Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Meteorologie und Klimatologie MeteoSchweiz

Veröffentlichung MeteoSchweiz Nr. 76

Probabilistic verification of operational monthly temperature forecasts

Daniel Baggenstos

Veröffentlichung MeteoSchweiz Nr. 76

ISSN: 1422-1381

Probabilistic verification of operational monthly temperature forecasts

Daniel Baggenstos

Bitte zitieren Sie diese Veröffentlichung folgendermassen
Baggenstos, D: 2007, Probabilistic verification of operational monthly temperature forecasts, *Veröffentlichungen der MeteoSchweiz*, **76**, 52 pp.

Herausgeber
Bundesamt für Meteorologie und Klimatologie, MeteoSchweiz, © 2007

MeteoSchweiz
Krähbühlstrasse 58
CH-8044 Zürich
T +41 44 256 91 11
www.meteoschweiz.ch

Weitere Standorte
CH-8058 Zürich-Flughafen
CH-6605 Locarno Monti
CH-1211 Genève 2
CH-1530 Payerne

Contents

ABSTRACT

Monthly forecasting bridges the gap between traditional medium-range weather forecasts and seasonal predictions. Forecasts in the range of 1 to 4 weeks are vital to many applications in the context of weather and climate risk management. Since 2004, the ECMWF runs a monthly ensemble prediction system operationally.

It is the aim of this study to assess prediction skill of this monthly forecasting system in dependence of lead time, season and geographical location. The study considers all 12 years of hindcasts available. Weekly averages of near-surface (2m) temperature predictions are verified. The monthly ensemble forecasts are evaluated grid point wise against ERA-40 reanalysis data (until 2001) and the ECMWF operational analysis (after 2001) respectively. While our investigations are carried out for the entire globe, special focus is on skill in Europe and North America.

As a probabilistic skill metric, the discrete ranked probability skill score $RPSS_D$, which is a modified version of the widely used RPSS, is applied. In contrast to the classical RPSS, the $RPSS_D$ is a debiased skill score in that it is insensitive to ensemble size. This is an important property in the present context, given the differing numbers of ensemble members in the datasets available (51 in the forecasts, 5 in all hindcasts).

The results indicate that monthly forecasts can be useful at lead times of 2 to 4 weeks depending on the region and season considered. Forecasts over sea remain skillful longer than over land, most notably in the ENSO region and the central Atlantic, where skill is high (up to a $RPSS_D$ of 0.5) even at four weeks lead time. Yet, there are land areas, e.g. tropical South America where the model seems to retain significant skill into the fourth week. For most of the extratropical landmasses, however, skill appears to drop sharply after 1 week. In these regions, the second week is characterized by a skill score of approximately 0.1 while being strongly dependent on seasons. In the third and fourth week, skill seems to be very low over extratropical land generally.

On the global scale, seasonal skill variations are dominated by El Niño and its teleconnections. In central Europe, monthly forecasts of weeks 2 and 3 indicate best predictability in fall and winter, while summer appears particularly difficult. In the North American domain, week 2 forecasts are commonly skillful but spatial and seasonal differences are large, with forecasts in winter at the east coast showing especially high skill ($RPSS_D$ of 0.3).

Ultimately, monthly forecasts perform better than or as well as reference strategies such as persistence and climatology consistently, encouraging their use worldwide.

1 Introduction

1.1 Motivation

Monthly forecasts fill the gap between traditional medium-range weather forecasts and seasonal predictions. Forecasts in the range of 1 to 4 weeks are essential to many applications in the context of weather and climate risk management and mitigation. They are expected to have some skill in forecasting persistent blocking events associated with heat waves or cold spells (Schwierz et al. 2006). Particularly in the light of climate change and a possible increase in the frequency of extreme temperature periods, monthly forecasts could serve as an early warning system for such events. However, even if it is not possible to give precise forecasts, information combined with uncertainty can be valuable to users.

Potential users of monthly forecasts include, amongst others, companies from the energy, agriculture and insurance sectors. For instance, the prospect of the next month being very dry might influence decision making of hydropower companies, while the knowledge of a coming cold spell could be of interest to electric power stations. Also, agriculture has a need for the prognosis of temperature and precipitation in order to maximize crop yields. Jones et al. (2000) and Hammer et al. (2001) show that the potential benefit of seasonal climate predictions for crop modeling is very high. The insurance industry could profit greatly from valid forecasts in the monthly range in the context of weather derivatives (Zeng 2000). As such forecasts become available, an increasing number of users will adopt them to their benefit.

Since 2004 the European Centre for Medium-Range Weather Forecasts (ECMWF) operationally runs a monthly ensemble prediction system based on a dynamic numerical model, i.e. on the integration of a set of prognostic equations describing the evolution of the atmosphere. A fundamentally different approach to obtain forecasts on these timescales are statistical models, which use correlations observed in long time series or teleconnections to infer a probable future. While this study deals exclusively with dynamical forecasts, it should be remembered that statistical forecasting can be a valid alternative, especially for extended-range predictions.

A systematic verification is important for every forecasting system because it assesses the quality of the forecast. It is crucial for the interpretation and application to know its strengths and weaknesses, seasonal and/or regional dependencies. Unless such information is available, a potential user cannot decide how much he should or should not trust a certain forecast.

Previous studies have already provided first estimates of the skill of the ECMWF 32-days forecasting system. Here we carry out a more comprehensive and fully probabilistic verification, considering all hindcast data available. It is the aim to provide a systematic skill evaluation and to assess prediction skill of the monthly forecasting system in dependence of lead time, season and geographic location. As a skill metric, the RPSS_D will be applied. While our evaluations are carried out for the entire globe, special focus will be on skill in Europe. Monthly forecasts can be issued for any parameter of the model (such as surface temperature, total precipitation or geopotential height). In this study, we assess forecasts of 2 meter temperature (T2), which is an important parameter for most users. The verification of other parameters is left for future research.

The thesis is organized as follows: The following sections in this chapter introduce the physical concepts of predictability and monthly forecasting. Chapter 2 describes the data basis and the methods used. In chapter 3 the results are presented and chapter 4 presents a discussion and a summary of the main points.

1.2 Predictability

On the threshold of the 20th century, scientists believed that one day it would be possible to predict weather exactly. The common belief was that, given enough computing power, the equations which precisely represent the evolution of the atmosphere will be solvable (Richardson 1922). It was only in 1963 when Lorenz (1969) showed that the atmosphere is a non-linear dynamic system which is now commonly understood as dynamic chaos. The main characteristic of such a system is that its evolution is highly sensitive on its initial state. Due to its non-linearity, even smallest disturbances from a given state can grow rapidly within a couple of days and therefore alter the predicted weather totally.

Figure 1.1 shows a schematic visualization of dynamic chaos, the Lorenz attractor. Every point in 3-D space resembles a different state of the system. The trajectories visualize the temporal evolution of the system. Obviously, the system circles on similar paths around two attractors. In this simple model, the attractors could e.g. stand for El Niño and no El Niño events, or for NAO+ and NAO- phases (NAO stands for North Atlantic Oscillation). Depending on the characteristics of the system, the trajectories will or will not follow predictable courses. Given that, in the real world, it will never be possible to accurately determine the exact state of the entire atmosphere at a given time, be it due to errors in terms of spatial coverage or accuracy of the individual measurement, the Lorenz attractor illustrates that there will never be something like a perfect forecast. Indeed, every forecast is uncertain to some degree. Traditionally, numerical weather prediction (NWP) models are deterministic, that is, capable of producing only a single forecast. However, given that a probability distribution would represent the situation more precisely, probabilistic approaches are becoming more and more common.

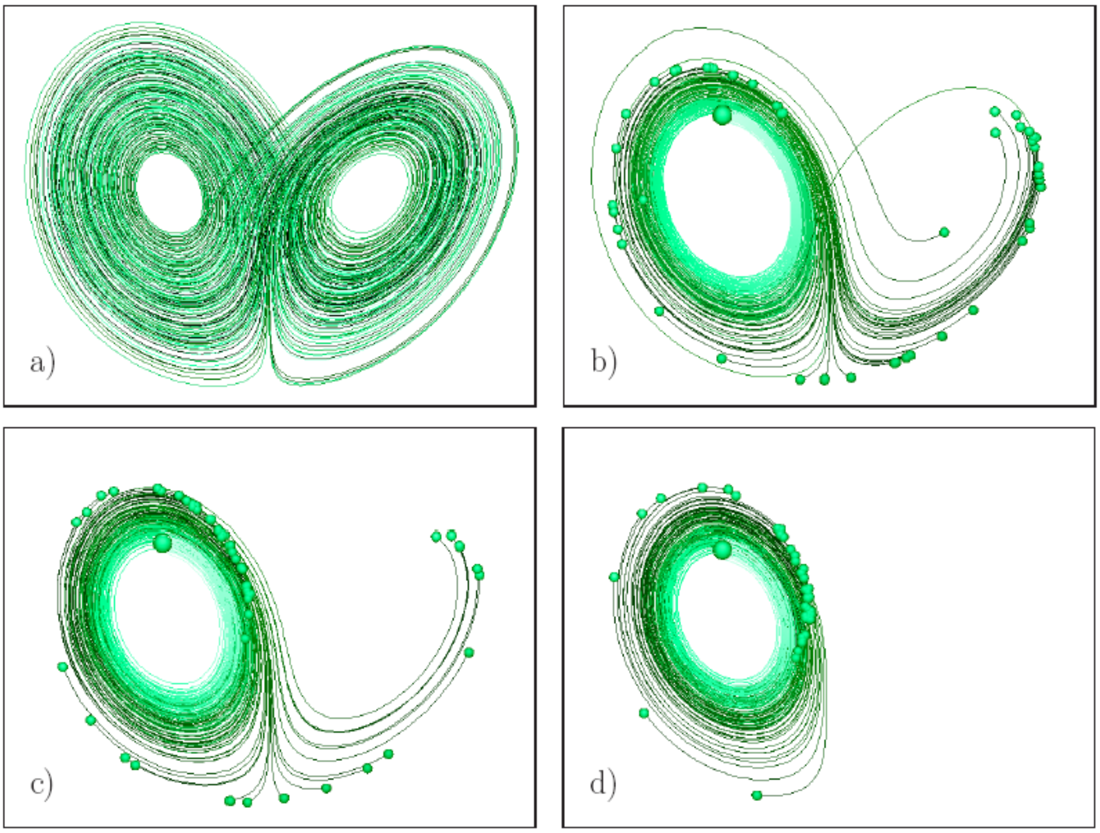

Figure 1.1: a) Illustration of a 3-D Lorenz attractor. b), c) and d): The large sphere resembles 50 slightly different initial values. The forecasts end at the small spheres. The initial values lie in a region of (b) poor, (c) moderate and (d) good predictability. From Liniger (2003).

Considering the non-linearity mentioned, atmospheric modeling is to a large degree an initial value problem. Meanwhile, sophisticated methods for estimating the initial conditions have been developed and the network of measurement stations has improved. As a consequence of that, NWPs have improved consistently over the last decades, as illustrated e.g. by Simmons and Hollingsworth (2002). Other sources of error include a simplification of the real world, resulting in the need for discretization and parameterization, or uncertainties in the boundary conditions. The latter become especially important when considering longer timescales. For example, Liniger et al. (2007) investigate the effect of greenhouse gas concentration trends on seasonal forecasts. The authors show that forecasts with realistic greenhouse gas concentrations are in better agreement with observed temperatures than forecasts with constant concentrations, illustrating the importance of correct boundary conditions.

The time range in which deterministic forecasts can still be considered to possess valuable information does not exceed 7-10 days in the extratropics. This is about the duration attributed to the memory of the atmosphere. There have been attempts, for example by Jung and Vitart (2006), to enhance the skill of such medium-range forecasts with the implementation of an interactive ocean. Their study, which focuses on northern hemispheric wintertime, indicates that an atmosphere/ocean coupling has little to no benefit in the medium-range. For longer (e.g. monthly) timescales, a fundamentally different approach needs to be applied which will be presented in section 1.3.

Probabilistic forecasting

In order to quantify the uncertainty inherent in deterministic forecasts, NWP models are often run several times from slightly perturbed initial conditions, thus yielding not only a single forecast value, but an entire 'ensemble' of possible forecast realizations (Wilks 1995). Such a system is termed ensemble prediction system (EPS) and its individual runs are called 'members'. The wider the members disperse from each other, the larger is the uncertainty due to inaccurate initial conditions.

EPS are employed on every timescale, but naturally their importance grows as integration time becomes longer and the effect of uncertain initial conditions destroys more and more of the predictive signal. Therefore, it makes sense to consider the ensemble distribution as a whole and to issue probabilistic forecasts. For example, if 30 out of 50 ensemble members indicate that the temperature will go above a given threshold, the probability of this event is issued as 60%.

This method of assessing the uncertainty involved is nowadays used routinely by most weather services, at least in the context of medium-range and long-term predictions. A description of three widely used medium-range EPS from the ECMWF, the Meteorological Service of Canada (MSC) and the National Centers for Environmental Prediction (NCEP), is provided by Buizza et al. (2005).

Seasonal forecasting

Above, it has been stated that the predictability of the atmosphere reaches not beyond 10 days. Longer-range forecasts therefore can not be seen as classic weather forecasts any more. Rather they have the nature of a climate prediction, i.e. the prediction of average weather conditions instead of explicit weather events. Such climate forecasts are possible to a certain degree, since the climate is influenced and determined by processes which can partially be predicted. On seasonal timescales, the oceans with their heat capacity and relative inertia present the most important source of memory which can be exploited by forecasters. Therefore, seasonal forecasts are computed with coupled ocean-atmosphere general circulation models.

Because seasonal forecasts must be understood as climate predictions, they are not issued for particular days but rather as information over a fixed period (usually one or several months). The idea behind this is that temporally averaged variables, such as temperature over a whole month, are only to a little degree dependent on day-to-day variations induced e.g. by the passing of cyclones. A larger input is provided by boundary conditions such as sea surface temperature, snow cover or soil moisture which have a significantly longer memory than the atmosphere itself. For example, Shongwe et al. (2007) have found a high seasonal predictive skill of cold spells in central and eastern Europe in winter and spring and attributes this skill to snow effects. They conclude that improvements in snow analysis and land surface parameterizations are expected to increase the skill of seasonal forecasts.

Meanwhile, seasonal forecasts are well-established and computed by many institutions or organizations. They have especially high skill in predicting El Niño phases. Van Oldenborgh et al. (2005) show that dynamic seasonal forecasts outperform statistical forecasting techniques especially during the initial stages of an El Niño. With its many links and teleconnections to areas outside the Pacific, El Niño is essentially a global phenomenon and much of the predictability on the globe in seasonal timescales can be associated to it.

1.3 Monthly forecasting

Monthly forecasts encompass predictions in the range of several weeks, and are often also referred to as subseasonal forecasts. It is a relatively new discipline in atmospheric modeling. Besides the ECMWF, only the Japan Meteorological Agency (JMA) issues monthly forecasts based on a dynamical system operationally. In terms of predictability, monthly forecasts fall into the gap between medium-range and seasonal forecasts. The former draw its validity from a forward-in-time integration of the initial state of the atmosphere itself, the latter from the extended predictability of boundary conditions determining the climate. Monthly forecasts comprise characteristics of both. This dualism is reflected by the fact that monthly forecasting systems reveal characteristics of both seasonal predictions and medium-range NWP: Similar to seasonal predictions, the monthly forecasting system builds upon the oceanic predictability and the boundary conditions of snow cover and soil moisture. Similar to medium-range forecasts, on the other hand, the model is applied in higher resolution and with more members.

Yet, monthly predictions are not only to be seen as a mixture of NWP and seasonal forecasting. Indeed, the atmospheric system reveals sources of predictability on exactly the timescales considered by monthly forecasts. In particular, these are the Madden-Julian Oscillation (MJO) and stratosphere-troposphere interaction.

The Madden-Julian Oscillation

Madden and Julian (1994) provide a review of the observational evidence gathered about the MJO. The 40 to 50 day oscillation results from large-scale circulation patterns that move eastward from at least the Indian Ocean to the Pacific. Disturbances in the wind field in the upper troposphere often propagate even around the globe. Associated to the circulation cells are complex convective regions that also show an eastward movement. The MJO is the dominant mode in intraseasonal timescales throughout the tropics. The oscillation has to be seen as a coupled atmospheric/oceanic phenomenon (Slingo et al. 1999). Although it is a strict tropical feature, it has an effect on the extratropics as well, as shown by Reichler and Roads (2005) with a perfect model approach. In their setting, the useful forecast skill of intraseasonal variability is about 4 weeks. They also stress that the oscillation has to be seen as a coupled feature, and that extended-range forecasting may require the use of fully interactive ocean models. Also Jones et al. (2004) highlights the importance of the MJO in modulating weather variability in the extratropics. It stays unclear, however, to what extent the MJO affects weather in the midlatitudes, in particular in Europe.

Stratosphere-troposphere interaction

Another known source of predictability is given by stratosphere-troposphere interaction. Evidence has been gathered that the stratosphere influences the troposphere and its circulation. Observational studies and numerical experiments (e.g. Hartley et al. 1998) showed that a coupling is likely. Another finding is that the NAO, the dominant mode of interseasonal variability in Europe, could be strongly subject to stratospheric forcing. Baldwin and Dunkerton (2001) describe how large variations in the strength of the stratospheric circulation descend within the stratosphere and are followed by anomalous tropospheric weather regimes. They suggest using stratospheric precursors as a predictor of tropospheric weather variability. Haynes (2005) summarizes the findings in this area and concludes that the stratosphere plays an active role in tropospheric variations. If such a connection exists, the slower dynamics of the stratosphere could increase the predictability in the troposphere by providing a longer memory of the initial state.

Skill of monthly forecasts

Some first verification studies on the skill of monthly predictions have been carried out by Vitart (2004). He shows that monthly forecasts do exhibit some moderate skill after the initial week, especially when being compared to climatology. The second forecast week performs generally better than both climatology and persistence. After about 20 days the performance depends strongly on the geographical location, with Europe being particularly difficult. He concludes that the scores show strong seasonal variability, with highest skill in winter. This thesis thematically builds upon the work of Vitart and continues with the systematic skill evaluation of the ECMWF's monthly prediction system.

As mentioned above, a fully probabilistic approach is applied, considering all hindcast data available.

Monthly forecasting at MeteoSwiss

The Federal Office of Meteorology and Climatology (MeteoSwiss) is currently testing and evaluating the use of monthly forecasts. A not yet operational product derived from the ECMWF monthly forecasts has been set up in 2004 and is running with minor modifications since then. Figure 1.2 shows a sample forecast from January 18, 2007. Due to the probabilistic nature of the forecasts, the results are typically displayed as probability maps, indicating, for example, how likely it is that the week's average temperature will be in the climatological upper tercile.

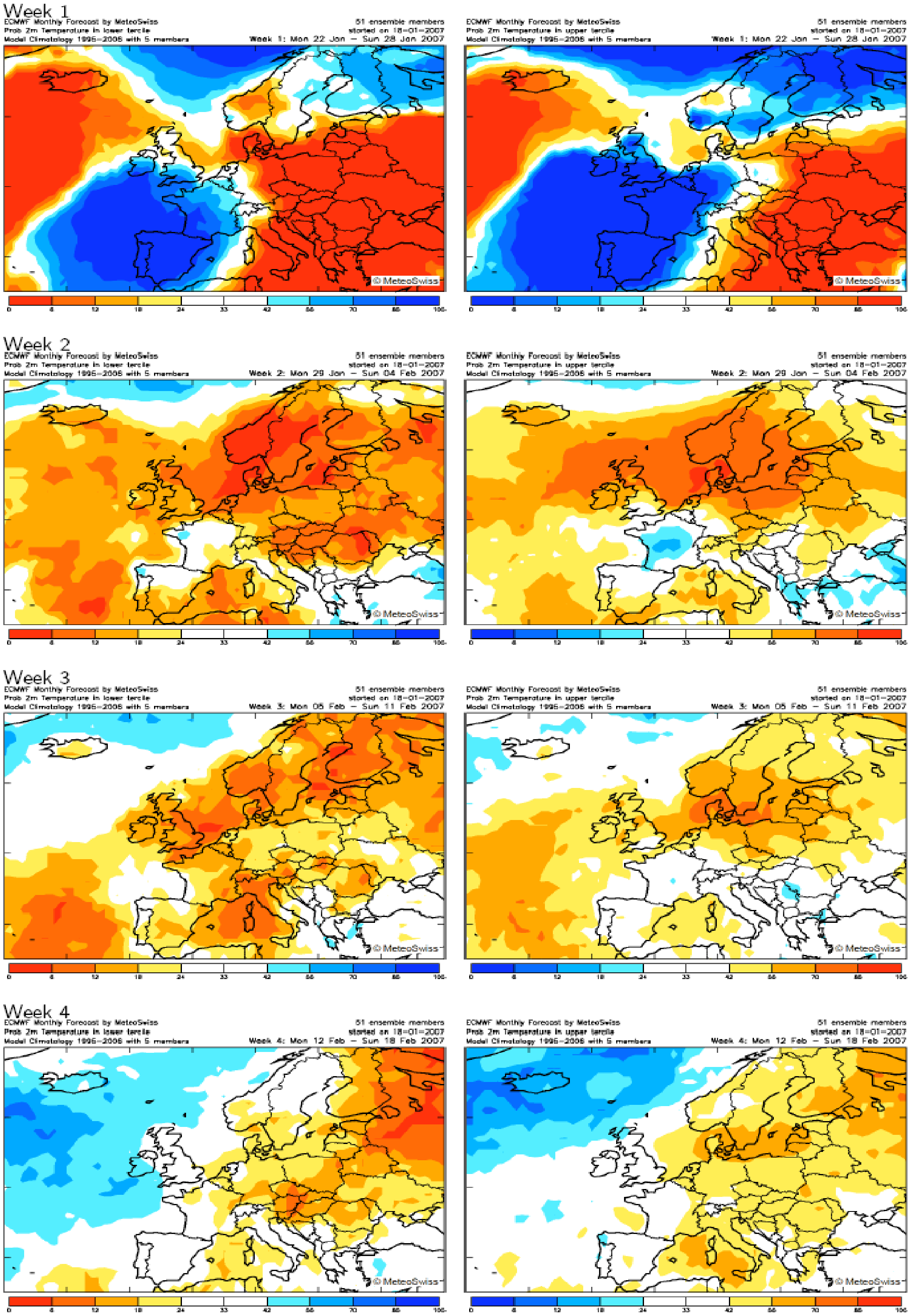

Figure 1.2: Probabilistic weekly temperature forecast for the European Domain. Shown are tercile probabilities for the upper and lower climatological third (left and right). The first week shows a distinct pattern. For weeks 2-4 the signal is more diffuse and converges to the climatological mean. Courtesy of MeteoSwiss.

2 Data and Methods

2.1 Data

2.1.1 ECMWF monthly forecasting system

The probabilistic forecasts evaluated in this study are based on the ECMWF monthly forecasting system, which is a coupled ocean-atmosphere, hydrostatic, global model. Monthly forecasts from ECMWF consist of 51 members. They are integrated over 32 days, starting each Thursday since October 2004. Thus the first forecast week corresponds to days 5-11, the second to 12-18, the third to 19-25 and the fourth to days 26-32 of the integrations. The horizontal resolution is approximately $1 \times 1°$, which lies between the resolution of medium-range and seasonal forecasts. The ocean model is HOPE from the Max Planck Institute for Meteorology in Hamburg. Atmosphere/ocean interaction is synchronized every hour through the OASIS interface. The atmospheric initial conditions are obtained from the ECMWF operational analysis.

Additionally, for each forecast, corresponding hindcasts of the previous 12 years are calculated (say the forecast starts on 13.4.2006, then the hindcasts are calculated from initial values of 13.4.1994, 13.4.1995, ..., 13.4.2005), using an ensemble size of 5 members. Such hindcasts are necessary to estimate the model climatology which is required to calibrate the forecasts w.r.t. the real climatology. ERA-40 and the operational analysis, both of which are introduced below, provide the initial conditions for the hindcasts.

A more detailed and technical description and some remarks on oceanic initial conditions and ensemble generation can be found in Vitart (2004) or the ECMWF online documentation (ECMWF 2006).

2.1.2 Observations

Every verification study needs a set of observations that can be treated as the real outcome of the forecasts under inspection. In this case we consider the European Re-Analysis (ERA-40) and the ECMWF operational analysis as the 'truth'. These datasets are not observations in a classic sense, but represent the best estimate of the state of the atmosphere at a time by the numerical model. They are obtained by feeding a multitude of observations (from surface stations, radiosondes, aircrafts, ocean-buoys, satellites etc.) from all around the world into a dynamical model, which interpolates these data on a grid and adjusts them in such a way that the result is the most physically consistent state of the atmosphere. Such data are often denoted as 'quasi-observations'.

One of the most comprehensive quasi-observational datasets currently available is the ERA-40 reanalysis. It covers the time from September 1957 to August 2002. In a reanalysis, all available data are interpolated with the same, up-to-date model, thus yielding coherent results that are not influenced by model adjustments. Uppala et al. (2005) give a thorough description of the data acquisition and the changes in data type and coverage over the period. They also highlight two major problems, namely excessive precipitation over the tropics and a too strong stratospheric circulation, but conclude that ERA-40 has nevertheless good quality. This is supported by a study from Simmons et al. (2004), who compare mean air surface temperature from ERA-40 with monthly station data of the CRUTEM2v dataset. They show that the 2m temperatures are in good agreement in terms of variability and trends from the 1970s onward, which justifies using ERA-40 in this study. Another strong indicator for the quality of ERA-40 is provided by Kunz et al. (2007) in a comparison of daily surface temperature between the reanalysis and Swiss station data. Overall, and especially in the 1990s, the two datasets are in good agreement.

For the time after September 2002, the ECMWF operational analysis is used. It is similar to the reanalysis in that it represents the best guess of the atmosphere at that time. The difference to a reanalysis lies in the amount of data assimilated into the system (because of nearly real-time computation) and the model version, which is updated continuously for the analysis. The analysis serves as the starting point for the traditional medium-range model forecasts.

2.2 Methodology

In order to assess the quality of the monthly forecasts, several techniques will be applied. Thereby, two important issues need to be considered.

First, it is unlikely that the monthly forecasting system has any value in predicting weather precisely on a certain day, several weeks in advance (Lorenz 1969). Rather, it is expected that the system has some potential in forecasting averages of several days. Typically, and also in this study, weekly means are considered and evaluated.

Second, the monthly forecasting system is an EPS. The ensemble allows to issue a probabilistic forecast. It has been shown that it is difficult to verify probabilistic forecasts in a way that includes all the information (Candille and Talagrand 2005). In particular, a probability forecast cannot be validated by a single observation. Therefore, it is not possible to assess model skill on the basis of a single probabilistic forecast. The validation can only be statistical, with a large sample of forecast realizations and verifying observations.

2.2.1 Skill assessment

Various measures can be used to assess the model's predictive capabilities. Here, we consider two aspects of model performance: (i) the systematic deviation from the climate (model drift, bias), and (ii) the skill of probabilistic categorical forecasts (RPSS_D).

Bias

The bias is a widely used and easily calculated attribute of forecast model quality. It measures the difference between the average forecast and the average observed value of the predictand and is therefore often referred to as systematic bias (Wilks 1995). It is defined as

$$Bias = \langle y \rangle - \langle o \rangle \tag{2.1}$$

with y denoting the forecast and o the observational value. The brackets indicate that the bias is usually an average of a number of forecast-observation pairs. The information going into the bias is stripped of all probabilities, since only the ensemble mean is considered. Therefore the bias is merely able to give a rough picture of model performance, e.g. whether the model is consistently too warm/cold, thus enabling the detection of possible model drifts.

RPS, RPSS & RPSS_D

As a skill metric, the ranked probability score (RPS) can be calculated (Epstein 1969). The RPS is a generalization of the well-known Brier score (Brier 1950) for multicategory events.

$$RPS = \sum_{k=1}^{K} (Y_k - O_k)^2 = (\boldsymbol{Y} - \boldsymbol{O})^2 \tag{2.2}$$

K = Number of categories, Y_k = k^{th} component of the cumulative forecast vector $\boldsymbol{Y}$,
O_k = same for the observation.

The RPS measures the squared errors of the forecast with respect to the observation. The 'observation' is thereby treated as a categorical vector which assumes a value 1 for the category the observation falls into, and 0 otherwise. The RPS compares the cumulative density function (CDF) of a probabilistic forecast to the CDF of the corresponding observation over a given number of discrete probability categories (in this study: 3 equiprobable categories). The squared difference is calculated for each category (e.g. warm / normal / cold) and added up to yield the RPS. Using CDFs instead of probability density functions makes the measure 'sensitive to distance', i.e. a forecast that is far off the observed value is punished stronger than one that is closer. Another RPS characteristic worth mentioning is its strict propriety, that means the score encourages forecasting one's true beliefs. A proper score is optimized if the forecast corresponds to the best judgement of the forecaster, making hedging useless (e.g. Wilks 1995 or Murphy and Epstein 1967).

To get an impression of the mechanics of the RPS, a numeric example is given. The temperature scale has been divided into K=3 categories. Assume that of all 51 members, 21 fall in the cold, 24 in the middle and 6 in the warm tercile. The observed temperature finally lies in the middle tercile. The corresponding CDFs are ($\frac{7}{17}$, $\frac{15}{17}$, 1) for the forecast and (0, 1, 1) for the observation. Calculating the squared errors yields

$$RPS = (\frac{7}{17} - 0)^2 + (\frac{15}{17} - 1)^2 + (1 - 1)^2 = 0.183. \tag{2.3}$$

The lower the RPS value is, the better is the forecast. A perfect forecast results in a RPS of 0. The worst possible score (observation falls into one of the outer categories while the forecast is 100% in the other outside category) is K-1.

Usually the RPS is determined as the average over n forecast/observation pairs:

$$\langle \, RPS \, \rangle = \frac{1}{n} \sum_{i=1}^{n} RPS_i \tag{2.4}$$

Although the RPS technically assigns a number to every forecast/observation pair, it is not trivial to assess model skill with it. More preferable is a relative skill measure, i.e. a skill measure which is independent of the verification context applied (e.g. number of categories). A common solution is to compare the RPS of the forecasting system with the RPS obtained from a reference strategy. This yields the so-called ranked probability skill score (RPSS) which is defined as

$$RPSS = 1 - \frac{RPS}{RPS_{ref}} \tag{2.5}$$

and measures the degree to which the forecast outperforms the reference forecast. Climatology is often used as the reference strategy, i.e. a probability of $\frac{1}{3}$ is given to each category (if 3 equiprobable categories are considered). A perfect forecast (RPS = 0) always yields a RPSS of 1, one that is equally good as the reference always gives a value of 0 and one exhibiting less skill than the reference strategy always provides a negative RPSS value.

One important property of the RPSS is its intrinsic correction for systematic biases. If the quantiles determining the forecast categories are obtained on the basis of the model climatology, and if the quantiles determining the observation categories are obtained from the observed climatology, then this quantile mapping implicitly calibrates the model. Thus, the results can be regarded as not influenced by a possible bias.

However, there is one major flaw to the RPSS. As pointed out e.g. by Richardson (2001), the skill score is negatively biased for small ensembles. This characteristic is undesired since skill scores should reflect model skill and not ensemble size. The bias was further investigated by Muller et al. (2005) and Weigel et al. (2007b) and found to be due to the 'intrinsic unreliability' of the EPS. An analytical formula has been derived quantifying the bias in dependence of category probabilities and ensemble size, and a correction formula for the RPSS has been obtained. If three equi-probable categories are considered, a debiased version of the RPSS, the so-called RPSS_D, is given by

$$RPSS_D = 1 - \frac{\langle\, RPS \,\rangle}{\langle\, RPS_{ref} \,\rangle + D}. \tag{2.6}$$

$$with \qquad D = \frac{1}{M} \cdot \frac{K^2 - 1}{6K}. \tag{2.7}$$

M = ensemble size, K = number of categories.

The RPSS_D is almost independent of ensemble size, a property that is essential in our study. Using the traditional RPSS on the hindcasts with only 5 members would yield severely biased results. Now we are able to verify the full 12 years of hindcasts as well as the 51 member forecasts without having to account for systematic biases or ensemble size dependent skill.

2.2.2 Quantiles

To be able to define three climatologically equiprobable categories in the RPSS_D, their boundaries have to be determined from the climatology of each week of the annual cycle. Calculating the weekly terciles with only 13 years of data, i.e. with a sample size of 12 (due to cross-validation, see below) for each week, turns out to be statistically problematic. Therefore, we included the week before and after the relevant week and created a centered three week quantile with equal weights for all weeks. By applying this technique, the effective sample size is increased from 12 to 36, thus enhancing the robustness of the quantile estimates (Figure 2.1).

For the observations, this method is straight-forward. The forecasts, however, do not have a single value denoting the mean temperature during a given week, but a multitude (51 resp. 5) of such values. This facilitates the quantile determination because a lot more values are available. Nevertheless, for consistency reasons, three weekly averages are considered for the forecast quantiles, too. To take into account the varying ensemble size, the contribution of the individual years to the quantile calculation has been

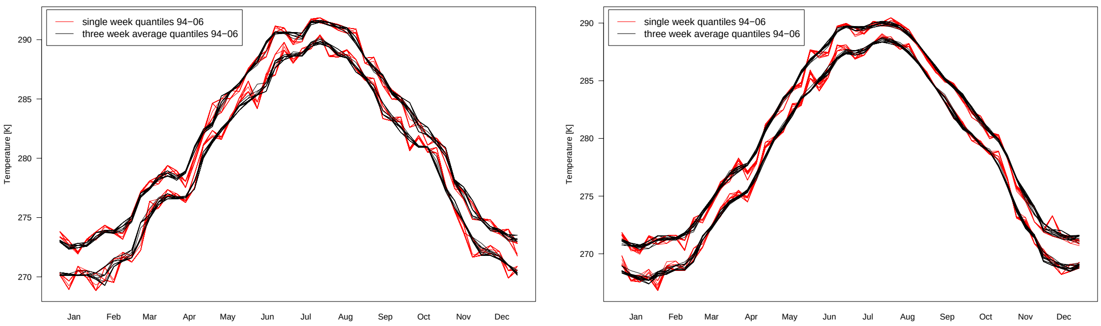

Figure 2.1: Seasonal cycle of the 33% and 66% quantiles from ERA-40 (left) and the monthly forecasts for week 1 (right) of the 1994-2006 period at a single grid point in northern Switzerland (8°E 47°N). Each line represents a time series of weekly quantile estimates for the period 94-06 in a leave-one-out cross-validation mode. Quantiles obtained from weekly averages (red) are fluctuating strongly, in contrast to quantiles obtained from 3-week averages (black).

weighted according to the respective ensemble size.

2.2.3 Cross-validation

Forecasts and observations are used both for the verification and for estimating the climatology. To guarantee that no information from a given forecast is used in the verification process of that very forecast, we apply the quantile calculation in a leave-one-out cross-validation mode. This means, for every year the quantiles are calculated separately, using information from all years except the one under consideration.

3 Results

3.1 Model drift

To get a first overview of the performance of the ECMWF monthly forecasting system, we examine the model's systematic bias. Figure 3.1 shows the annual mean bias of near-surface (2 meter) temperature for forecasting weeks 1 through 4. The model has developed a significant bias in many regions already in the first week (which corresponds to days 5-11). For parts of Greenland, South America, Southeast Asia and Antarctica the model is more than 1 K colder than the observations. Large patches in continental North America and the polar oceans show a warm anomaly of at least 1 K. Other areas of large differences between the model and the observations include many coastal regions (California, Chile, and Namibia). Away from the coastline, the oceans are generally too cold by 0-1 K depending on the location.

For weeks 2-4 the anomalies further amplify, while revealing the same spatial pattern as in week 1. In particular, the cold anomalies over Greenland are nearly doubled from week 1 to week 2, and over Antarctica a maximum bias of nearly 2 K is developed after 4 weeks. For the non-polar oceans, an increase of the negative bias from 0.3 to 0.5 K on average can be detected after the full integration time.

In order to characterize the seasonal dependence of the model bias, the annual cycle of observed and simulated T2 is illustrated in Figure 3.2, both globally averaged and for a single grid point in northern Switzerland. We start by discussing the left hand panel. The amplitude of the global seasonal cycle is 3 K with the maximum occurring during boreal summer. This asymmetry is due to the fact that landmasses cover a much bigger portion of the Northern Hemisphere (NH) than of the Southern Hemisphere (SH), combined with the different heat capacities of land and sea. The forecasts consistently underestimate the observed temperatures. Except for the leap in late August and November, which will be explained in the next paragraph, the negative bias is most pronounced during boreal summer (JJA, 0.5 K), while during the other seasons it assumes a fairly constant value of about 0.3 K. The results are consistent with the findings from the annual mean maps shown in Figure 3.1. A large first week error is followed by smaller drifts for the second week. Forecast weeks 3 and 4 are almost identical to forecast week 2, suggesting that the prediction system has assumed a state of quasi-equilibrium after two weeks.

In September, a significant reduction in negative bias down to a value of 0.1 K occurs with one week offset for each additional forecast week. The model mean temperatures

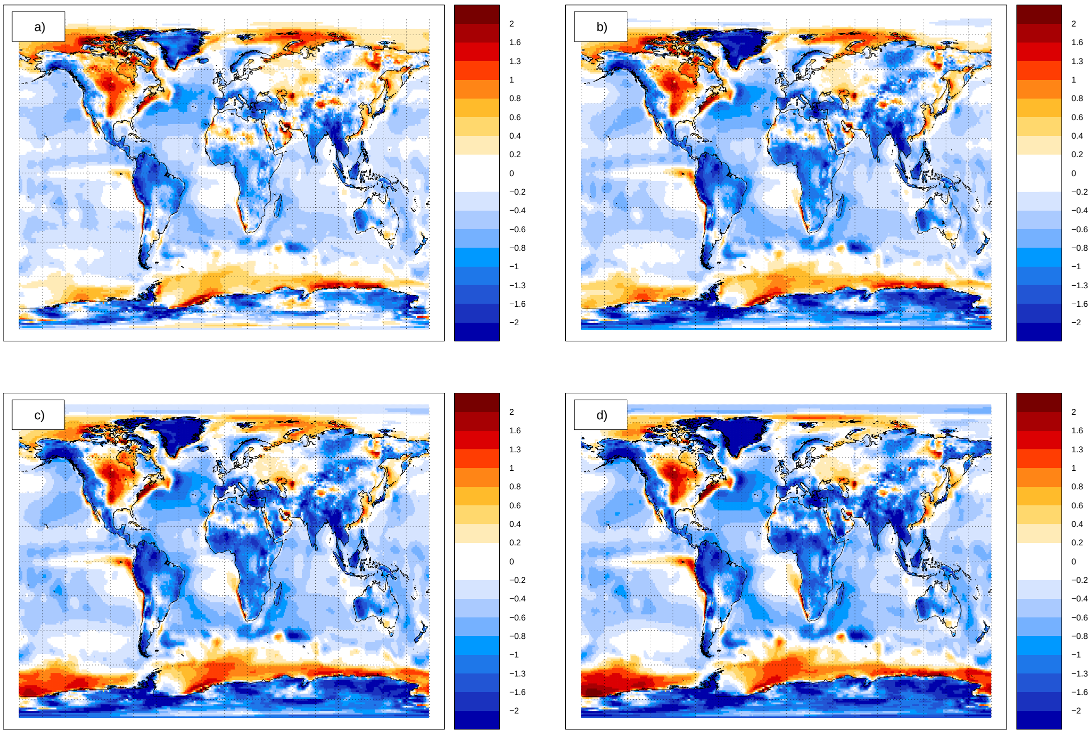

Figure 3.1: Annual mean bias (forecast−observation) of temperature [in K] for forecast weeks 1 to 4 (a-d). Data used is from all 13 years (forecasts + available hindcasts). Grid spacing is 20° in latitude and longitude.

remain in much better agreement with the observed values until end of November, when the bias jumps back to a value of ∼0.3 K. This strange feature can be attributed to a change in the model version, more precisely to the introduction of Cycle 31r1 in the ECMWF forecasting system. The staggered appearance of the improvement can be explained in the following way: Changing the system at a given time will not affect the week 4 forecast launched 2 weeks before the change. A shift observed for the week 1 forecast at a given time is expected to manifest itself in the week 4 forecast three weeks later. The unanimous drop in the end of November back to pre-September quality is due to the time-frame of the data used. More precisely: the forecast data considered range from December 2005 to November 2006. Thus, with the new system running only since September 2006, all values between December and September (including the corresponding hindcasts) are based on the older model version.

On a first look, it seems that the introduction of the new cycle has highly improved the model. To investigate what has changed in spatial terms we compare maps of the average bias for the period covered by the old model to the period with the new model (Figure 3.3). Although the averaging period is not the same, we assume that the dif-

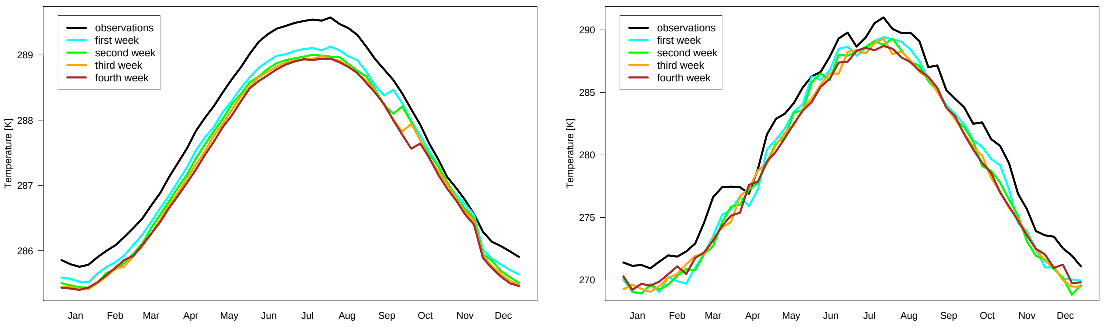

Figure 3.2: Seasonal cycle of observed and model mean weekly T2 for global average (left) and 8°E 47°N (right), which represents a grid point in northern Switzerland. Data covers the period from December 2005 to November 2006 including all hindcasts.

ferences found are due to the model version, because it appears unlikely that the bias should be substantially different in fall than during the rest of the year. The differences between model versions are small for most parts of the world. The main discrepancy is found in high latitudes (the Arctic Ocean, Siberia, the Southern Ocean and Antarctica), where a significant warming is detected. The newer model version seems to exhibit more regions of strong positive biases, but in a global mean (as in Figure 3.2) the intense warm anomalies over the polar regions cancel much of the cold bias found in the rest of the world. As far as the midlatitudes are concerned, the version upgrade of the forecasting system appears to have only minor effects on model climate.

The characteristics of the yearly temperature cycle of a single grid point in northern Switzerland (Figure 3.2, r.h.s) are similar to those found on the global scale. However, the curves are less smooth, because the small sample size of 13 years is not large enough to average out the interannual variability. The amplitude of the seasonal cycle amounts to 18 K which is in good agreement with climate diagrams for Zürich (Begert et al. 2005). The annual mean cold bias is roughly 1 K, which is in good agreement with the value for central Europe in Figure 3.1. The bias is largest (2 K) in winter and smallest (1 K) from April to June. There is no significant difference between the individual weeks. As mentioned above, the introduction of the new model version does not seem to have an impact on the local model bias in Switzerland.

3.2 Lead times, seasonal and regional dependence

In this section the skill of the monthly forecasts is examined with respect to lead times and seasonal cycle, and a closer look is provided for the European and North American domain.

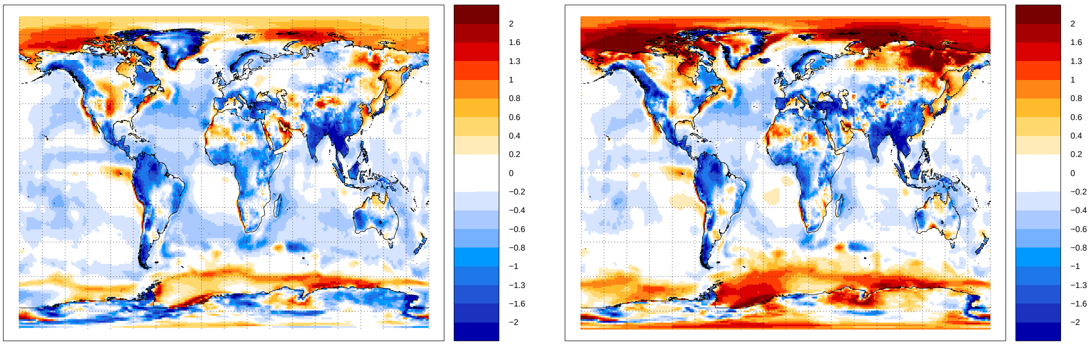

Figure 3.3: Mean bias of T2 [in K] for December to August (left) and October to November (right) for forecast week 1. Weeks 2-4 show similar patterns and are therefore not shown.

3.2.1 Annual mean skill

Figure 3.4 shows the average $RPSS_D$ for forecasting weeks 1 through 4. For the first week, the skill distribution closely resembles the skill of medium-range EPS, which is not surprising since a similar atmospheric model is employed on a similar timescale. Important characteristics of the distribution are generally high skill (>0.3) for almost the whole globe except the equatorial Atlantic, Indian, and west Pacific Ocean. In these regions the monthly forecasting system exhibits almost no skill. On a first look, this may appear surprising, because it is particularly the tropics which are known to reveal good longer-term predictability (see also below). Why should they have a reduced short-term predictability? This supposed paradox can be resolved by considering the small inter-annual climatological variability. In those regions, the 33% and 66% quantile are very close to each other, making a forecast more difficult. For example, the average tercile boundaries for a week in January in the central Atlantic lie at 299.50 K and 299.80 K. If the difference between a warm and a cold week is only 0.3 K, any categorical temperature forecast is error-prone.

The best forecast area is the central and eastern Pacific, the region of the El Niño Southern Oscillation (ENSO). The temperature regime in this area is dominated by the El Niño phenomenon, yielding large interannual differences in T2. Here, the average tercile boundaries for a week in January lie at 296.07 K and 297.87 K, respectively. The difference of 1.8 K tops the same measure in the Atlantic by a factor of 6. Essentially, there are very warm and very cold years, which causes a large spread in the terciles. Good predictability for El Niño events (Latif et al. 1998) determining the interannual variability and its persistent nature result in very high skill.

Not surprisingly, forecasts for the second week have less skill. Continental skill is generally much lower with an $RPSS_D$ of about 0.1. South America and Africa thereby reveal slightly better predictability than North America, Eurasia and Australia. The

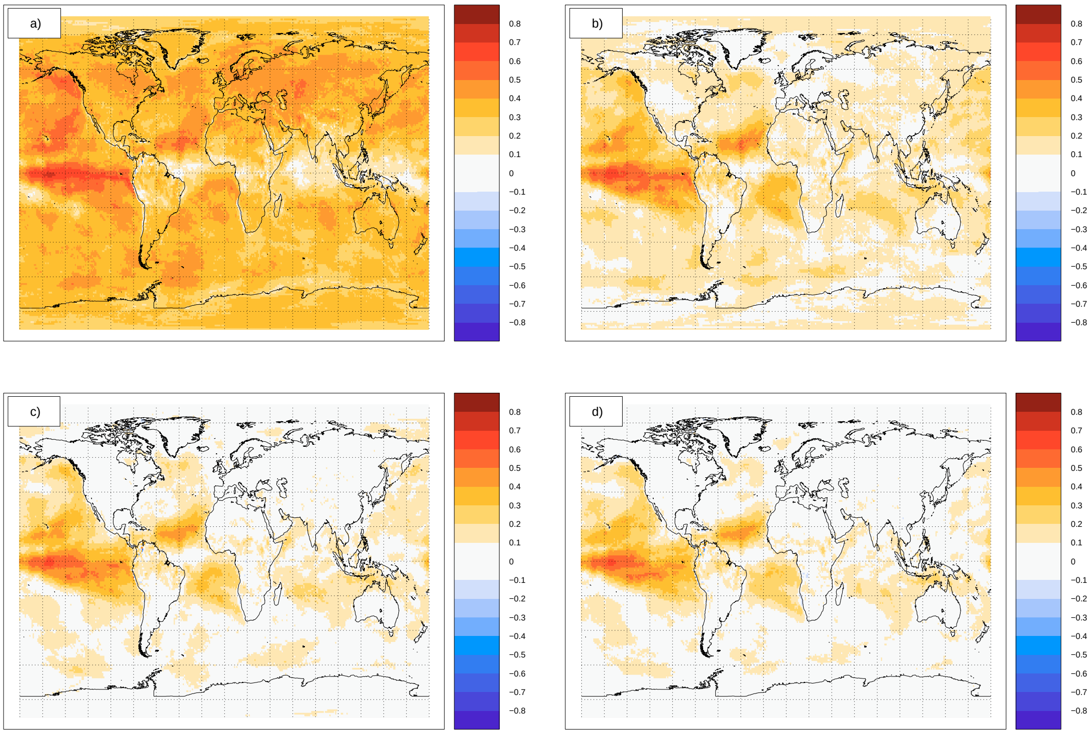

Figure 3.4: Annual mean RPSS$_D$ of monthly forecasts of T2 for forecast weeks 1 to 4 (a-d).

oceans show considerably more skill, especially in the trade wind regions on both sides of the equator, but the decrease to week 1 is large.

The third and fourth forecasting weeks are characterized by the loss of all skill on the continents except for the tropical region. The northern part of South America is clearly the best forecast land territory for this lead time. Over the oceans, the ENSO region retains its high predictability well into week 4. Generally, the pattern seen in week 2 remains, but the magnitude of skill is lower. It seems noteworthy that there are almost no places on the entire globe, where the monthly forecasts show significant negative skill even after a month of integration time. This indicates that after a certain time, which corresponds to about 15-20 days over the continents and about 30 days over extratropical oceans, the ensemble members have spread apart such that the forecast can be considered a climatological forecast with similar probabilities in all terciles.

Motivated from the large skill drop between forecasting weeks 1 and 2, the next set of plots (Figure 3.5) illustrates this drop in absolute (l.h.s.) and relative (r.h.s.) terms. In absolute numbers, the reduction in skill has its maximum in the center of the Eurasian continent with an RPSS$_D$ skill loss of more than 0.4. The skill drop is also very pronounced (0.3-0.4) for most parts of Europe and North America. Generally, the NH is affected more strongly than the SH. Most of the extratropical oceans loose 0.1-0.3 of skill. Throughout the tropics, the drop in skill is negligibly small.

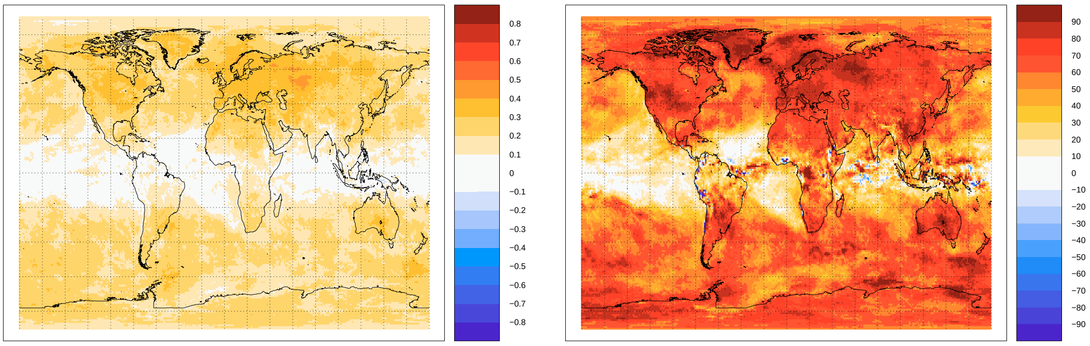

Figure 3.5: Drop of annual mean skill from week 1 to week 2 in absolute (left) and relative (right) numbers. Absolute drop is in RPSS_D values, relative in percent change.

A consistent picture is revealed by the relative loss of skill: In almost all non-tropical continental regions, skill drops down by 60-100%. Among others, northern Europe, northwestern USA or central Australia loose virtually all skill they had in week 1. The values in the vicinity of the equatorial Atlantic, Indian and west Pacific Ocean must be considered with care because skill is very low from the beginning in these regions. The definition of a relative drop is therefore problematic and may lead to misleading results. For some regions, most notably the North American west coast, there is a very distinct difference between land and sea grid points. While the skill decreases by about 40% in the Northern Pacific, the drop is at least 70% on the mainland, which is consistent with the skill maps shown in Figure 3.4. There is almost no difference between skill in the first and second week in the ENSO region and the central Atlantic.

3.2.2 Seasonal dependence of skill

So far we have focused on average model performance. It is known from seasonal and medium-range forecasting systems that the seasonal variability of weather and climate patterns may result in a seasonal variability of skill. We therefore proceed by examining the annual cycle of skill for different regions. Figure 3.6 displays the annual cycle of forecast skill for the northern and southern extratropics and the tropics. The time series are smoothed by applying a circular (1,2,3,2,1) convolution filter. This means that the new value in a given week is calculated by weighting that very week by a factor of 3, the two adjacent weeks by 2 and their respective adjoining weeks by 1. First of all, we recognize that the large loss of skill between the first and second forecasting week (described above) is independent of the time of the year. This drop is less distinct in the tropics, mainly due to the influence of El Niño. The difference between the weeks 2,

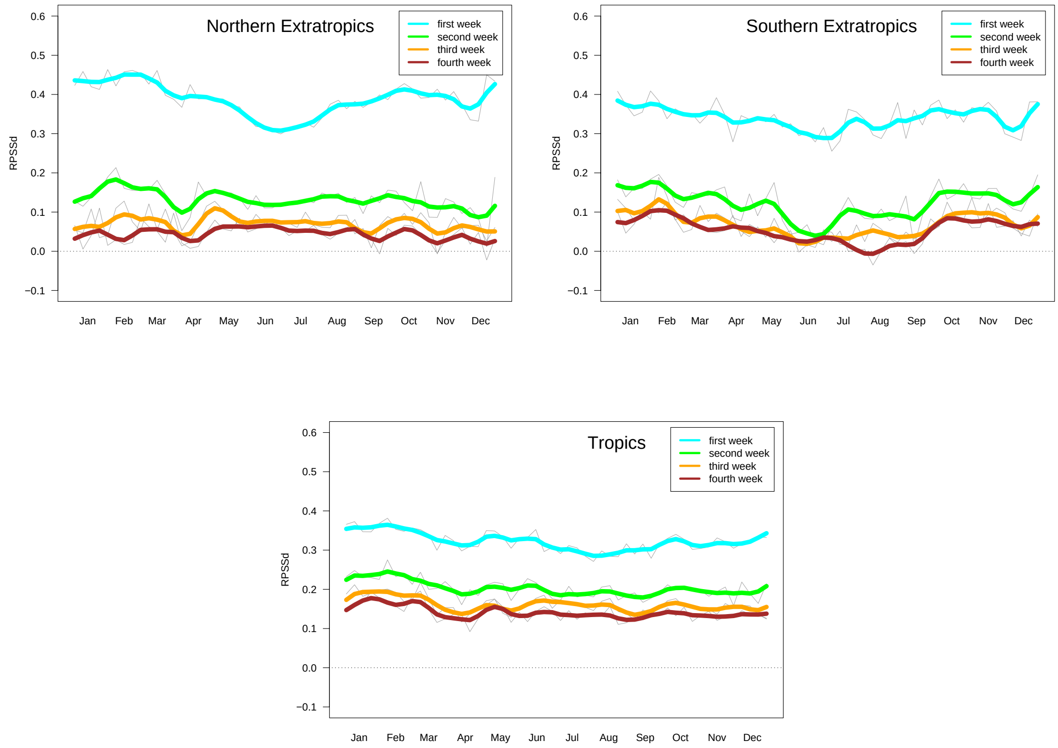

Figure 3.6: Smoothed annual cycle of average skill in the northern extratropics (30°-60°N), the southern extratropics (30°-60°S), and the tropics (30°S-30°N) for forecast weeks 1 to 4. Unfiltered curves in gray. A (1,2,3,2,1) convolution filter is applied. Grid points on land and sea are equally considered.

3 and 4 is comparatively small throughout the year. In the northern extratropics, one week forecasts in summer are less skilled than in the other seasons. There is no pronounced annual cycle apparent in weeks 2-4. However, at least in the second forecasting week, the model appears to perform slightly better in late winter (February, March) as compared to the rest of the year. The physical reason for the enhanced predictability in winter could be persistent blocking situations (Scherrer et al. 2006, Croci-Maspoli et al. 2007). These large-scale flow patterns can stay in place for days up to weeks. It is possible that in times of persistent blockings (as is the case in winter on the NH) the skill is enhanced. In the SH, week 1 forecasts show a less distinct seasonal cycle than in the NH, but still reveal a slight maximum in DJF. This is surprising, since in terms of local seasons, this is exactly opposite to the observation made in the NH, because DJF corresponds to summer in the SH. For forecast weeks 2-4 this pattern is more pronounced with particularly little skill in wintertime (JJA).

Within the tropics (Figure 3.6, center), forecast skill is constant throughout the entire year. This is plausible since it represents an average over an area centered around the

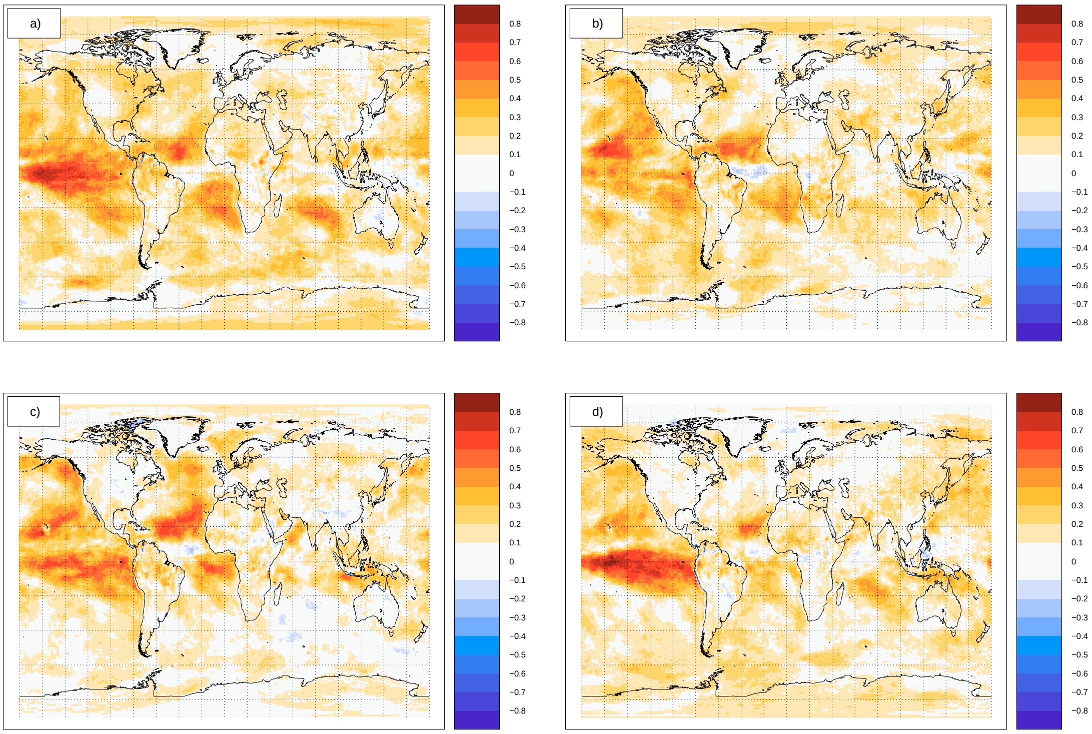

Figure 3.7: $RPSS_D$ skill of week 2 forecasts in seasonal resolution. Winter (a), spring (b), summer (c), and fall (d) are shown.

equator. The skill in this region does not degrade as quickly as in the extratropics and seems to saturate at 0.15 i.e. at 15% improvement w.r.t. the climatology.

Another possibility to display seasonal influences are maps which show skill averages of selected seasons. This technique helps to identify spatial differences in the course of the year. Figure 3.7 shows the model skill for the second forecast week in seasonal resolution. Weeks 3 and 4 are qualitatively very similar and therefore not shown but can be found in the Appendix. The most noticeable feature is the seasonal dependence of skill in the central and eastern Pacific. Skill is extraordinarily high (~0.7) during fall and winter, but falls to average oceanic values in spring before recovering slightly in summer. This corresponds well to the known timeline of an El Niño event: SST anomalies onset in summer, increase in fall and peak in winter. In spring the system goes back to its normal state and displays low interannual variability. Another interesting feature can be discovered in the central Indian Ocean, where skill also very strongly depends on the season. In JJA, the area shows slightly negative skill. In DJF, however, the area shows substantial skill of about 0.5. This oscillation is in phase with ENSO and could well be connected to it.

To simultaneously evaluate the latitudinal and seasonal dependence of forecast skill while removing the impact of local topographic features, we consider zonal mean plots as shown in Figure 3.8. The plot shows zonally averaged $RPSS_D$ values for different lead-times and seasons for only land, only sea, and all grid points. Poleward of 60° North and South, there are only few independent grid points due to the convergence of the circles of longitude. The results in these high latitudes should therefore be considered with care. The 12 plots consolidate some findings we have already made. Seasonal dependence of forecast quality is strongest in the NH and the first forecasting week, where skill in summer is inferior to the other seasons. Skill over land drops quickly and seems to saturate at a $RPSS_D$ value of 0 in the extratropics in week 3 to 4. Over sea, skill is generally higher and decreases slower. The integrated land and sea plot contains characteristics of both. What emerges in all columns is that the region of least skill in week 1 forecasts (the tropics) becomes the best predicted area by week 3 and even more pronounced in week 4.

Also new aspects of the skill distribution can be found, though. In the tropical region of the land-only plots for week 3 and 4, the latitude of maximum model performance seems to oscillate around the equator in a regular manner. Enhanced skill might be related to the migration of the intertropical convergence zone (ITCZ), but there is a shift of 2-3 months between the migration of the ITCZ and the oscillation seen in our data. For example, while in boreal summer the maximum skill is at 5°S, the ITCZ is located in the NH. Given that this phenomenon is only observed over land, and given that in these latitudes continental skill is mainly given by South America, this oscillation my well be a local feature. Another particularity involves the two peaks in week 4 of the sea-only zonal mean at approximately 60° North and South. The increase in skill in these latitudes in the respective summertime is very distinct. This anomaly is not observed over land at all and might be due to sea ice effects, since it matches the region where the annual variation of sea ice cover is largest (at least in the SH). This feature can also be found in the seasonal maps (Figure 3.7) of the second week as a band of enhanced skill and is even more marked in the week 4 seasonal maps (cf. Appendix).

3.2.3 The European and North American domain

The above analysis has been carried out globally. Given the particular importance of Europe and North America to users, we now evaluate these regions more in detail. Figure 3.9 presents the annual mean skill for the European domain. For the first forecasting week, skill reveals a gradient from southwest to northeast, increasing from 0.3 to 0.5. Superimposed on this pattern is a patch of reduced skill over the British Isles and one of increased skill over central Spain. In the annual mean, only few regions in Europe (notably the Mediterranean and the North Sea) exhibit significant skill values in the second week. These areas of skill all lie over or close to the sea, and it is likely that the persistence of sea surface temperatures contributes to the predictability in these places. After 20 days of integration time, virtually all skill is lost.

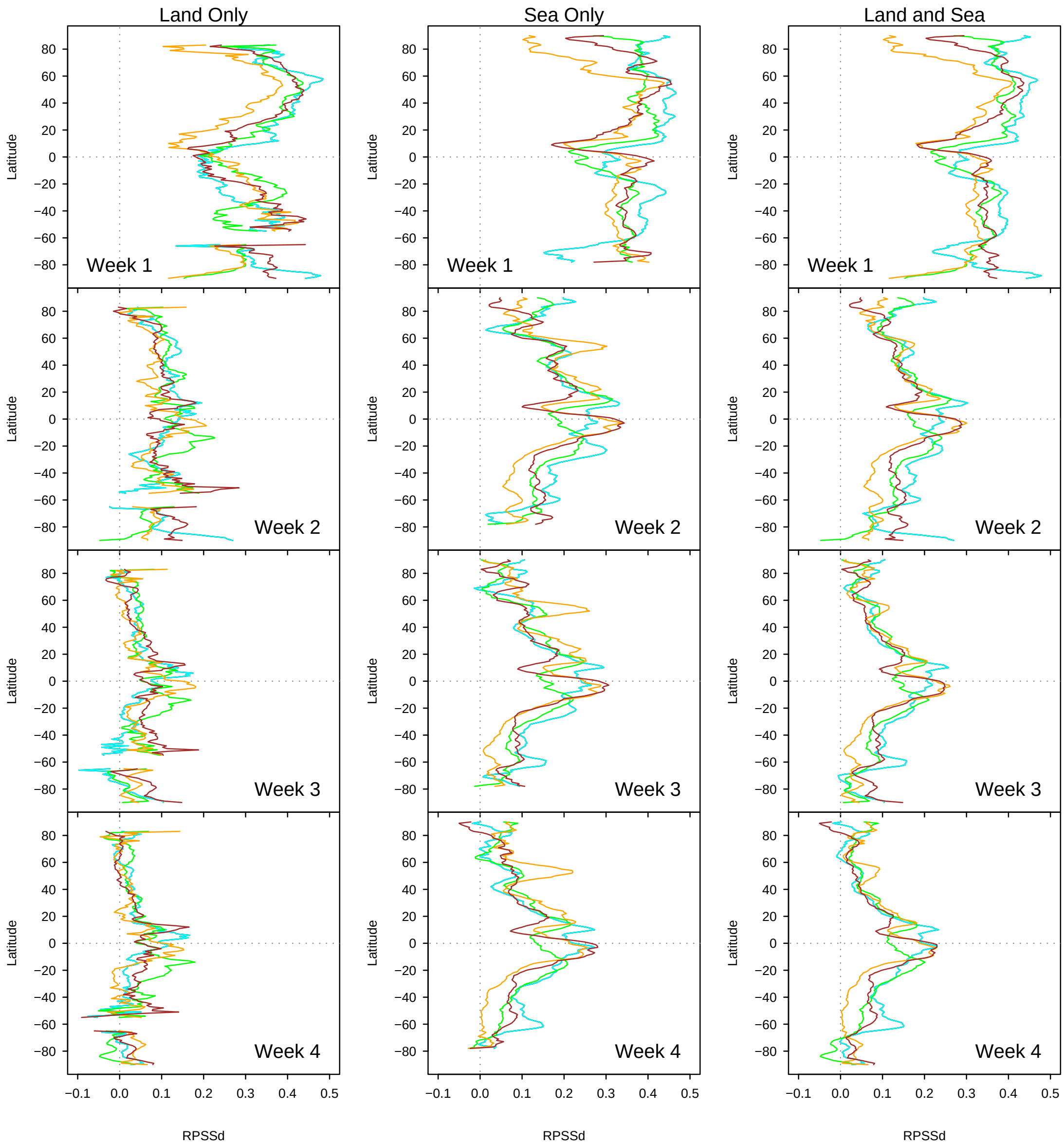

Figure 3.8: Zonal mean $RPSS_D$ skill for forecast weeks 1 to 4 for land (left), sea (center), and all grid points (right). The colors are chosen to represent boreal seasons, even though this might be misleading in the SH: spring (green, MAM), summer (orange, JJA), fall (red, SON), winter (blue, DJF). In the plots featuring land-only points, the absence of data around 60°S and poleward of 80°N is due to the absence of land on these latitudes. Similar, there is no sea-only data from 80°S poleward.

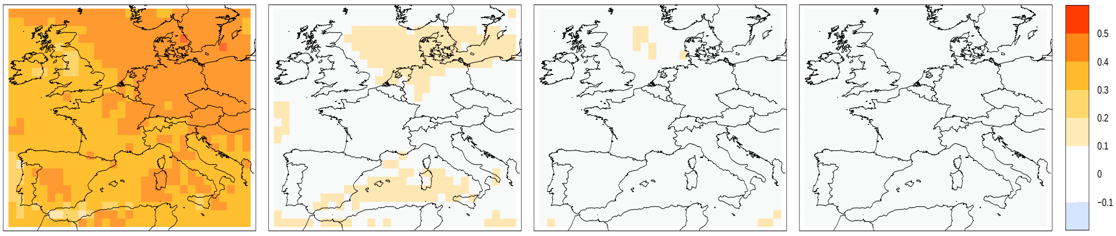

Figure 3.9: Annual mean RPSS_D of monthly forecasts for weeks 1 to 4 (left to right) in the European domain.

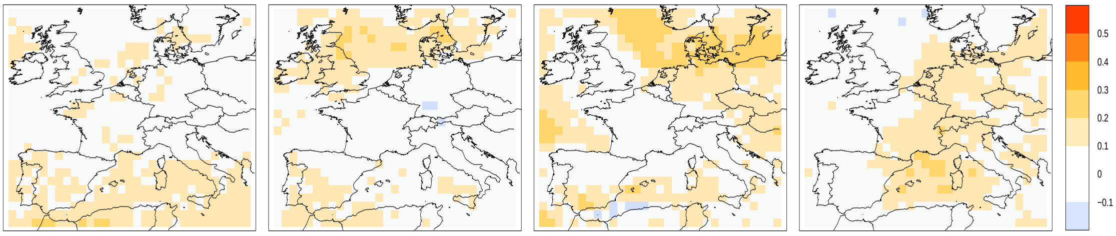

Figure 3.10: RPSS_D skill of week 2 forecasts as seasonal averages in the European domain. Winter, spring, summer and fall are shown (from left to right).

Seasonally resolved skill maps of the individual seasons for the second forecasting week also display interesting features in Europe (Figure 3.10). First, forecasts in the Mediterranean and southern Europe are more skillful in fall and winter than in spring and summer. In the North Sea, exactly the opposite is found: There is higher skill in summer and spring compared to the other seasons. Reasons for this pattern could be the thermal memory of the heated Mediterranean which contributes to skill in fall and winter. Maybe small interannual variability similar as found in the tropics hinders skillful forecasts in summer. In northeastern Europe snow feedbacks (as described in the Introduction, or Shongwe et al. 2007) could play an important role for predictability in this region. In the Alpine region, an annual cycle can also be detected. Slightly negative skill in spring is followed by very low skill summer. In fall relatively high skill of up to 0.2 can be found before the skill matches the climatology again in winter. Forecasts in the third and fourth forecasting week exhibit very low skill generally and are therefore not shown (cf. Appendix).

These findings are refined with the consideration of the annual cycle of skill averaged over all land points of that domain (Figure 3.11, l.h.s.). Skill remains significant ($\sim$0.1) in wintertime for the second and third forecast week. Equally, the good predictability in September and October carries well into the third forecasting week. For week 4, however, there is no time of the year when skill in Europe reaches at least 0.1 in two consecutive weeks.

As a second region of special interest, we now consider the North American domain. The first week in annual average (Figure 3.12) is characterized by high skill (0.4-0.5)

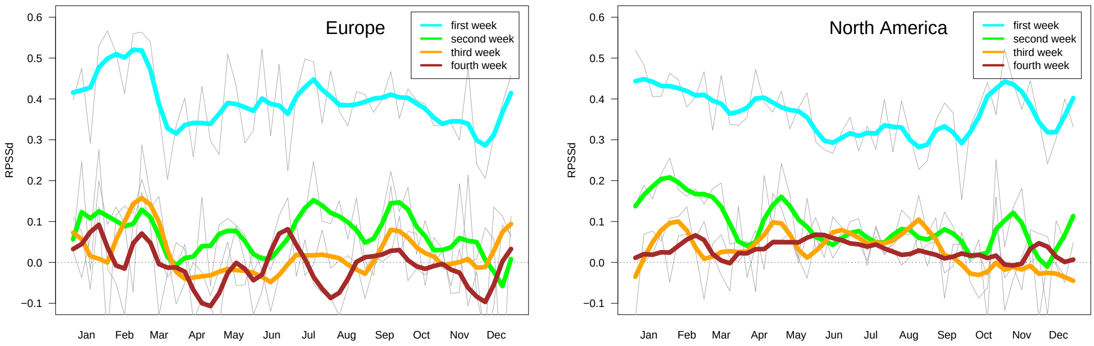

Figure 3.11: Smoothed annual cycle of average skill in the European (left) and North American (right) domain for forecast weeks 1 to 4. Unfiltered curves in gray. A (1,2,3,2,1) convolution filter is applied (see section 3.2.2 for description). Only land points are considered.

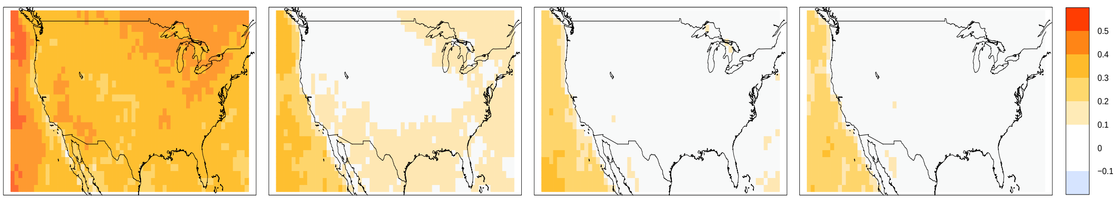

Figure 3.12: As Figure 3.9, but for the North American domain.

in the northeastern and the southwestern part of the continent as well as the Pacific Ocean. Relatively low skill is found in central USA as well as on the west coast. In the second forecasting week, only the east coast, the South and the Northeast still exhibit significant skill. By the third week, forecasts in the North American domain do not perform better than a climatological forecast, except for the Pacific Ocean where high predictability is expected.

Similar to results in Europe, strong seasonal variations in skill can be found in North America (Figure 3.13). The east coast, for example, shows a very pronounced skill signal only in winter. During spring, many parts of the US, with the exception of the Southeast and the Northwest, show substantial skill. In an annual cycle plot (Figure 3.11, r.h.s.), smoothed weekly values of skill in the individual weeks are shown. Also in this domain, winter is clearly the most skillful season in the second week. In contrast to Europe, the smoothed skill curves go almost never below 0 although in absolute terms they are very low. The difference between week 3 and week 4 seems random which means that there is no significant change in skill over the last 10 days of the forecast.

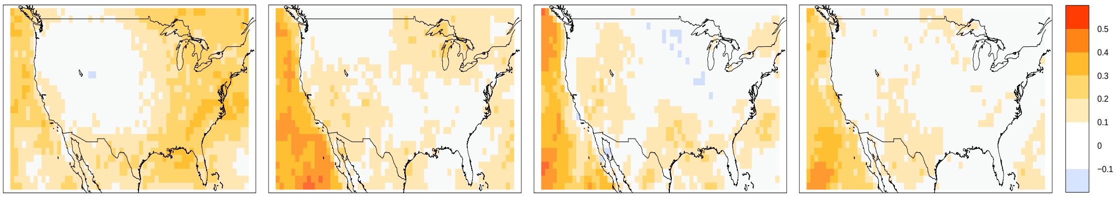

Figure 3.13: As Figure 3.10, but for the North American domain.

3.3 The impact of persistence

So far, the monthly forecasting system has been tested against a "perfect" climatology. Another common reference strategy is persistence. In this section, the value of persistence forecasts is analyzed. They have the benefit of not needing many resources since they are independent of time. Particularly when skill is comparatively low, it is of interest to know whether a certain quality of forecast could just as well be achieved by a simple persistence forecast. We apply a forecast-persistence approach, that is the first week probability forecast is persisted and used to predict the second, third and fourth week. This corresponds closely to persisting the medium-range EPS a month into the future. The persistence forecasts are verified with the same scheme as standard monthly forecasts using the RPSS_D with climatology as the reference strategy.

The annually averaged results are shown in Figure 3.14. Note that the first panel, i.e. the first forecasting week is not yet a persistence forecast; it is equivalent to the normal forecast shown in Figure 3.4 a. Already in the second forecast week, persistence fails to outperform climatology on all continental landmasses except for equatorial South America. Over sea, persistence forecasts have only skill in the ENSO region and parts of the Atlantic. These are the regions where the monthly forecasting system equally performs well (cf. Figure 3.4). In the third and fourth week, these characteristics become even more pronounced. We find negative skill on the continents and also over most parts of the oceans. Only the primary El Niño areas still display positive skill on these timescales. The negative values over the continents imply that the persistence forecasts are overconfident. While the uncertainties grow with increasing forecast time, ensemble spread remains narrow. This means, the ensemble forecasts, while being sharp, are likely to be centered at the wrong value. Overconfidence is punished heavily by the RPSS_D, resulting in negative skill (Weigel et al. 2007a).

Finally, we quantify the degree to which the monthly forecasts outperform persistence. For this purpose, Figure 3.15 shows the difference in annual skill of monthly and persistence forecasts. Again, the evaluation will start with the second week, since in the first week the monthly forecast and persistence are identical. In week 2, monthly forecasts outperform persistence forecasts already significantly worldwide apart from the tropics. By week 4, only small patches in the Pacific and Atlantic Ocean are equally well forecast with both techniques. Thus, while monthly forecasts have very little skill four weeks in advance, they nevertheless outperform persistence because persistence yields strongly negative skill values in that time-frame, as has been discussed above.

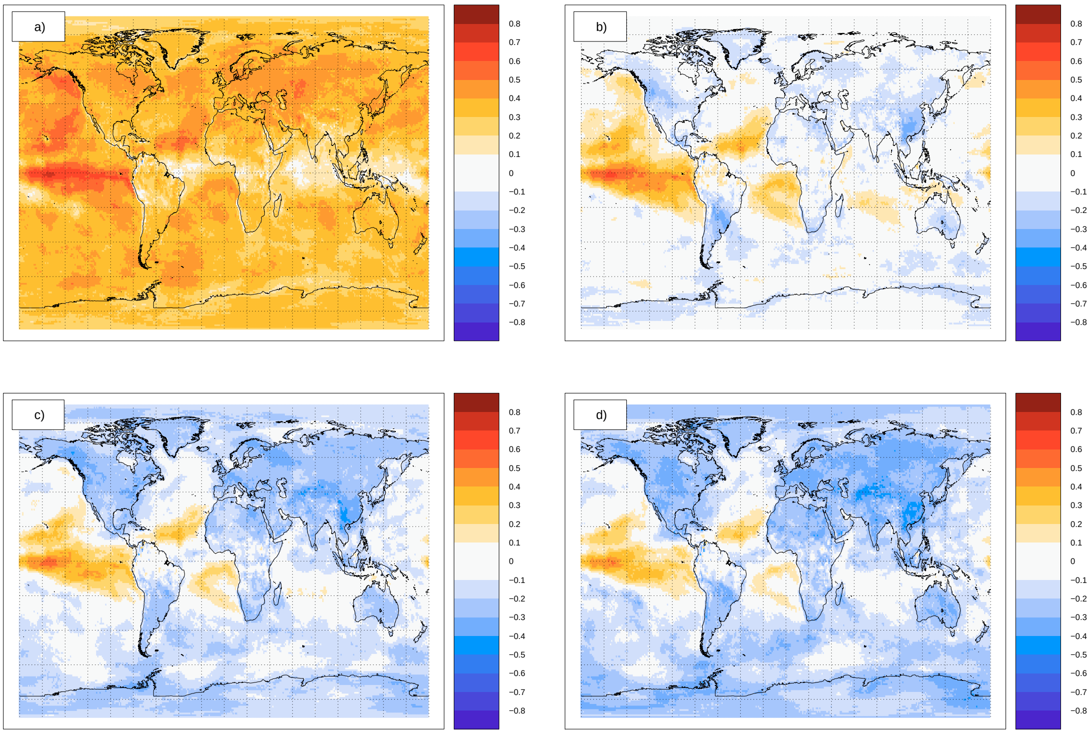

Figure 3.14: Annual mean $RPSS_D$ of first week persistence forecasts for weeks 1 to 4 (a-d).

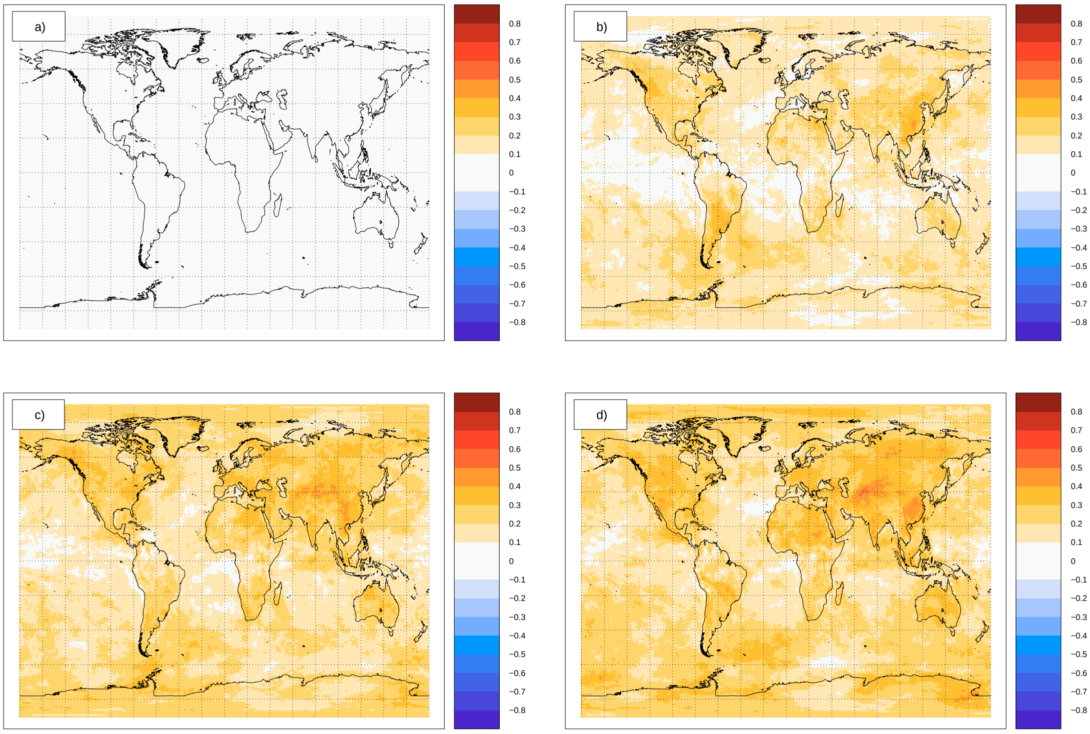

Figure 3.15: Annual mean RPSS_D of monthly forecasts minus annual mean RPSS_D of persistence forecasts for weeks 1 to 4 (a-d).

4 Concluding Discussion and Outlook

In this study, a systematic, probabilistic verification of the ECMWF monthly forecasting system for 2m temperature has been provided. From the results presented in the previous chapter, the following conclusions can be drawn.

4.1 Performance of the ECMWF monthly forecasting system

The bias of the model is unexpectedly high in some regions and there are substantial biases in absolute temperatures already after week 1. Thus, the absolute values provided by the monthly forecasting system should be treated with caution. While the probabilistic forecasting procedure implicitly corrects for such biases by mapping the terciles of the model and the observed climatology separately, the magnitude of the anomalies nevertheless raises the question of their origin. Potential error sources may lie in the model itself, the verification dataset or in incompatibilities between the reanalysis and the forecasting system. However, given that the bias changed in a coherent manner upon introduction of a new model version in September 2006, it seems likely that the systematic bias is a characteristic of the forecasting system itself. Further investigation of the reasons for this bias is subject to future research. A comparison to the systematic errors of the atmospheric circulation in the ECMWF forecasting system presented by Jung (2005) for geopotential height shows that the spatial structures observed in geopotential height and surface temperature seem not to be correlated. He notes, however, that error growth in the north Pacific saturates beyond forecasting day 20. This characteristic is also observed in our analysis of absolute model temperatures.

The probabilistic monthly forecasts have been shown to generally outperform climatological and persistence strategies. Their skill depends highly on regions and seasons. There is a substantial loss of skill from week 1 to week 2 especially over continents. In week 2, skill on most continents reveals seasonal fluctuations, which is likely due to the enhanced predictability of seasonally varying dynamical regimes, such as persistent blockings in winter, and the presence or absence of dominant surface boundary conditions such as snow cover. On the oceans, forecasts remain skillful longer than over land, because the boundary condition provided by ocean surfaces have a comparatively high degree of predictability due to the relative inertia and high heat capacity of the large water bodies. Indeed, it has been shown that those areas where the system exhibits high skill into the fourth forecasting week benefit from steady boundary conditions such as particularly high sea surface temperatures during an El Niño event. While it is also in these regions, i.e. the eastern Pacific, that persistence forecasts perform well, the monthly forecasts still achieve a slightly higher score, justifying their use.

While persistent boundary conditions surly is a key component of monthly predictability, it must be strongly assumed that other sources of predictability, such as stratosphere-troposphere interaction and the Madden-Julian Oscillation mentioned in the introduction, might as well play a role. However, it was beyond the scope of this study to quantify their impact. The effects of the dynamic ocean scheme as investigated by Jung and Vitart (2006) for short-range and medium-range weather forecasting cannot be assessed as no similar forecasts without an interactive ocean were analyzed in this study.

4.2 Usefulness of monthly forecasts

The central question from a user perspective is, to what extent are monthly forecasts valuable and to be preferred to simple forecast strategies such as climatology or persistence. The applicability of monthly forecasts depends on the region and timescale of interest. As already stated above, generally good predictability is found over the oceans, where many regions are well predicted even four weeks into the future, particularly in the central and eastern Pacific and in the trade wind zones of the Atlantic. On the other hand, there are some places, e.g. the tropical Atlantic and Indian Ocean, where categorical temperature forecasts are of reduced value since interannual variability is very small and the terciles differ only a few tenths of a degree; indeed, in these regions the system performs no better than a climatological forecast. Of all landmasses, South America exhibits the highest skill, especially in the proximity of the equator, where forecasts are still skillful after 4 weeks of integration time. Also in tropical Africa, monthly forecasts show considerable skill in this range. On the other hand, over most of the extratropical landmasses, week 2 forecasts have low, week 3 forecasts almost no skill. However, if only certain seasons are considered rather than the entire year, also some areas of poor predictability may reveal some prediction skill. For example, in Europe, the system performs well over the British Isles in spring, and in eastern Europe in summer. Central Europe and the Mediterranean are best forecast in fall while in winter significant skill is only found in the Mediterranean Sea and northern Africa. Week 3 forecasts appear not to possess substantial skill in the European domain. The performance of monthly forecasts in North America is comparable to the situation found in Europe. While week 2 forecasts exhibit skill in some seasons and some regions, forecasts of the third week produce results that feature only a marginal gain in comparison to a climatological forecast.

These results are in good agreement with a verification study by Vitart (2004). He also concludes that monthly forecasts generally outperform persistence and climatology and further attributes little skill to the European domain in forecast week 2. However, he finds useful skill after 20 days over North America, Asia and the southern extratropics, which contrasts with the results from this study, where skill in these regions seems to be low after 2 weeks, at least in the annual mean. The reason for this discrepancy might be the fact that Vitart (2004) evaluated only the forecasts without the corresponding hindcasts. His sample size is therefore (a) smaller and (b) likely biased due to an El Niño event taking place during his verification period. The latter would probably boost the skill over North America, explaining some of the differences.

Another verification study relevant in this context was conducted by Hamill et al. (2004). They carry out a probabilistic verification of week 1 and 2 temperature forecasts produced by a NCEP, medium-range, low resolution forecast model in the North American domain. The RPSS skill is slightly negative already for the second week, which differs from our results and is likely due to the low resolution of their model. However, they show that the skill of their forecasts can be greatly improved by applying a model output characteristics (MOS) approach to the forecasting system.

All in all, the results show that monthly forecasts have valuable skill in predictions up to forecast week 2 (i.e. 18 days). Beyond that, skill is almost exclusively found over the oceans. However, given that the monthly predictions are never worse than climatology or persistence forecasts, it is advisable to use them on the entire globe. In the best case, they provide valuable information, in the worst case, they are just as good as a climatological forecast.

4.3 Outlook

Based on the findings of this study, several recommendations can be made for future research:

First of all, a more detailed analysis of the skill's lead time dependence is necessary. Using 1, 3, or 5 day averages, the question of whether the skill drops sharply or continuously could be addressed.

Secondly, the reasons of the regional and seasonal fluctuations discussed in this study should be investigated more in detail. For example, why are forecasts in fall significantly better than in other seasons in central Europe? What flow patterns are typical for high respectively low skill? Furthermore, to what extent do stratospheric-tropospheric exchange and the Madden-Julian Oscillation contribute to the skill in the extratropics?

Thirdly, the focus of the verification has been on 2m temperature only. An important scientific and applied approach would be to assess the system's ability in predicting derived quantities, such as heat waves and cold spells, several weeks in advance. If monthly forecasts have skill in such situations, they might be used as an early warning system and thereby help the authorities and decision makers in the affected region. Also of interest is the performance of the monthly forecasts for different parameters. In some parts of the world (mainly the tropics) other parameters such as precipitation are of equal or greater importance as temperature. It is quite possible that monthly forecasts for precipitation show substantial skill also in regions where temperature forecasts seem to perform mediocre.

In this study, the monthly forecasts are verified against ERA-40 and the ECMWF operational analysis. Reanalysis data is very convenient for verifying model output because it is available on a global, regular grid and is consistent in time. Yet, it is essentially model-based data which only provide an estimate of reality. The next step would be a similar verification against real observations, e.g. station data, similar to what Kunz et al. (2007) have provided for the ERA-40 dataset.

A further motivation not investigated in this study is the impact of ocean-atmosphere coupling on medium-range forecasts. In the first 10 days, the main difference between conventional medium-range and monthly predictions is that the latter apply a full 3-D ocean model while the former do not. A comparison between the two might improve the understanding of the influence of such a coupling in the medium-range which could be important in cyclone track forecasts, for instance.

On the other end of monthly predictions, they could also help in getting a handle on the first month in seasonal predictions. By comparing the two during the first month, the impact of the atmospheric resolution on seasonal forecasting could be evaluated. Besides, monthly forecasts allow for a more detailed look at the model evolution during the initial weeks in seasonal forecasts, which is where a significant part of the error develops.

Although the bias is of minor importance in this categorical verification, it could influence physical processes in the model due to its magnitude. In order to improve the performance of the monthly forecasting system, efforts should be undertaken to reduce the bias of the system. Also, temperature trends should be examined. Possible trends in the time series could affect the tercile determination and therefore the skill of the monthly forecasts as investigated by Liniger et al. (2007) for seasonal forecasts, respectively.

Using only 12 years of hindcasts to determine the weekly terciles appears not very robust. In regions of large interannual variability (extratropics) 12 values is a rather small sample size. The forecasts would surly benefit from an extended hindcast period. An alternative approach to enhance the robustness of the tercile determination without using more computing time would be to apply a centered three week mean. This technique (cf. section 2.2.2) uses the hindcasts of the week before and after a given week to blow up the sample size. Obviously, this is not possible with the current monthly forecasting setup, since the hindcasts starting a week after a given forecast have not been computed yet. However, since the initial values of these 'next week hindcasts' are available (ERA-40 and operational analysis), one could imagine to always calculate the terciles of the coming week instead of those that start that very week. This simple change in the configuration of the monthly forecasting system could increase the robustness of the tercile determination and therefore improve monthly forecasts in general.

Finally, it should be remembered that the RPSS_D is only one of many measures of skill. Each measure has its own characteristics and it is not trivial to assess to what degree the results are influenced by the skill score's aptness for a given setting. A study based on a different skill score, such as e.g. anomaly correlation, could come to different conclusions about the usefulness of monthly forecasts.

Acknowledgements

I thank Andreas Weigel and Mark Liniger for the supervision, their kind help and support. I also thank Paul della Marta, Hubert Mathis and the rest of the NCCR team at MeteoSwiss for helpful comments and their comradeship.

Further, I thank Christof Appenzeller for directing this thesis.

Appendix

Global maps of RPSS_D of monthly forecasts for individual seasons

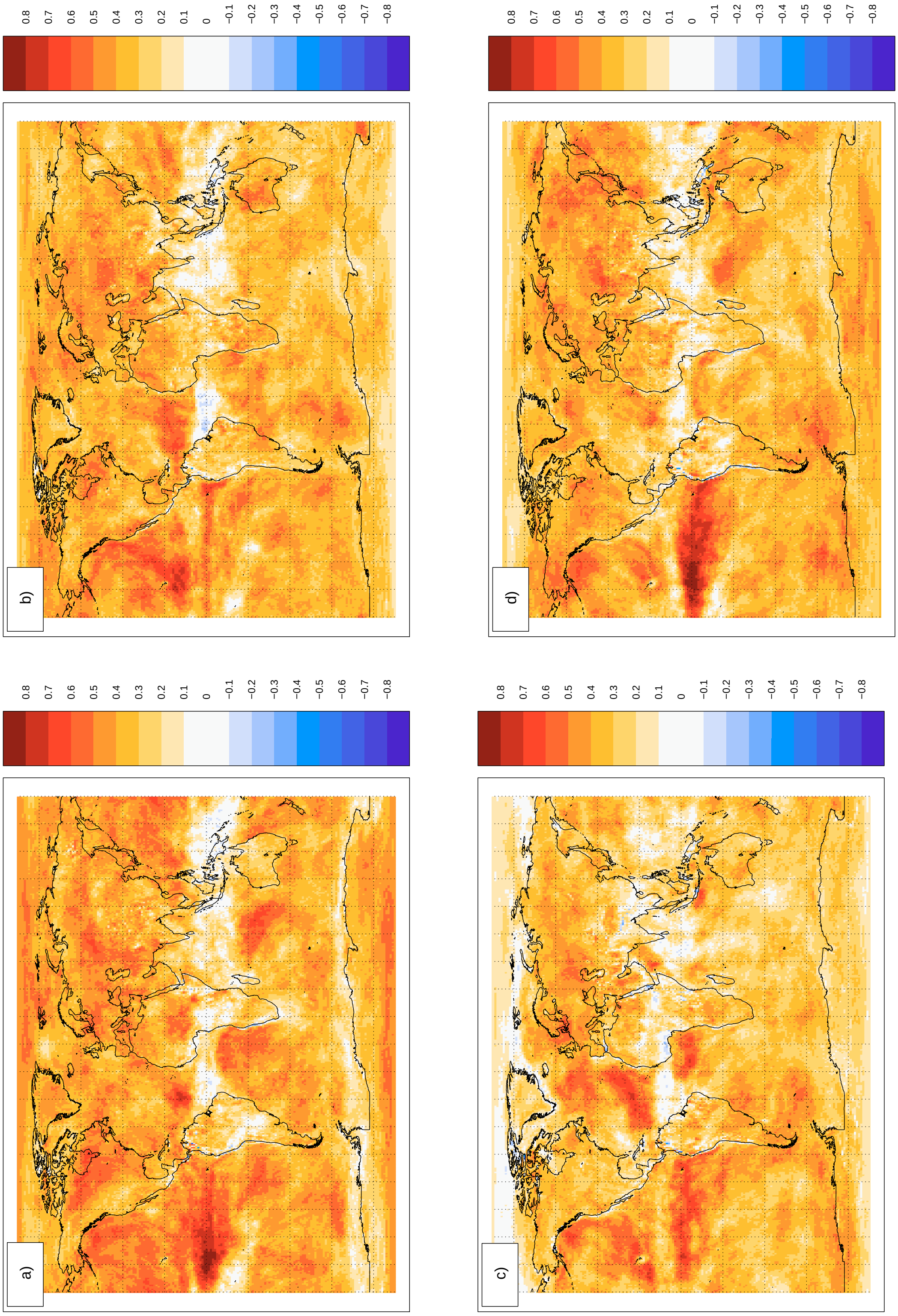

Figure App.1: Average RPSS$_D$ skill of week 1 forecasts in seasonal resolution. Winter (a), spring (b), summer (c), and fall (d) are shown.

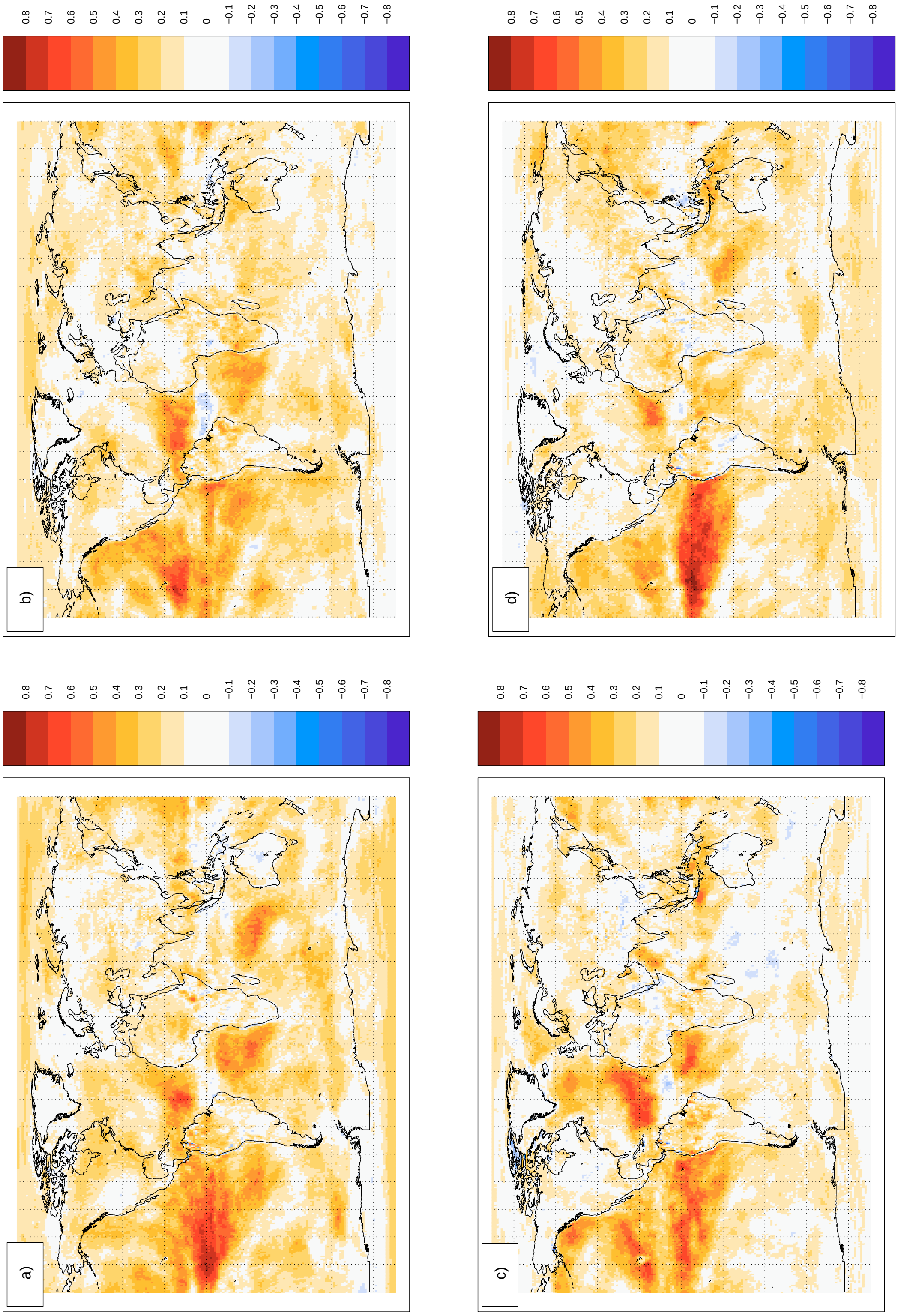

Figure App.2: Average RPSS_D skill of week 2 forecasts in seasonal resolution. Winter (a), spring (b), summer (c), and fall (d) are shown.

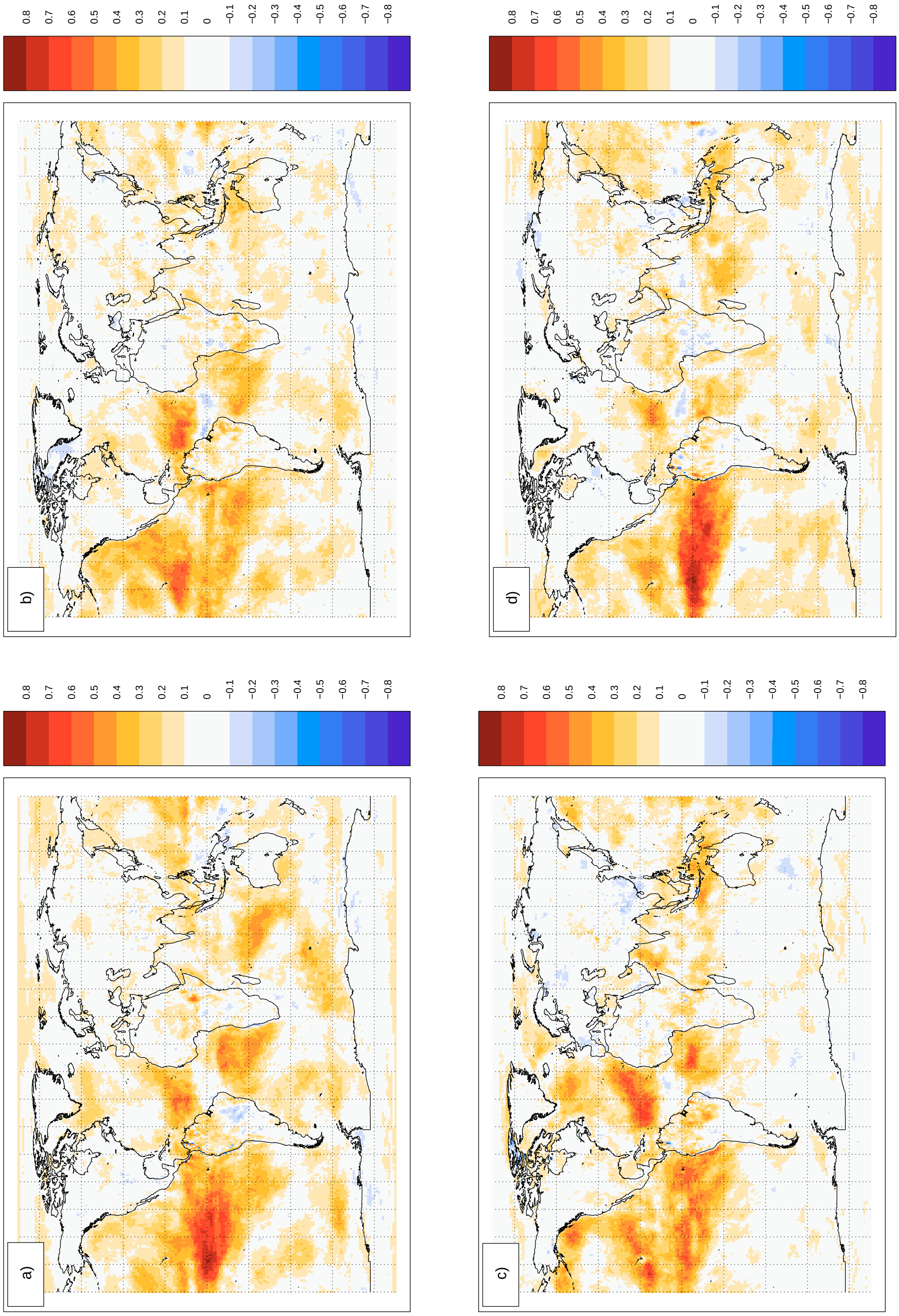

Figure App.3: Average $RPSS_D$ skill of week 3 forecasts in seasonal resolution. Winter (a), spring (b), summer (c), and fall (d) are shown.

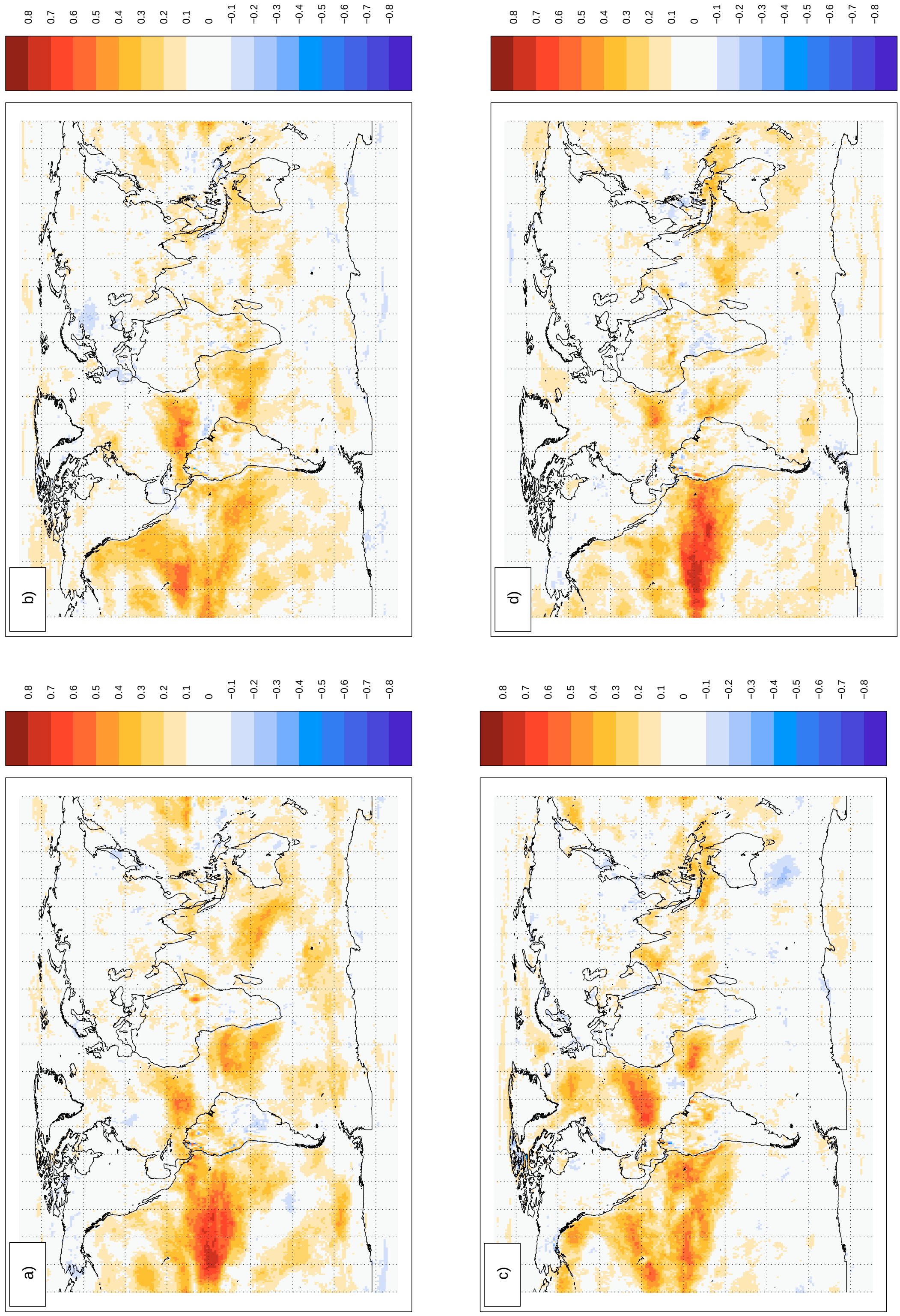

Figure App.4: Average $RPSS_D$ skill of week 4 forecasts in seasonal resolution. Winter (a), spring (b), summer (c), and fall (d) are shown.

Regional maps of $RPSS_D$ of monthly forecasts for individual seasons

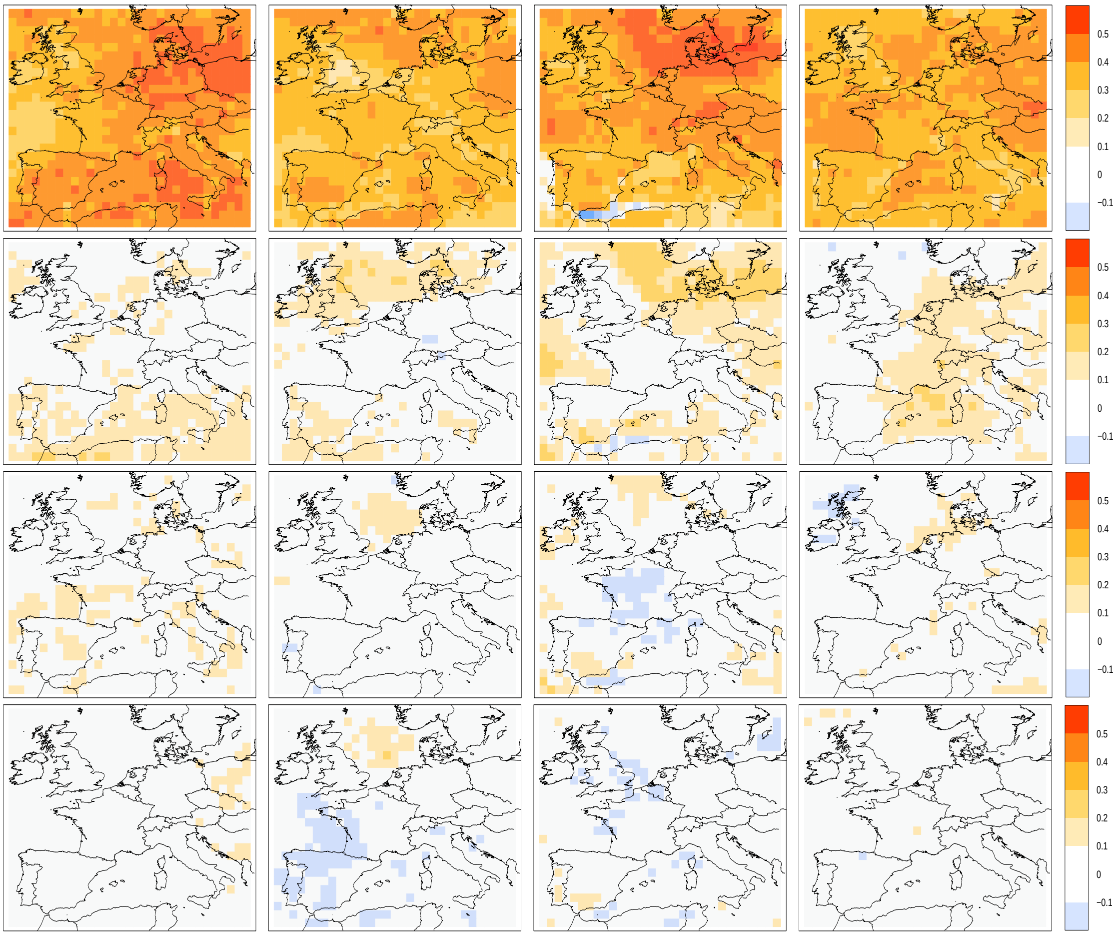

Figure App.5: Seasonally averaged skill maps for the European domain. Winter, spring, summer and fall are shown in this order from left to right. The top row contains week 1 forecasts, the second week 2 etc.

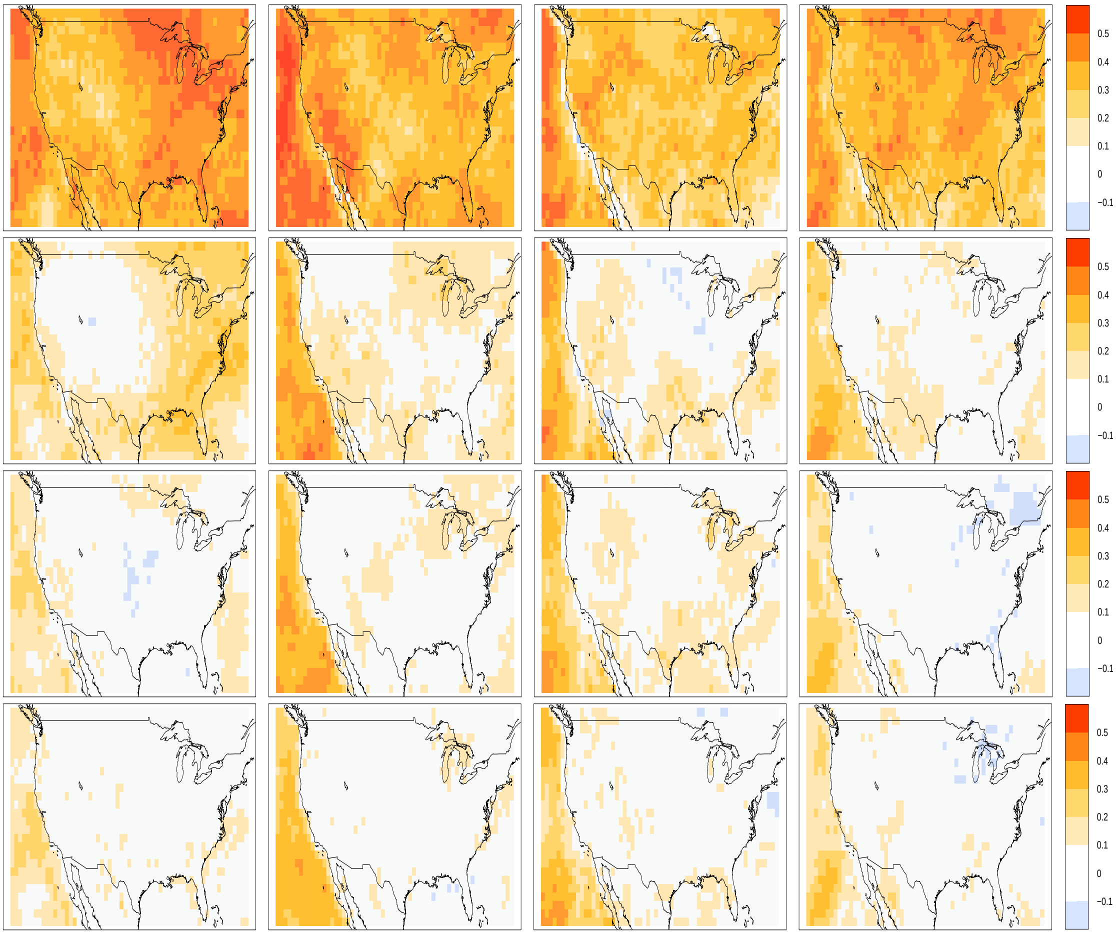

Figure App.6: Seasonally averaged skill maps for the North American domain. Winter, spring, summer, and fall are shown in this order from left to right. The top row contains week 1 forecasts, the second week 2 etc.

Bibliography

Baldwin, M. P. and T. J. Dunkerton (2001, October). Stratospheric harbingers of anomalous weather regimes. *Science 294*(5542), 581–584.

Begert, M., T. Schlegel, and W. Kirchhofer (2005, January). Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal Of Climatology 25*(1), 65–80.

Brier, G. (1950). Verification of forecasts expressed in terms of probabilities. *Monthly Weather Review 78*, 1–3.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Z. Wei, and Y. J. Zhu (2005, May). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review 133*(5), 1076–1097.

Candille, G. and O. Talagrand (2005, July). Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal Of The Royal Meteorological Society 131*(609), 2131–2150.

Croci-Maspoli, M., C. Schwierz, and H. C. Davies (2007, February). A multifaceted climatology of atmospheric blocking and its recent linear trend. *Journal Of Climate 20*(4), 633–649.

ECMWF (2006, February). Monthly Forecasting Documentation. *http://www.ecmwf.int/research/monthly_forecasting/Documentation.html*. last accessed March 15, 2007.

Epstein, E. (1969). A scoring system for probability forecasts of ranked categories. *Journal Of Applied Meteorology 8*, 985–987.

Hamill, T. M., J. S. Whitaker, and X. Wei (2004, June). Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review 132*(6), 1434–1447.

Hammer, G. L., J. W. Hansen, J. G. Phillips, J. W. Mjelde, H. Hill, A. Love, and A. Potgieter (2001, November). Advances in application of climate prediction in agriculture. *Agricultural Systems 70*(2-3), 515–553.

Hartley, D. E., J. T. Villarin, R. X. Black, and C. A. Davis (1998, January). A new perspective on the dynamical link between the stratosphere and troposphere. *Nature 391*(6666), 471–474.

Haynes, P. (2005). Stratospheric dynamics. *Annual Review Of Fluid Mechanics 37*, 263–293.

Jones, C., D. E. Waliser, K. M. Lau, and W. Stern (2004, June). The Madden-Julian oscillation and its impact on Northern Hemisphere weather predictability. *Monthly Weather Review 132*(6), 1462–1471.

Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina (2000, December). Potential benefits of climate forecasting to agriculture. *Agriculture Ecosystems & Environment 82*(1-3), 169–184.

Jung, T. (2005, April). Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quarterly Journal Of The Royal Meteorological Society 131*(607), 1045–1073.

Jung, T. and F. Vitart (2006, July). Short-range and medium-range weather forecasting in the extratropics during wintertime with and without an interactive ocean. *Monthly Weather Review 134*(7), 1972–1986.

Kunz, H., S. C. Scherrer, M. A. Liniger, and C. Appenzeller (2007). The Evolution of ERA-40 Surface Temperatures and Total Ozone compared to observed Swiss time series. *Meteorologische Zeitschrift, in press.*

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider (1998, June). A review of the predictability and prediction of ENSO. *Journal Of Geophysical Research-Oceans 103*(C7), 14375–14393.

Liniger, M. A. (2003). Weather Storms and Climate Anomalies. *GAiA 12*(4), 241–336.

Liniger, M. A., H. Mathis, C. Appenzeller, and F. J. Doblas-Reyes (2007). Realistic Greenhouse Gas Forcing and Seasonal Forecasts. *Geophysical Research Letters, in press.*

Lorenz, E. N. (1969). Predictability Of A Flow Which Possesses Many Scales Of Motion. *Tellus 21*(3), 289–&.

Madden, R. A. and P. R. Julian (1994, May). Observations Of The 40-50-Day Tropical Oscillation - A Review. *Monthly Weather Review 122*(5), 814–837.

Muller, W. A., C. Appenzeller, F. J. Doblas-Reyes, and M. A. Liniger (2005, May). A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal Of Climate 18*(10), 1513–1523.

Murphy, A. H. and E. S. Epstein (1967). A note on probability forecasts and "hedging". *J. Appl. Meteor. 6*, 1002–1004.

Reichler, T. and J. O. Roads (2005, March). Long-range predictability in the tropics. Part II: 30-60-day variability. *Journal Of Climate 18*(5), 634–650.

Richardson, D. S. (2001, October). Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal Of The Royal Meteorological Society 127*(577), 2473–2489.

Richardson, L. F. (1922). *Weather Prediction by Numerical Process.* Cambridge Mathematical Library. Cambridge University Press.

Scherrer, S. C., M. Croci-Maspoli, C. Schwierz, and C. Appenzeller (2006, February). Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal Of Climatology 26*(2), 233–249.

Schwierz, C., C. Appenzeller, H. C. Davies, M. A. Liniger, W. Muller, T. F. Stocker, and M. Yoshimori (2006, November). Challenges posed by and approaches to the study of seasonal-to-decadal climate variability. *Climatic Change 79*(1-2), 31–63.

Shongwe, M., C. Ferro, C. Coelho, and G. van Oldenborgh (2007). Predictability of cold springs in Europe. *Journal Of Climate, in press.*

Simmons, A. J. and A. Hollingsworth (2002, January). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal Of The Royal Meteorological Society 128*(580), 647–677.

Simmons, A. J., P. D. Jones, V. D. Bechtold, A. C. M. Beljaars, P. W. Kallberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004, December). Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal Of Geophysical Research-Atmospheres 109*(D24), D24115.

Slingo, J. M., D. P. Rowell, K. R. Sperber, and E. Nortley (1999, January). On the predictability of the interannual behaviour of the Madden-Julian Oscillation and its relationship with El Nino. *Quarterly Journal Of The Royal Meteorological Society 125*(554), 583–609.

Uppala, S., P. Kallberg, A. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. Allan, E. Andersson, K. Arpe, M. Balmaseda, A. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B. Hoskins, L. Isaksen, P. Janssen, R. Jenne, A. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. Rayner, R. Saunders, P. Simon, A. Sterl, K. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen (2005). The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc. 131*, 2961–3012.

Van Oldenborgh, G. J., M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. T. Anderson (2005, August). Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *Journal Of Climate 18*(16), 3240–3249.

Vitart, F. (2004, December). Monthly forecasting at ECMWF. *Monthly Weather Review 132*(12), 2761–2779.

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2007a). Can multi-model combination really enhance prediction skill of probabilistic ensemble forecasts? *in preparation.*

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2007b). The discrete Brier and ranked probability skill scores. *Monthly Weather Review 135*(1), 118–124.

Wilks, D. (1995). *Statistical Methods in the Atmospheric Sciences*, Volume 59 of *International Geophysics Series*. Academic Press.

Zeng, L. X. (2000, September). Weather derivatives and weather insurance: Concept, application, and analysis. *Bulletin Of The American Meteorological Society 81*(9), 2075–2082.

Veröffentlichungen der MeteoSchweiz

75 Fikke S, Ronsten G, Heimo A, Kunz S, Ostrozlik M, Persson PE, Sabata J, Wareing B, Wichura B, Chum J, Laakso T, Säntti K, Makkonen L: 2007, COST 727: Atmospheric Icing on Structures Measurements and data collection on icing: State of the Art, 110pp, CHF 83.-

74 Schmutz C, Müller P, Barodte B: 2006, Potenzialabklärung für Public Private Partnership (PPP) bei MeteoSchweiz und armasuisse Immobilien, 82pp, CHF 76.-

73 Scherrer SC: 2006, Interannual climate variability in the European and Alpine region, 132pp, CHF 86.-

72 Mathis H: 2005, Impact of Realistic Greenhouse Gas Forcing on Seasonal Forecast Performance, 80pp, CHF 75.

71 Leuenberger D: 2005, High-Resolution Radar Rainfall Assimilation: Exploratory Studies with Latent Heat Nudging, 103pp, CHF 81.-

70 Müller G und Viatte P: 2005, The Swiss Contribution to the Global Atmosphere Watch Programme – Achievements of the First Decade and Future Prospects, 112pp, CHF 83.-

69 Müller WA: 2004, Analysis and Prediction of the European Winter Climate, 115pp, CHF 34.

68 Bader S: 2004, Das Schweizer Klima im Trend: Temperatur- und Niederschlagsentwicklung seit 1864, 48pp, CHF 18.-

67 Begert M, Seiz G, Schlegel T, Musa M, Baudraz G und Moesch M: 2003, Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961-1990, Schlussbericht des Projektes NORM90, 170pp, CHF 40.-

66 Schär Christoph, Binder Peter, Richner Hans (Eds.): 2003, International Conference on Alpine Meteorology and MAP Meeting 2003, Extended Abstracts volumes A and B, 580pp, CHF 100.

65 Stübi R: 2002, SONDEX / OZEX campaigns of dual ozone sondes flights: Report on the data analysis, 78pp, CHF 27.-

64 Bolliger M: 2002, On the characteristics of heavy precipitation systems observed by Meteosat-6 during the MAP-SOP, 116pp, CHF 36.-

63 Favaro G, Jeannet P, Stübi R: 2002, Re-evaluation and trend analysis of the Payerne ozone sounding, 99pp, CHF 33.-

62 Bettems JM: 2001, EUCOS impact study using the limited-area non-hydrostatic NWP model in operational use at MeteoSwiss, 17pp, CHF 12.-

61 Richner H, et al.: 1999, Grundlagen aerologischer Messungen speziell mittels der Schweizer Sonde SRS 400, 140pp, CHF 42.-

60 Gisler O: 1999, Zu r Methodik einer Beschreibung der Entwicklung des linearen Trends der Lufttemperatur über der Schweiz im Zeitabschnitt von 1864 bis 1990, 125pp, CHF 36.-

59 Bettems J-M: 1999, The impact of hypothetical wind profiler networks on numerical weather prediction in the Alpine region, 65pp, CHF 25.-

58 Baudenbacher, M: 1997, Homogenisierung langer Klimareihen, dargelegt am Beispiel der Lufttemperatur, 181pp, CHF 50.-

57 Bosshard, W: 1996, Homogenisierung klimatologischer Zeitreihen, dargelegt am Beispiel der relativen Sonnenscheindauer, 136pp, CHF 38.-

56 Schraff, C: 1996, Data Assimilation and Mesoscale Weather Prediction: A Study with a Forecast Model for the Alpine Region, 138pp, CHF 38.-

Frühere Veröffentlichungen und Arbeitsberichte finden sich unter
www.meteoschweiz.ch » Forschung » Publikationen

Arbeitsberichte der MeteoSchweiz

214 Schmucki D., Weigel A., 2006, Saisonale Vorhersage in Tradition und Moderne: Vergleich der "Sommerprognose" des Zürcher Bööggs mit einem dynamischen Klimamodell, 214, 46pp, CHF 68.-

213 Frei C: 2006, Eine Länder übergreifende Niederschlagsanalyse zum August Hochwasser 2005. Ergänzung zu Arbeitsbericht 211, 10pp, CHF 59.-

212 Z'graggen, L: 2006, Die Maximaltemperaturen im Hitzesommer 2003 und Vergleich zu früheren Extremtemperaturen, 74pp, CHF 75.-

211 MeteoSchweiz: 2006, Starkniederschlagsereignis August 2005, 63pp., CHF 72.-

210 Buss S, Jäger E and Schmutz C: 2005: Evaluation of turbulence forecasts with the aLMo, 58pp, CHF 70.–

209 Schmutz C, Schmuki D, Duding O, Rohling S: 2004, Aeronautical Climatological Information Sion LSGS, 77pp, CHF 25.–

208 Schmuki D, Schmutz C, Rohling S: 2004, Aeronautical Climatological Information Grenchen LSZG, 73pp, CHF 24.–

207 Moesch M, Zelenka A: 2004, Globalstrahlungsmessungen 1981-2000 im ANETZ, 83pp, CHF 26.–

206 Schmutz C, Schmuki D, Rohling S: 2004, Aeronautical Climatological Information St.Gallen LSZR, 78pp, CHF 25.–

205 Schmutz C, Schmuki D, Ambrosetti P, Gaia M, Rohling S: 2004, Aeronautical Climatological Information Lugano LSZA, 81pp, CHF 26.–

204 Schmuki D, Schmutz C, Rohling S: 2004, Aeronautical Climatological Information Bern LSZB, 80pp, CHF 25.–

203 Duding O, Schmuki D, Schmutz C, Rohling S: 2004, Aeronautical Climatological Information Geneva LSGG, 104pp, CHF 31.–

202 Bader S: 2004, Tropische Wirbelstürme – Hurricanes –Typhoons – Cyclones, 40pp, CHF 16.-

201 Schmutz C, Schmuki D, Rohling S: 2004, Aeronautical Climatological Information Zurich LSZH, 110pp, CHF 34.-

200 Bader S: 2004, Die extreme Sommerhitze im aussergewöhnlichen Witterungsjahr 2003, 25pp, CHF 14.-

199 Frei T, Dössegger R, Galli G, Ruffieux D: 2002, Konzept Messsysteme 2010 von MeteoSchweiz, 100pp, CHF 32.-

198 Kaufmann P: 2002, Swiss Model Simulations for Extreme Rainfall Events on the South Side of the Alps, 40pp, CHF 20.-

197 WRC Davos (Ed): 2001, IPC - IX, 25.9. - 13.10.2000, Davos, Switzerland, 100pp, CHF 32.-

196 Hächler P et al.: 1999, Der Föhnfall vom April 1993, 139pp, CHF 40.-

195 Urfer Ch, Vogt R, 1999, Die Niederschlagsverhältnisse in Basel 1964-1998, 43pp, CHF 40.-

194 Courvoisier HW: 1998, Statistik der 24-stündigen Starkniederschläge in der Schweiz 1901-1996, 20pp, CHF 11.-

193 Defila C, Vonderach G: 1998, Todesfälle und Wetterlagen in Schaffhausen, 72pp, CHF 25.-

Frühere Veröffentlichungen und Arbeitsberichte finden sich unter
www.meteoschweiz.ch » Forschung » Publikationen