended to be simple random samples of the same distribution as the observation. Thus, a scoring rule that favors ensembles that behave as if they were drawn from the same distribution as the observation is desirable [42]. However, since CRPS

verifies the ensemble probability forecast, random sample ensembles may not actually appear optimal based on CRPS [43]. This suggests that the CRPS is sensitive to the average ensemble spread including the frequency and magnitude of the outliers [41]. In this regard, Fricker et al. [44] introduced the concept of Fair CRPS that rewards ensemble with members that behave as if they were randomly drawn from the same distribution as the observation. In the next section, I discuss the Fair CRPS further.

### 2.2.3 Fair Continuous Ranked Probability Skilled Score (FCRPS)

A scoring rule is not *fair* if the score can be increased by forecasting what is not the forecaster's true belief. For example; as in the case of Finely's tornado forecast (Section 1.3). For probability forecasts, a proper scoring rule is fair when it is interpreted as the forecaster's belief [44]. So, if the expectation of the scoring rule $s(p, x)$, is taken with respect to any true probability distribution, $q$, for the verifying observation, $x$, the score is optimized when $p = q$. Thus, if $q$ represents a forecaster's true belief about $x$, then the forecaster cannot do better (on average) by issuing a forecast that is other than $q$ [40]. For ensemble forecasts, it is more complicated. Fair scoring rule (FSR) for ensemble is one that elicits random samples of forecasts. However, no such scoring rules exists [44] because expectation of a scoring rule, $s(\hat{\boldsymbol{x}}, x)$, with respect to any distribution, $q$, for the verifying observation, $x$, is a deterministic function of ensemble, $\hat{\boldsymbol{x}}$. This means optimizing values can always be determined and issued as the forecast instead of issuing a random sample [43].

This raises a question about how we should verify an ensemble? Ferro [43] suggests that ensemble should be verified according to its intended use. So, if probability distribution based on the empirical distribution of the forecast is used then we must verify the empirical distribution with a proper score, such as CRPS. If we want to use ensemble mean as the forecast, then we should verify squared error based on the ensemble mean. To verify whether the ensemble are random samples of the same distribution as the observation, Fricker et al. [44] suggests a FSR such that the expected value of the score is over all possible ensembles.

Given that an ensemble, $\hat{\boldsymbol{x}}$, is a random sample from a probability distribution, $p$, the expectation of the scoring rule, $s(\hat{\boldsymbol{x}}, x)$, with respect to **both** $p$ and any distribution, $q$, for the verifying observation, $x$, is optimized when $p = q$. The scoring rule is *strictly* fair if its expectation is *uniquely* optimized when $p = q$ (Fricker et al. 2013). Thus, choosing $p$ such that it issues random samples from a *support-subset* of $\widehat{X}$ will optimize the expectation with respect to both $p$ and $q$. If the *support-subset* of $q$ is a subset of $\widehat{X}$, then any distribution, $p$, with support equal to that of $q$ will optimize the expected score [43]. In other words, the expected value of the score is optimized over all possible ensembles.

To see this applied in our data, where the $m$ ensemble members and the observation $x$, can take any value on a real line, let $p$ denote the probability density function (PDF) for the ensemble distribution and $q$ denote the PDF for the distribution of the observation, $x$. Furthermore, let $x_t = \mathbb{1}(x \leq t)$ and $\hat{\boldsymbol{x}}_t = (\hat{x}_{t,1}, ..., \hat{x}_{t,m})$,

where $\hat{x}_{t,i} = \mathbb{1}(\hat{x}_i \leq t)$, represent the step-function verifying the observation and ensemble forecast based on the outcome of the event $\{x \leq t\}$ for a threshold, $t$. For each $t$, let the scoring rule, $s_t(\hat{\boldsymbol{x}}_t, x_t)$, be fair, negatively oriented and also bounded for all $\hat{\boldsymbol{x}}$, $x_t$ and $t$. Then, FSR can be obtained as:

$$FCRPS = s(\hat{\boldsymbol{x}}, x) = \int_{-\infty}^{\infty} \left\{ \frac{i(t)}{m} - \frac{j(t)}{n} \right\}^2 dt - \frac{\sum_{i \neq j} |x_i - x_j|}{2m^2(m-1)}$$

where $i(t)$ members and $j(t)$ are verifications that predict the event $\{y \leq t\}$.

Hence, FCRPS effectively evaluates the underlying ensemble distribution and not just the empirical distribution as in the case of CRPS. FCRPSS is computed using the same skill score formula (Equation 1):

$$FCRPSS = \frac{FCRPS_f - FCRPS_{clim}}{FCRPS_{perf} - FCRPS_{clim}}$$

where $FCRPS_f$, $FCRPS_{clim}$ and $FCRPS_{perf}$ stand for FCRPS of forecast of interest, FCRPS of reference climatology and FCRPS of perfect forecast, respectively. It is interpreted similar to CRPSS.

# 3 Results

## 3.1 Correlation

The standard procedure of forecast verification for temperature and rainfall is to compute correlation between forecast and the observations obtained for the same time period of 1992 - 2012. It is used to quantify the maximum skill that can be obtained in a particular region given a forecast system. For the evaluation of this technique, temporal correlation was computed for the European region specified as $20^o$ W to $70^o$ E and $25^o$ N to $75^o$ N over the 21 years. Seasonal temperature and precipitation forecast for each model was obtained by averaging over all ensemble members and over three months, JJA (initialized on 1st of May) and DJF (initialized on 1st of November). Multi-model forecast was obtained by averaging the forecasts of the four models. Correlation was computed on these seasonal datasets for each grid point. Figure 10, 11 and 12 (following pages) present the result of correlation analysis. The correlation maps can be interpreted as:

- Areas covered in red are indicative of positive relationship and suggest better skill compared to climatology.
- Areas covered in blue indicate worse skill than climatology.
- Based on one-tailed Student t-distribution test, forecasts for both temperature and precipitation are considered statistically significant if they lie in the prediction interval of 95% confidence. These areas are represented by a dot on each grid point.
- The data over sea is masked, which are the white areas on the maps.

### Correlation Skill for Winter Season

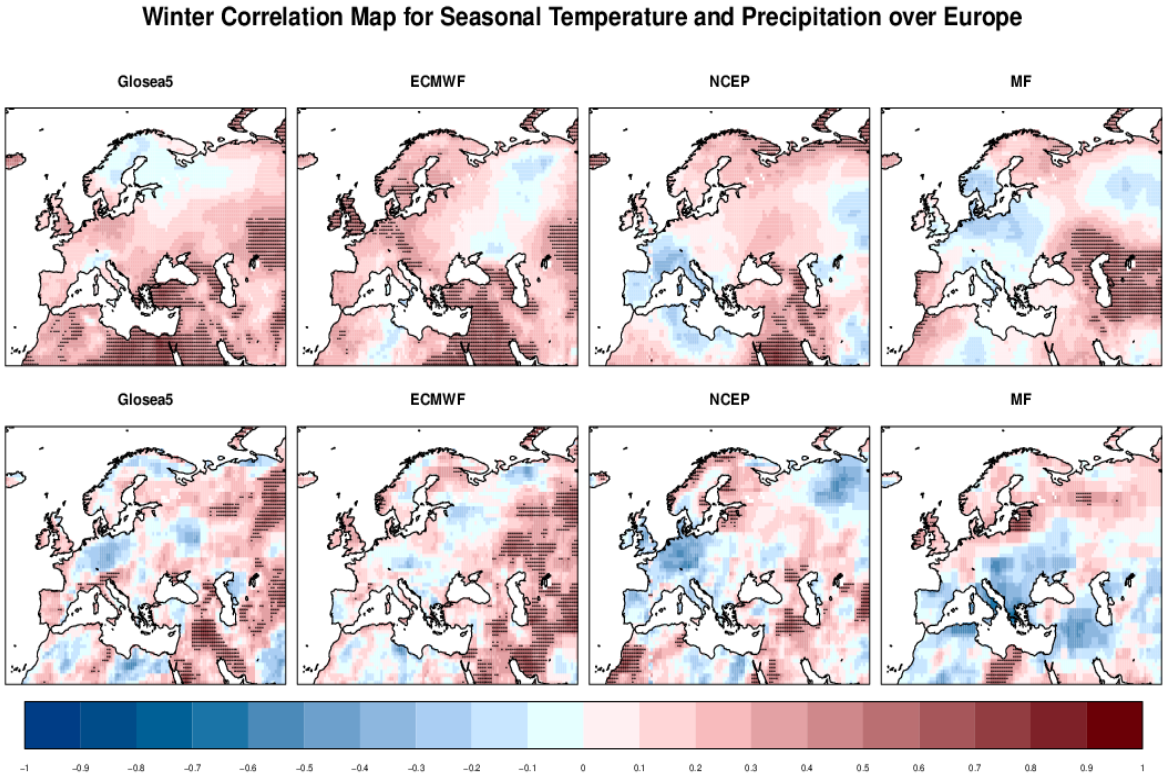


Figure 10: Seasonal (DJF) temporal correlation for European region between 1992-2012. The four maps on the top and the bottom correspond to seasonal temperature and precipitation, respectively. Red indicates positive correlation and blue indicates negative correlation. Dots mark the areas where the skill is significant at 95% confidence level.

Based on the maps in Figure 10, the forecast skill for both seasonal temperature and precipitation for winter is limited in Europe. The skill is the South-Eastern region mostly. The seasonal precipitation skill is much lower and sporadic compared to seasonal temperature. This is because given its high variability, precipitation is hard to observe and to forecast. The skill is mostly in Glosea5 and ECMWF.

It is interesting to note that the area where the models are skillful do not overlap very much. This suggests that the mechanisms leading to the skill in forecast might vary from one model to another. The overall lack of significance in the correlation in the western region of Europe could be because of the low predictability of NAO. However, caution must be exercised in the interpretation of these results as the seasonal data is only based on the years 1992 - 2012.

### Correlation Skill for Summer Season

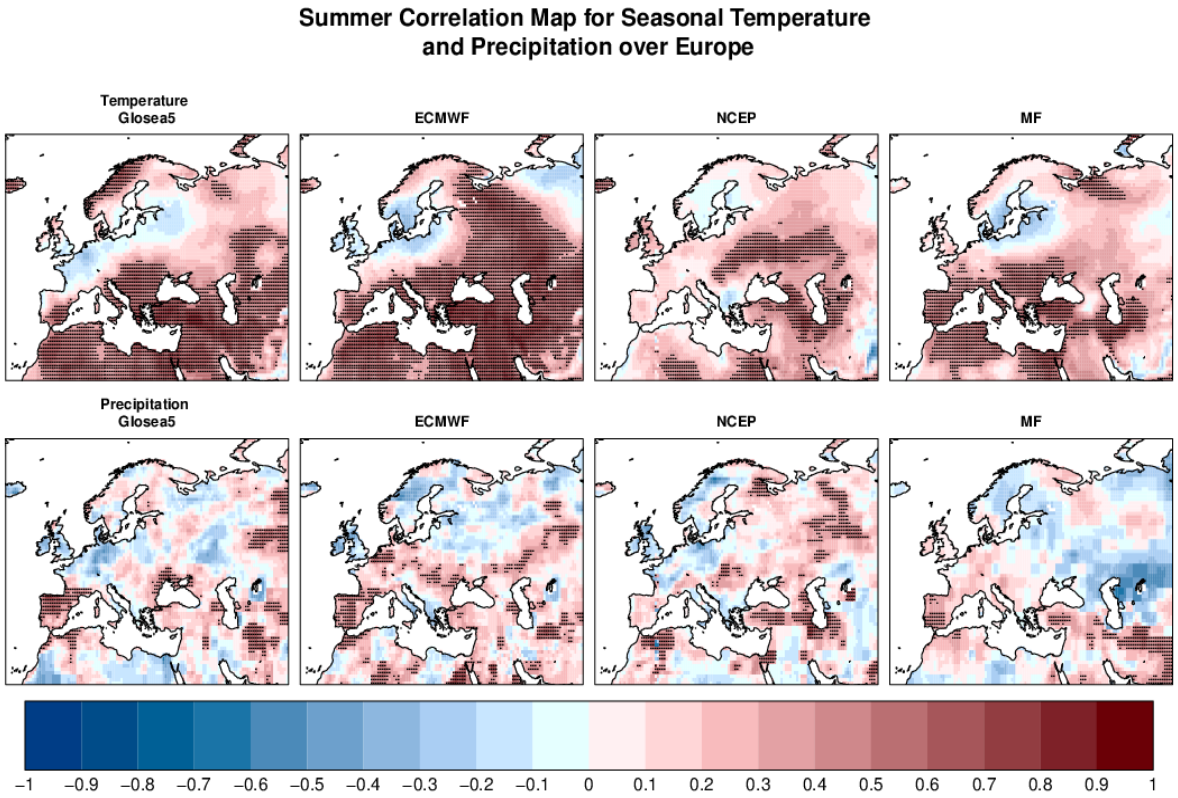


Figure 11: Seasonal(JJA) temporal correlation for European region between 1992-2012. The four maps on the top and the bottom correspond to seasonal temperature and precipitation, respectively. Red indicates positive correlation and blue indicates negative correlation. Dots mark the areas where the skill is significant at 95% confidence level.

Compared to the winter season, the skill for summer temperature in Europe is higher. ECMWF forecast system has highest positive and significant skill at 95% confidence level for summer seasonal temperature, covering most of the central and southern Europe (Fig 11). Glosea5 forecasting system also has significant and positive skill over south of Europe. MF has higher skill in the South-Western region compared to the rest of the Europe. Finally, the skill for seasonal precipitation in summer is more sporadic and lower compared to winter.

**Correlation Skill by Multi-Model Forecasting System**

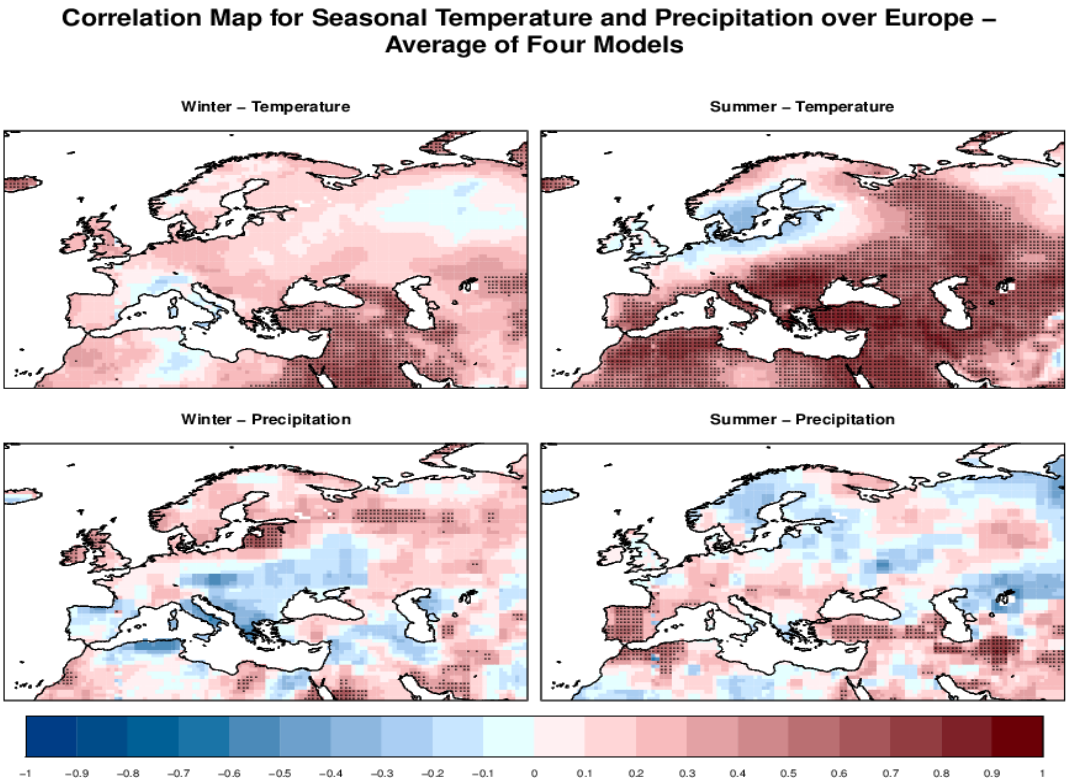


Figure 12: Seasonal temporal correlation for European region between 1992-2012 for multi-model forecast system consisting of Glosea5, ECMWF, NCEP and MF. The top two maps correspond to winter(DJF) and summer(JJA) seasonal temperature and the bottom are seasonal precipitation for winter(DJF) and summer(JJA). Red indicates positive correlation and blue indicates negative correlation. Dots mark the areas where the skill is significant at 95% confidence level.

Next, I take a multi-model approach to assess the skill of seasonal temperature and precipitation over Europe. The average of ensemble forecasts from all four model was computed to obtain a single forecast per grid point and the correlation was computed on this new multi-model forecasts. The results in Figure 12 show significant skill for predicting summer temperature over large area. However, some skill exhibited by Glosea5 and ECMWF individually in the western and eastern Europe are not exhibited by the multi-model forecast (MMF) system. There does not seem to be a noticeable improvement on the skill of MMF system for seasonal precipitation but this is consistent throughout the individual models.

### Correlation Skill by Model and Location

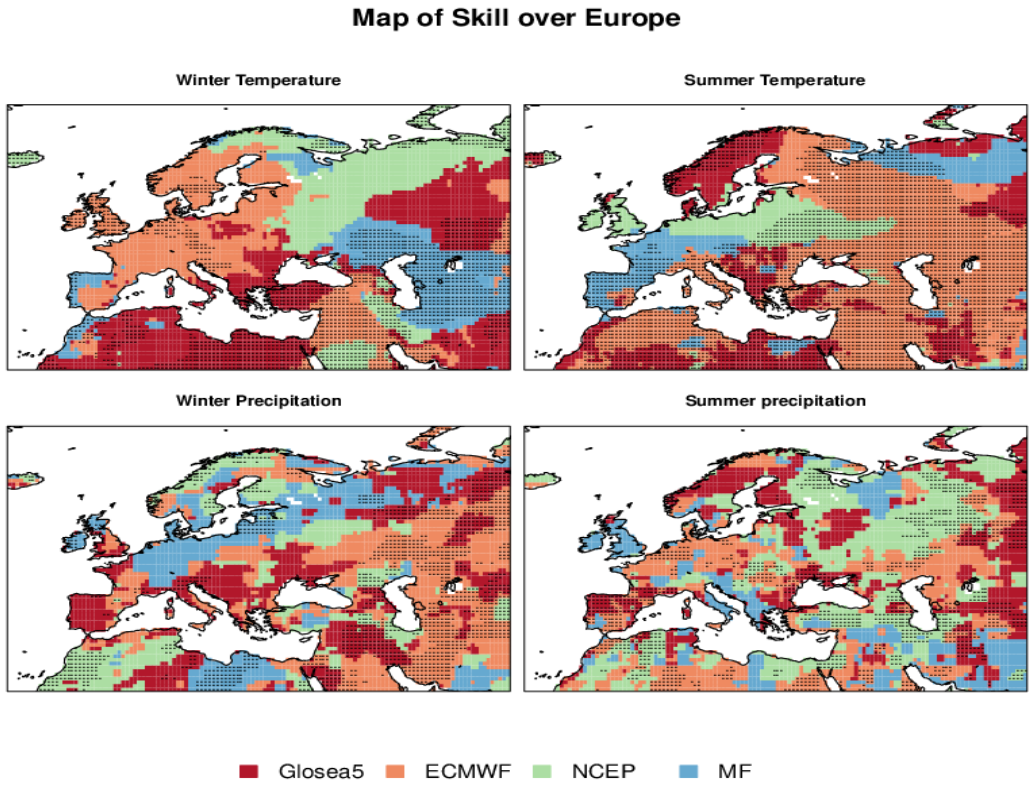


Figure 13: Model with maximum temporal correlation between 1992-2012 over various geographical regions in Europe. Competing models are Glosea5 (red), ECMWF(orange), NCEP(green) and MF(blue). The top two maps correspond to winter(DJF) and summer(JJA) seasonal temperature and the bottom are for winter(DJF) and summer(JJA) seasonal precipitation. Dots mark the areas where the skill is significant at 95% confidence level.

The final analysis based on correlation is aimed at identifying which model has highest skill for seasonal temperature and precipitation over various geographic regions in Europe (Fig. 12). The four colors in the map correspond to the four forecasting systems; Glosea5 (red), ECMWF (orange), NCEP (green) and MF (blue). Out of the correlation computed for each individual forecasting system, the maximum correlation among them was computed for each grid point. Then, each grid point was assigned to the model that had highest correlation among the four.

The skill for seasonal winter temperature seems to be more or less evenly divided among the four models (Fig. 13). ECMWF is the best performing forecast system for summer season. For winter season, the four model have best performance in different regions over Europe. The correlation skill among the four models for seasonal precipitation is scattered throughout Europe. ECMWF performs best in the eastern region for winter season. NCEP forecasts exhibit best skill in comparatively few locations.

## 3.2 Continuous Ranked Probability Skill Score (CRPSS)

CRPSS is a skill score based on the CRPS. It is a better measure of relationship between ensemble forecast and observation because it takes into account the full probability distribution obtained from the ensemble members and compares it with the verifying observation. It can also be interpreted as probabilistic generalization of the mean absolute error.

- CRPSS evaluates the percentage of forecasts that are more skillful than the reference forecast, climatology. For example, a value of CRPSS=0.2 indicates that the probabilistic forecast error is 20% less than the climatological forecast error.
- Negative values (in blue) imply that the skill of estimated forecast probabilities is worse than the use of climatological frequencies as forecast.
- Positive values of CRPSS (in red) indicate that the model is better than climatological probabilities.
- White areas on land surface represents scores that are lower than -1 showing particularly worse relationship between the ensemble distribution and the observation. The data over sea is masked and also represents white areas on the maps.

### CRPS Skill for Winter Season

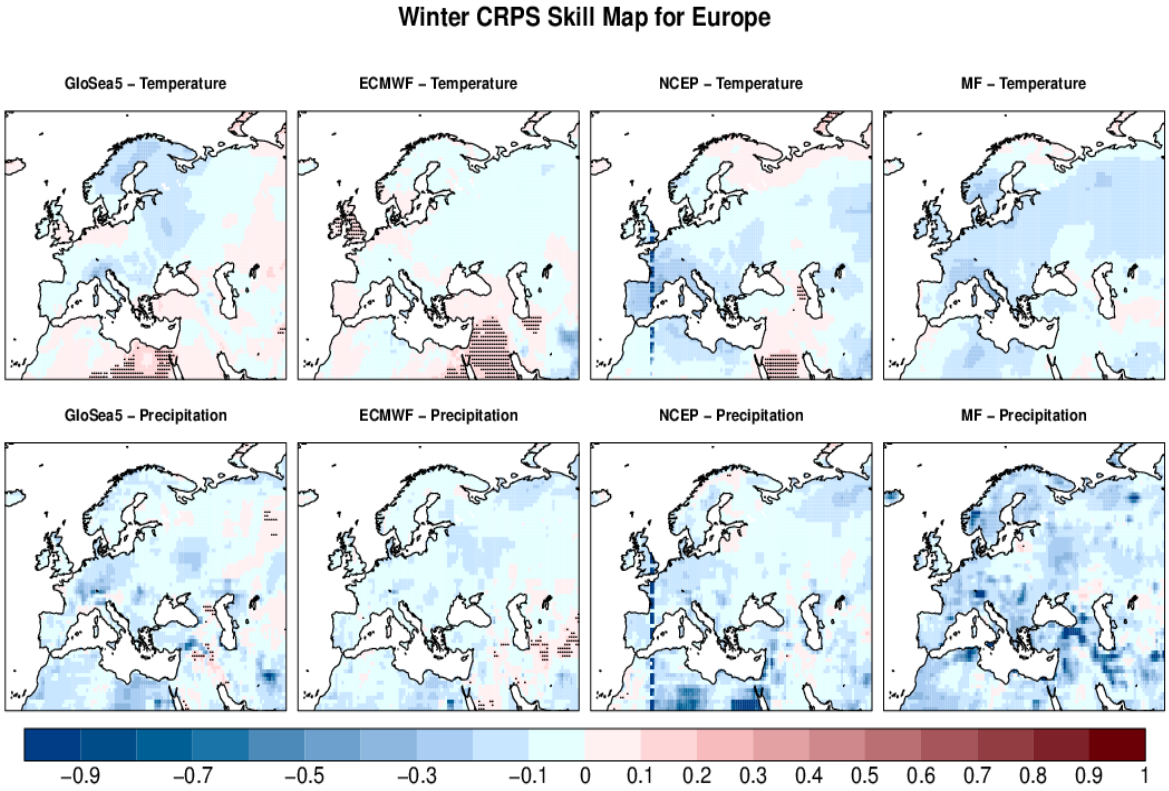


Figure 14: CRPSS for seasonal (DJF) temperature(top) and precipitation(bottom) between 1992-2012 in the European region. Red indicates skill higher than climatology and blue indicates worse skill than climatology. Dots mark the areas where the skill is significant at 95% confidence level.

The result of CRPSS on the assessment of seasonal temperature and precipitation forecast over Europe for winter show very limited correspondence between the probability distribution of ensemble members and the verifying observations. There is some positive and significant skill in ECMWF, which is limited to the south of Europe by the Mediterranean coast (Fig.14). The CRPSS skill for precipitation remains very low.

### CRPS Skill for Summer Season

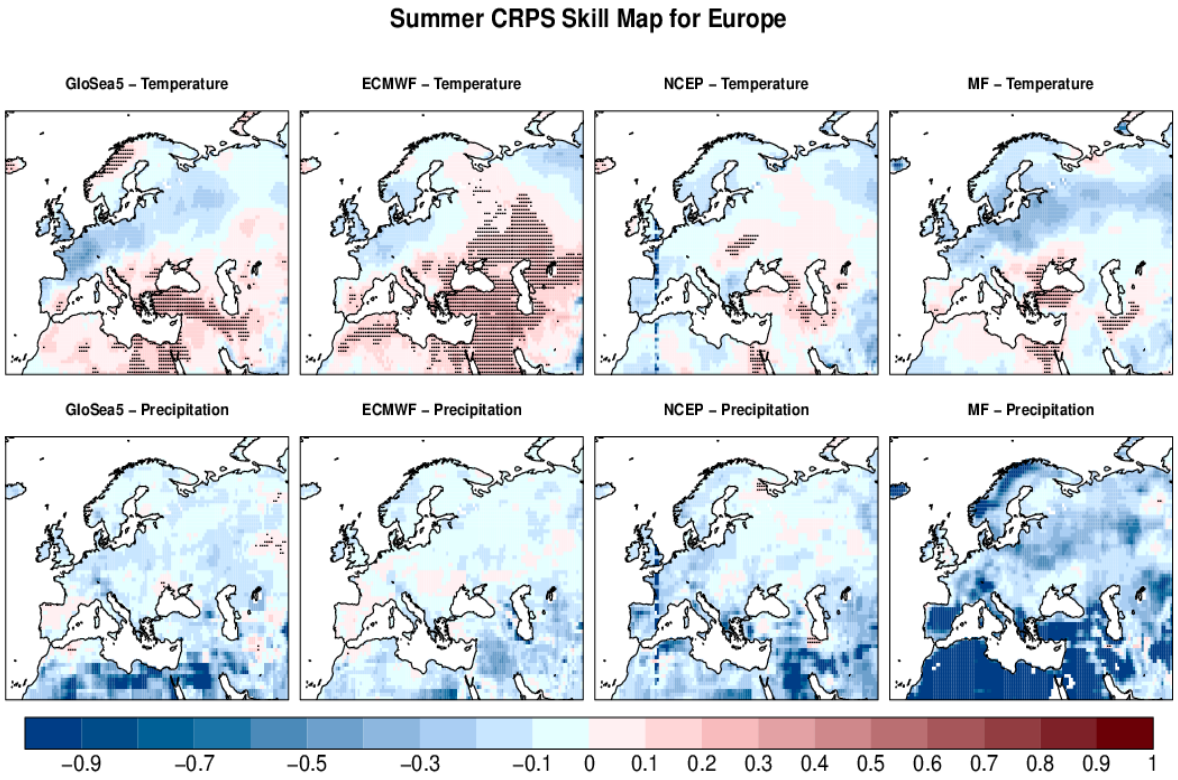


Figure 15: CRPSS for seasonal (JJA) temperature(top) and precipitation(bottom) between 1992-2012 in the European region. Red indicates skill higher than climatology and blue indicates worse skill than climatology. Dots mark the areas where the skill is significant at 95% confidence level.

Consistent with the case of correlation, CRPSS for summer is better compared to that of winter. This suggests that seasonal temperature predictability is higher for summer than for winter season in general. ECMWF has positive significant skill over central Europe (Figure 15). Some positive significant skill is also in Glosea5 in the South of Mediterranean.

None of the models exhibit significant skill for seasonal precipitation within Europe. Some positive skill is gained but very sporadically in just a few regions.

**CRPS Skill by Multi-Model Forecasting System**

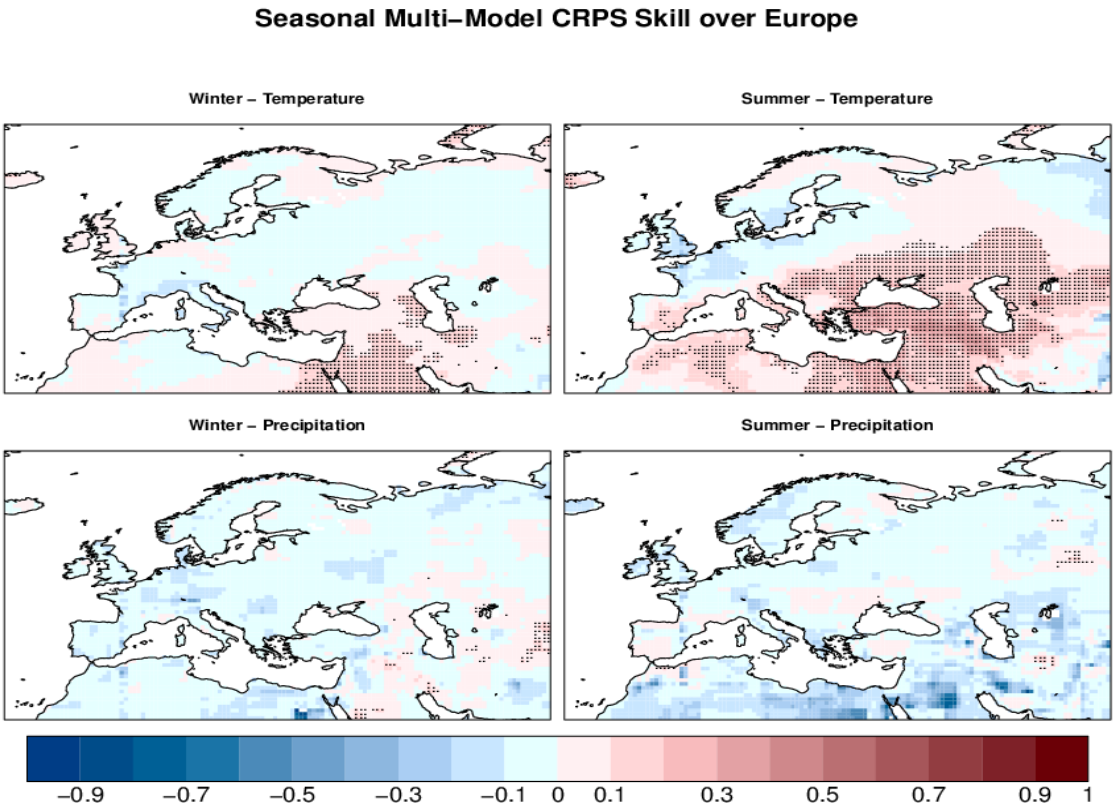


Figure 16: CRPSS for multi-model forecast system consisting of Glosea5, ECMWF, NCEP and MF for seasonal temperature(top) and precipitation(bottom) between 1992-2012 in Europe. Maps on left correspond to winter(DJF) season and to right correspond to summer(JJA). Red indicates better skill than climatology and blue indicates worse skill than climatology. Dots mark the areas where the skill is significant at 95% confidence level.

Finally, I compute CRPSS for MMF consisting of the four models to explore if they out-perform the individual forecast systems. As seen in Figure 16, the result is consistent as in the case of correlation where MMF has improved skill is central Europe. However, MMF skill decreased in some other areas such as Scandinavia.

As mentioned in the earlier sections, CRPS is not a *fair* scoring rule as it is sensitive to average ensemble spread including the frequency and magnitude of the outliers. When ensemble forecasts are compared with each other, the ensemble size affects the forecast skill. Note that ECMWF, which performs best on this skill measure compared to the rest of the model, has the highest number of ensemble, 51. On the other hand, MF which has lowest performance on this measure only has 15 members. Glosea5 and NCEP both have 24 ensemble members. In the next section, I discuss the results of FCRPS which adjusts these drawbacks of CRPS to some extent.

### 3.3 Fair Continuous Ranked Probability Skill Score (FCRPSS)

I computed FCRPSS for all four models as a fairer measure of scoring rule. FCRPSS is a skill score that measures whether the ensemble members are random samples of the same distribution as the observation. The interpretation of FCRPSS is similar to CRPSS.

The adjustment made by FCRPSS, lowered the skill for ECMWF suggesting the skill in CRPSS was driven in part due to the high number of ensemble members. The skill comparison of CRPSS (Fig. 14) and FCRPSS (Fig. 17) show only slight change in skill. The change is due to adjustments based on number of ensemble and its spread. The adjustment is particularly notable for ECMWF. Note that the significance of skill in the northern region of Caspian sea remains for ECMWF for the summer season even after the adjustment (Fig. 15 and Fig. 18). MMF performance (Fig. 16 is consistent with correlation skill and CRPSS. It is better in some areas but not in some other areas. The seasonal precipitation over Europe for winter and summer has low skill and this has been consistent throughout the analyses using all three methods.

### FCRPS Skill for Winter and Summer Season

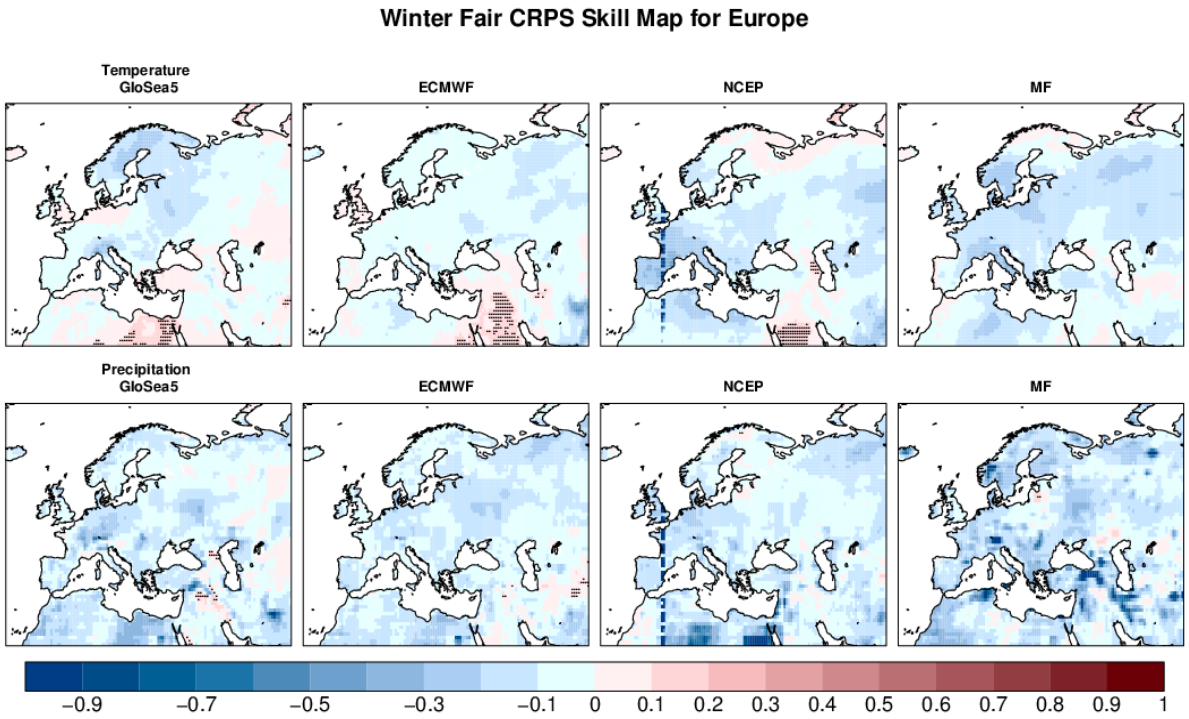


Figure 17: Same as for (Fig.14) but for FCRPSS values.

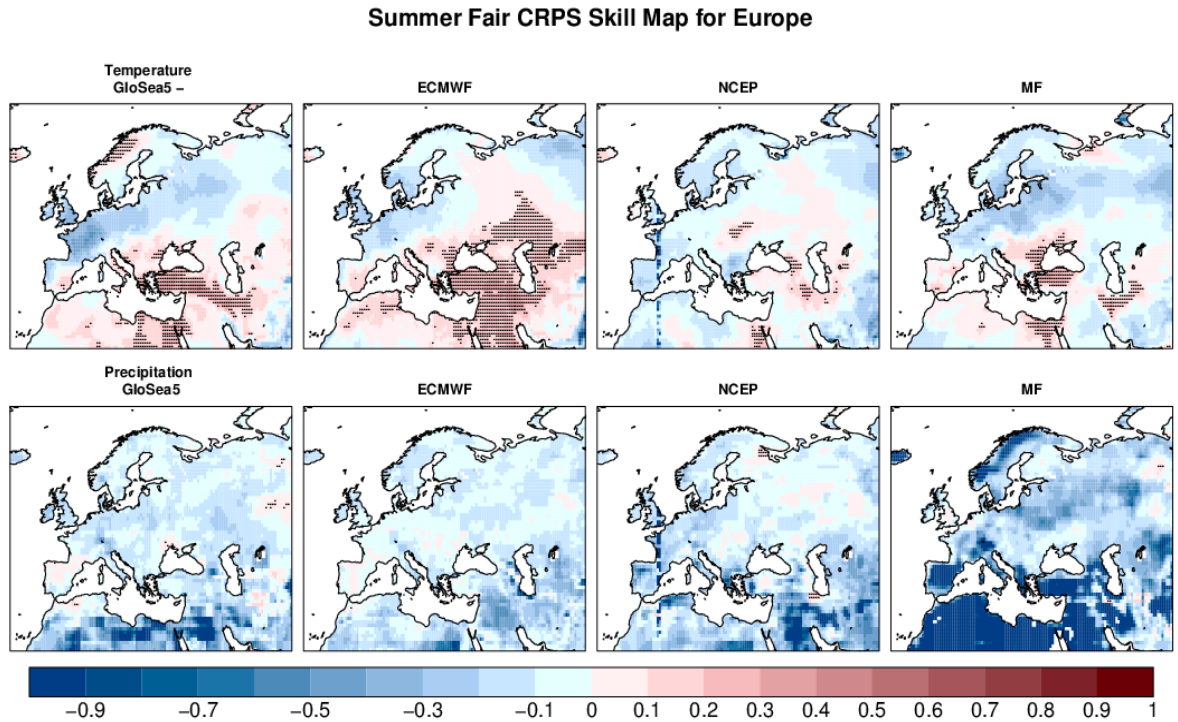


Figure 18: Same as for (Fig.15) but for FCRPSS values.

**FCRPS Skill for Multi-Model Forecasting System**

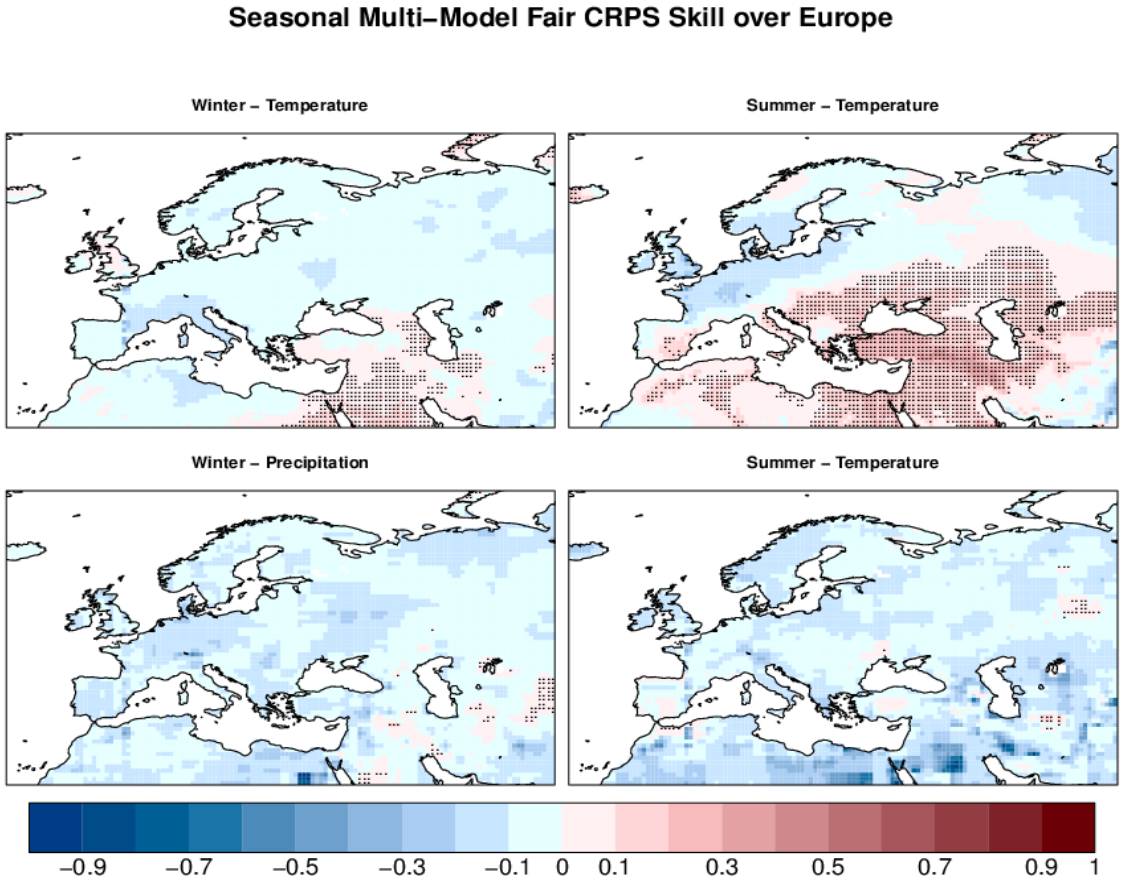


Figure 19: Same as for (Fig.16) but for FCRPSS values.

# 4 Conclusion

The goal of this thesis is to assess the seasonal temperature and precipitation forecasts provided by EUROSIP against observations dataset of temperature (ERA-Interim) and precipitation (GPCP) over Europe. SCFs are forecasts of climatic events at timescales of few weeks up to a few months. SCFs are currently underutilized in Europe, mostly due to limited predictability in the European region. In this thesis, however, I have shown that the potential for SCF exists and is particularly high during summer season in the central and southern regions of Europe.

In this thesis, I have highlighted the value of SCFs, which exists in various sectors ranging from agriculture, hydrology and insurance. There is a high potential of SCFs in aiding our decision making processes to be better prepared for the future. Furthermore, I have also highlighted the need for seasonal forecast verification. Due to the inherent uncertainties that exists in the atmospheric conditions of the Earth's climate system as well as due to the complexity of accurately depicting these conditions in the dynamical models, prediction are often penalized. Research and development in the scientific community has led to improvements in the forecasting technique. Ensemble is one such technique, the forecasts of which, I assessed in this thesis to determine the skill over the European region. Ensemble generation and modelling represents the "true" atmospheric state more accurately and accounts for some of these inherent uncertainties.

I used temporal correlation coefficient, CRPSS and FCRPSS as measures to assess the skill of SCFs. Correlation quantifies the maximum skill that can be obtained given the forecasts and observations. CRPSS measures the quality of ensemble by taking the full distribution of ensemble into account. However, it is sensitive to the number of ensemble members and its spread. Thus, CRPS favors the ensemble that has distributions similar to the distribution of the observations. FCRPS is a fairer scoring rule compared to CRPS. Unlike CRPS, FCRPS measures the underlying distribution of the ensembles and not just the empirical distribution. FCRPS assess whether the ensemble behave as a random sample generated from the same distribution as that of the observation. Thus, FCRPS measures how well each ensemble perform based on the observation.

The results of the assessment based on all three measures show that the skill for forecasting seasonal temperature exists in Europe. The skill is particularly notable for ECMWF, which remains even after adjustments over the number of ensemble members. Glosea5, NCEP and MF, all exhibit skill in different areas throughout Europe. Therefore, it is recommended to choose the forecast model based on the geographical area where they perform best. However, the skill for seasonal precipitation is particularly low. This is because in general precipitation is hard to observe and to forecast due to its high variability.

MMF assessment showed that multi-model system performs better in some areas but not throughout Europe. This is because the construction of multi-model is based on the average of the four models and

therefore is affected by the low skill of one or more of these models in a particular area. Hence, it would be interesting to observe the results of a weighted multi-model, which puts higher weight to the particular model that performs best in each grid point. After making for such adjustments, similar assessment using the three measures can be performed to see if multi-model can result in improved skill. This can be an area for further research.

Finally, significantly high correlation but fairly low CRPSS skill indicates that although the skill to predict the seasonal temperature and precipitation over Europe is high in many regions, predicting the full distribution of the seasonal temperature and precipitation is still a challenge. This requires re-calibration of the models to reduce the negative skill. However, more robust assessment of the sources of predictability is also required to better understand the atmospheric phenomena that contribute to seasonal climate anomalies. To conclude, while statistical methodologies can aid in climate forecast verification process, the key remains in our ability to understand and represent underlying physical processes more accurately.

# References


[1] Bart J.J.M. Van Den Hurk, Laurens M. Bouwer, Carlo Buontempo, Ralf Döscher, Ertug Ercin, Cedric Hananel, Johannes E. Hunink, Erik Kjellström, Bastian Klein, Maria Manez, Florian Pappenberger, Laurent Pouget, Maria-Helena Ramos, Philip J. Ward, Albrecht H. Weerts, and Janet B. Wijngaard. Improving predictions and management of hydrological extremes through climate services: www.imprex.eu. Climate Services, 1:6 – 11, 2016.

[2] Soares Marta Bruno and Dessai Suraje. Exploring the use of seasonal climate forecasts in europe through expert elicitation. Climate Risk Management, 10:8 – 16, 2015.

[3] Christine J. Kirchhoff, Maria Carmen Lemos, and Scott Kalafatis. Narrowing the gap between climate science and adaptation action: The role of boundary chains. Climate Risk Management, 9:1 – 5, 2015. Boundary Organizations.

[4] Francisco J. Doblas-Reyes, Javier García-Serrano, Fabian Lienert, Aida Pintó Biescas, and Luis R. L. Rodrigues. Seasonal climate predictability and forecasting: status and prospects. Wiley Interdisciplinary Reviews: Climate Change, 4(4):245–268, 2013.

[5] M. Balmaseda, D. Anderson, and A. Vidard. Impact of argo on analyses of the global ocean. Geophysical Research Letters, 34(16), 2007. L16605.

[6] Youmin Tang, Richard Kleeman, and Andrew M. Moore. Reliability of enso dynamical predictions. Journal of the Atmospheric Sciences, 62(6):1770–1791, 2005.

[7] E. Ebert, L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, and A. Watkins. Progress and challenges in forecast verification. Meteorological Applications, 20(2):130–139, 2013.

[8] Sarah J. Murphy, Richard Washington, Thomas E. Downing, Randall V. Martin, Gina Ziervogel, Anthony Preston, Martin Todd, Ruth Butterfield, and Jim Briden. Seasonal forecasting for climate hazards: Prospects and responses. Natural Hazards, 23(2):171–196, 2001.

[9] Hammer GL, Holzworth DP, and Stone R. The value of skill in seasonal climate forecasting to wheat crop management in a region with high climatic variability. Australian Journal of Agricultural Research, 47(5):717–737, 1996.

[10] Ouedraogo M., Zougmore R., Barry S., Some L., and Baki G. The value and benefits of using seasonal climate forecasts in agriculture: evidence from cowpea and sesame sectors in climate-smart villages of burkina faso. CCAFS Info Note. Copenhagen, Denmark: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), 2015.

[11] S. Singla, J.-P. Céron, E. Martin, F. Regimbeau, M. Déqué, F. Habets, and J.-P. Vidal. Predictability of soil moisture and river flows over france for the spring season. Hydrology and Earth System Sciences, 16(1):201–216, 2012.

[12] Mohamed Taher Kahil, Ariel Dinar, and Jose Albiac. Modeling water scarcity and droughts for policy adaptation to climate change in arid and semiarid regions. Journal of Hydrology, 522:95 – 109, 2015.

[13] Steven W. Martin, Barry J. Barnett, and Keith H. Coble. Developing And Pricing Precipitation Insurance. Journal of Agricultural and Resource Economics, 26(01), July 2001.

[14] Harvey Stern. The application of weather derivatives to mitigate the financial risk of climate variability and extreme weather events. Australian Meteorological Magazine, 282:171–182, 2001.

[15] Daniel E. Osgood Miguel A. Carriquiry. Index insurance, probabilistic climate forecasts, and production. The Journal of Risk and Insurance, 79(1):287–299, 2012.

[16] G. J. Boer. Climate trends in a seasonal forecasting system. Atmosphere-Ocean, 47(2):123–138, 2009.

[17] Michael A. Alexander, Ileana Bladé, Matthew Newman, John R. Lanzante, Ngar-Cheung Lau, and James D. Scott. The atmospheric bridge: The influence of enso teleconnections on air–sea interaction over the global oceans. Journal of Climate, 15(16):2205–2231, 2002.

[18] Timothy N. Stockdale, David L. T. Anderson, Magdalena A. Balmaseda, Francisco Doblas-Reyes, Laura Ferranti, Kristian Mogensen, Timothy N. Palmer, Franco Molteni, and Frederic Vitart. Ecmwf seasonal forecast system 3 and its prediction of sea surface temperature. Climate Dynamics, 37(3):455–471, 2011.

[19] J. Shukla. Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. Science, 282:728, October 1998.

[20] L. Goddard, S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher, and M.A. Cane. Current approaches to seasonal to interannual climate predictions. International Journal of Climatology, 21(9):1111–1152, 2001.

[21] Retish Senan, Yvan J. Orsolini, Antje Weisheimer, Frédéric Vitart, Gianpaolo Balsamo, Timothy N. Stockdale, Emanuel Dutra, Francisco J. Doblas-Reyes, and Droma Basang. Impact of springtime himalayan–tibetan plateau snowpack on the onset of the indian summer monsoon in coupled seasonal forecasts. Climate Dynamics, pages 1–17, 2016.

[22] Chloé Prodhomme, Francisco Doblas-Reyes, Omar Bellprat, and Emanuel Dutra. Impact of land-surface initialization on sub-seasonal to seasonal forecasts over europe. Climate Dynamics, pages 1–17, 2015.

[23] Virginie Guemas, Edward Blanchard-Wrigglesworth, Matthieu Chevallier, Jonathan J. Day, Michel Déqué, Francisco J. Doblas-Reyes, Neven S. Fučkar, Agathe Germe, Ed Hawkins, Sarah Keeley, Torben

Koenigk, David Salas y Mélia, and Steffen Tietsche. A review on arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Quarterly Journal of the Royal Meteorological Society, 142(695):546–561, 2016.

[24] M. an Co-Authors Latif. Dynamics of decadal climate variability and implications for its prediction. in Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society, 2:21–25, 2009.

[25] A A Scaife, A Arribas, E. Blockley, A. Brookshaw, R. T. Clark, N. Dunstone1, R. Eade, D. Fereday, C. K. Folland, M. Gordon, L. Hermanson, J. R. Knight, D. J. Lea, C. MacLachlan, A. Maidens, M. Martin, A. K. Peterson, D. Smith, M. Vellinga, E. Wallace, J. Waters, and A. Williams. Skillful long range prediction of European and North American winters. Geophysical Research Letters, 5:2514–2519, 2014.

[26] Yvan J. Orsolini and Nils G. Kvamstø. Role of eurasian snow cover in wintertime circulation: Decadal simulations forced with satellite observations. Journal of Geophysical Research: Atmospheres, 114(D19), 2009. D19108.

[27] R. D. Koster, S. P. P. Mahanama, T. J. Yamada, Gianpaolo Balsamo, A. A. Berg, M. Boisserie, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J.-H. Jeong, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. van den Hurk, F. Vitart, and E. F. Wood. The second phase of the global land–atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. Journal of Hydrometeorology, 12(5):805–822, 2011.

[28] F. J. Doblas-Reyes, R. Hagedorn, T. N. Palmer, and J.-J. Morcrette. Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. Geophysical Research Letters, 33(7):n/a–n/a, 2006. L07708.

[29] M. A. Liniger, H. Mathis, C. Appenzeller, and F. J. Doblas-Reyes. Realistic greenhouse gas forcing and seasonal forecasts. Geophysical Research Letters, 34(4):n/a–n/a, 2007. L04705.

[30] Brian D. Giles. Climate: into the 21st century, w. burroughs (editor). cambridge university press for world meteorological organization, 2003. 240 pages. isbn 0-521-79202-9. International Journal of Climatology, 25(13):1806–1807, 2005.

[31] G. Bala, K. Caldeira, M. Wickett, T. J. Phillips, D. B. Lobell, C. Delire, and A. Mirin. Combined climate and carbon-cycle effects of large-scale deforestation. 104(16):6550–6555, 2007.

[32] Allan H. Murphy. The finley affair: A signal event in the history of forecast verification. Weather and Forecasting, 11(1):3–20, 1996.

[33] Ian T. Jolliffe and David B. Stephenson, editors. John Wiley and Sons, Ltd, 2011.

[34] Allan H. Murphy. What is a good forecast? an essay on the nature of goodness in weather forecasting. Weather and Forecasting, 8(2):281–293, 1993.

[35] Harold E. Brooks and Charles A. Doswell III. A comparison of measures-oriented and distributions-oriented approaches to forecast verification. Weather and Forecasting, 11(3):288–303, 1996.

[36] Acacia S. Pepler, Leandro B. Díaz, Chloé Prodhomme, Francisco J. Doblas-Reyes, and Arun Kumar. 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, 2015. The World Climate Research Program Grand Challenge on Extremes – WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events.

[37] The era-interim archive version 2.0. http://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20. Accessed: 2016-05-30.

[38] P. Berrisford, D.P. Dee, P. Poli, R. Brugge, K. Fielding, M. Fuentes, P.W. Kållberg, S. Kobayashi, S. Uppala, and A. Simmons. The era-interim archive version 2.0. Shinfield Park, Reading, November 2011.

[39] Robert F. Adler, George J. Huffman, Alfred Chang, Ralph Ferraro, Ping-Ping Xie, John Janowiak, Bruno Rudolf, Udo Schneider, Scott Curtis, David Bolvin, Arnold Gruber, Joel Susskind, Philip Arkin, and Eric Nelkin. The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979–present). Journal of Hydrometeorology, 4(6):1147–1167, 2003.

[40] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477):359–378, 2007.

[41] Hans Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15(5):559–570, 2000.

[42] Anthony S. Tay Francis X. Diebold, Todd A. Gunther. Evaluating density forecasts with applications to financial risk management. International Economic Review, 39(4):863–883, 1998.

[43] C. A. T. Ferro. Fair scores for ensemble forecasts. Quarterly Journal of the Royal Meteorological Society, 140(683):1917–1923, 2014.

[44] Thomas E. Fricker, Christopher A. T. Ferro, and David B. Stephenson. Three recommendations for evaluating climate predictions. Meteorological Applications, 20(2):246–255, 2013.