Analysis and prediction of the European winter climate

Author(s):
Müller, Wolfgang Alexander

Publication Date:
2004

Permanent Link:

https://doi.org/10.3929/ethz-a-004828877 →

This Thesis

Analysis and Prediction of the European Winter Climate

by
WOLFGANG A. MÜLLER

was submitted as a dissertation to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY (ETH)
ZÜRICH
(ETH No 15540)

and accepted on the recommendation of
Prof. Dr. Christoph Schär, examiner
PD Dr. Christof Appenzeller, co-examiner
Prof. Dr. Mojib Latif, co-examiner

2004

Contents

Abstract

In this thesis a dynamical forecast approach is considered to evaluate the potential seasonal predictability in the European-Atlantic region with emphasis on the mean winter climate. Two state-of-the-art seasonal forecast systems are used, namely the Seasonal Forecast System 2 from the European Centre for Medium Range Weather Forecast (ECMWF) and a multi-model system developed within the joint European project DEMETER (Development of a European Multi-Model Ensemble Prediction System for Seasonal to Interannual Prediction). The predictions are verified with the ERA-40 re-analysis data.

Seasonal forecasts are probabilistic in nature and hence require verification techniques based on probabilistic skill measures. Here a multi-category skill score, namely the ranked probability skill score (RPSS) is applied. The RPSS is sensitive to the shape and the shift of the predicted probability density distribution. However, the RPSS shows a negative bias for ensemble systems with small ensemble sizes. It is shown that the negative bias can be attributed to a discretization and squaring error in the quadratic norm of the RPSS. In the following two strategies are explored to tackle this flaw. First, it is shown that the $RPSS_{L=1}$ based on the absolute rather than the squared norm is unbiased. Nevertheless, it is not strictly proper in a statistical sense. Second, an unbiased and strictly proper skill score can be defined based on the quadratic norm, along with the reference forecast reduced to sub-samples of the same size as the forecast ensemble size. This is denoted as the de-biased ranked probability skill score ($RPSS_D$). Based on a hypothetical set up comparable to the ECMWF hindcast system (40 members, 15 hindcast years) the $RPSS_D$ is used to show, that statistically significant skill scores can only be found for climate anomalies with a signal-to-noise ratio larger than ~0.3.

Furthermore, the seasonal predictability is evaluated using a forecast approach (FA) based on 2m mean temperature predictions on grid-point scale for the years 1987-2001. The ECMWF Seasonal Forecast System 2 provides a marked improvement in skill relative to climatological forecasts over the North-Atlantic Ocean with maximum values of up to 30 %. Over Europe no significantly positive skill scores are found. The DEMETER multi-model has higher forecast skills than individual models. Moreover, the potential predictability is investigated applying a perfect model approach (PMA). Such approach assumes that the climate system is fully represented by the model physics. The potential winter predictability over the European continent amounts to approximately ~10%.

The 3rd part of the thesis examines the potential seasonal predictability is examined via the leading mode of the European winter climate variability, namely the North Atlantic Oscillation (NAO). The PMA shows that the mean winter NAO and the NAO-

temperature related impact is potentially predictable for lead time 1 month, but with a gain in skill of only ~8 % compared to climatology. Using the FA, the results are quite different. For the period 1959-2001, the NAO skill score is not statistically significant, while the skill score is surprisingly large (16 % to 27 % relative to the observed climatology) for the period 1987-2001. For this period a weak relation between the strength of the NAO amplitude and the skill score of the NAO is found. This contrasts with ENSO variability where the amplitude dependent forecast skill is strong.

Finally, the seasonal forecasts are examined from the end user's perspective. A so-called "Klimagram" is introduced to assess seasonal climate forecasts for particular cities or regions. A first analysis reveals that the forecast skills can be improved in a relative sense, looking at spatial and temporal averaged quantities.

Overall, this study suggests a positive potential seasonal predictability in the European-Atlantic domain in winter. However, the potential benefit is rather small and constitutes a fraction only, compared to currently possible results in the tropics.

Zusammenfassung

In dieser Arbeit wird die potentielle saisonale Vorhersagbarkeit des Europäischen-Atlantischen Winterklimas mit Hilfe dynamischer Vorhersagemethoden untersucht. Hierfür werden zwei saisonale Vorhersagesysteme benutzt: das saisonale Vorhersagesystem 2 des Europäischen Zentrum für mittelfristige Wettervorhersage (EZMW) und das Multi-Model System DEMETER (Development of a European Multi-Model Ensemble Prediction System for Seasonal to Interannual Prediction). Die Vorhersagen werden mit den Beobachtungen aus den ERA40 Re-analysen verifiziert.

Saisonale Vorhersagen sind probabilistischer Natur und benötigen daher Verifikationstechniken die auf Wahrscheinlichkeitsmaße zurückgreifen. Die Verifikation in dieser Arbeit basiert auf einem mit mehreren Klassen umfassendes Maß: der Vorhersagegüte für abgestufte Wahrscheinlichkeiten (Ranked Probability Skill Score, RPSS). Das RPSS berücksichtigt die Form und Verlagerung der Wahrscheinlichkeits-funktion. Das RPSS zeigt jedoch einen negativen Bias für Ensemble-Systeme mit kleiner Ensemblegröße, der auf ein Diskretisierungs- und Quadrierungsproblem der quadratischen Norm des RPSS zurückführbar ist. Zwei Strategien werden untersucht, die dieses Problem angehen. Zum einen wird gezeigt, dass im Gegensatz zur quadratischen Norm, die absolute Norm keinen Bias aufweist. Es zeigt sich jedoch das dieses Maß im statistischen Sinne nicht streng genug maßgebend ist. Zum anderen wird ein Maß definiert, welches auf die quadratische Norm des RPSS zurückgreift, keinen Bias aufweist und maßgebend ist. Hierbei wird die Referenzvorhersage auf eine Unterstichprobe, der gleichen Größe wie die Ensemblevorhersagen, reduziert (de-biased Ranked Probability Skill Score, $RPSS_D$). Mittels einer hypothetischen Konfiguration, vergleichbar der des ECMWF System 2 (40 Ensemblemitglieder und 15 Vorhersagen), wird gezeigt, dass statistisch signifikante $RPSS_D$ nur dann für Klimaanomalien erreicht werden können, wenn das Verhältnis zwischen Signal und Rauschanteil ungefähr 0.3 ist.

Die saisonale Vorhersagbarkeit wird zunächst im Vorhersagemodus (FA) für die 2m Temperatur und den Zeitraum 1987-2001 an jedem Gitterpunkt bewertet. Das EZMW Vorhersagesystem 2 zeigt über dem Nordatlantischen Ozean eine deutliche Verbesserung gegenüber klimatologischen Vorhersagen, mit Werten bis zu 30 %. Über Europa können keine signifikant positive $RPSS_D$ gefunden werden. Das DEMETER Multimodell weist insgesamt eine höhere Vorhersagegüte auf als das Mittel der einzelnen Modelle. Mit Hilfe eines perfekten Modellansatzes (PMA) wird die potentielle Vorhersagbarkeit bestimmt. Diese Methode setzt ein Klimasystem voraus, dass vollständig durch die Modellphysik beschrieben wird. Die potentielle Vorhersagbarkeit des Winterklimas über dem Europäischen Kontinent erreicht nun etwa 10 %.

Die saisonale Vorhersagbarkeit wird alternativ mit dem dominanten Mode der Euro-Atlantischen Klimavariabilität, der Nordatlantischen Oszillation (NAO), untersucht. Der PMA zeigt, dass die Winter gemittelte NAO sowie der NAO bezogene Einfluss auf die Temperatur potentiell vorhersagbar sind, zumindest für eine Vorlaufzeit von einem Monat. Der Gewinn gegenüber den klimatologischen Vorhersagen ist jedoch gering und beträgt lediglich ~8 %. Für den FA werden unterschiedlich Resultate erzielt. Für den Zeitraum 1959-2001 ist die Vorhersagegüte der NAO statistisch nicht signifikant, während sie für den Zeitraum 1987-2001 überraschenderweise groß ist (16 % bis 27 % gegenüber der beobachteten Klimatologie). Für diesen Zeitraum wird weiterhin ein Zusammenhang zwischen der Stärke der NAO Amplitude und der Vorhersagegüte gefunden. Dieser ist jedoch schwach und steht im Gegensatz zur ENSO Variabilität, bei der es eine starke Abhängigkeit der Vorhersagegüte von der Amplitude gibt.

Schließlich werden die saisonalen Vorhersagen von der Perspektive des Endverbrauchers untersucht. Das so genannte Klimagramm wird eingeführt, um saisonale Klimavorhersagen für bestimmte Städte oder Regionen abzuschätzen. Eine erste Analyse zeigt, dass die Vorhersagegüte für räumlich und zeitlich gemittelte Größen verbessert werden kann.

Zusammenfassend zeigt diese Studie eine positive potentielle Vorhersagbarkeit des Europäischen-Atlantischen Winterklimas. Der potentielle Nutzen ist jedoch gering und lediglich ein Bruchteil dessen, was in den Tropen erreicht werden kann.

Chapter 1:

Seasonal to Interannual Climate Variability

1.1. Motivation

The European summer of 2003 has been one of the most extraordinary of its kind, in the climatological but also in the socio-economic context. A strong persistent anticyclonic circulation, only interrupted by a few synoptic frontal disturbances, dominated the summer months (Grazzini et al. 2003). A warm period in late spring and the subsequently developed massive heat wave affected the European continent during the entire summer season. Temperature anomalies were found to exceed the climatological mean by at least ~3°C which corresponds to about five standard deviations (Figure 1-1a – Schär et al. 2004). Record-breaking values of sea and land surface temperatures were also observed in the western Mediterranean Sea and its circumjacent countries (Grazzini et al. 2003, Grazzini and Viterbo 2003). An extremely dry period with associated high temperature places the summer heat wave on top of the largest natural catastrophes in 2003, with an economic damage of approximately 13bn US$ for Europe (MunichRe[1], 2003). Preliminary analysis suggests an enhanced mortality with about >20000 attributed deaths alone in France, Italy and Portugal (Kovats et al. 2004).

The question is whether timely seasonal climate forecasts could have limited the disastrous effects. The probabilistic summer forecast of the ECMWF Seasonal Forecast System 2 has predicted a signal one month in advance some what similar to the observed one. Figure 1-1b shows the probability forecast of summer mean 2m temperature anomalies above mean. Much of southern and the northwestern Europe were covered by a significant signal which indicated a 70 % chance for the 2m temperature to be above normal. Low probabilities over Eastern Europe and the sub-tropical Atlantic were forecasted indicating colder than normal summer conditions. The forecasts, however, failed to predict the intense amplitudes (not shown). Only anomalies with a magnitude of about 1.5°C were predicted in those areas where the heat wave reached its maximum. Although the system was able to predict the tendency of the anomalies, it failed to point out the relevance of the event. Moreover, subsequent forecasts initialized in June did not support the May predictions.

[1] TOPICSgeo Annual Review: Natural Catastrophes 2003; MunichRe

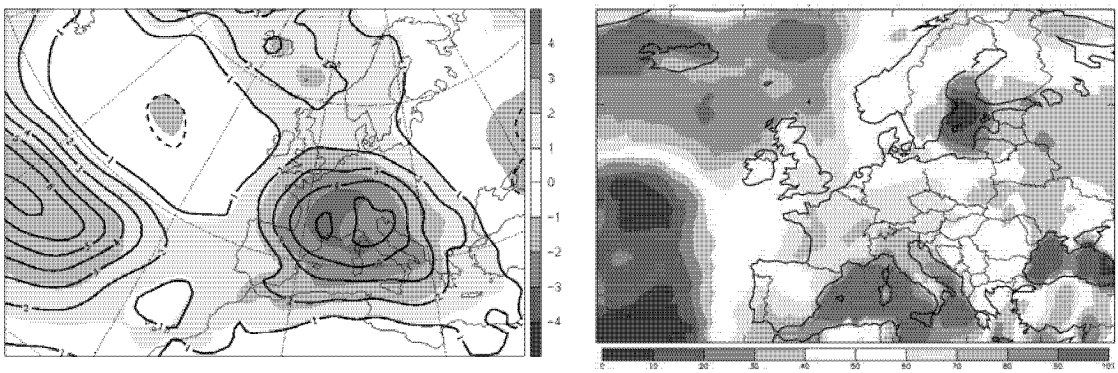


Figure 1-1: The observed (left) and predicted (right) summer (June-July-August) 2003. Shown are the 2m temperature anomalies of the ERA40 Re-Analysis (shadings 1°C) and the standard deviations (contours 1σ) based on the climatology for the period 1961-1990 (from Schär et al 2004). The prediction results from the ECMWF Seasonal Forecast System 2 initialised on the 1st May 2003. The model climatology is based on the period from 1987-2001 (times 40 ensemble members). Shown is the probability of the 2m temperature being greater than the long-term mean.

The prediction of seasonal climate anomalies has recently become operational at various meteorological centres. Several validation studies suggest potentially useful information from seasonal climate forecasts in many parts of the world. In this context it is of special importance to explore the variability and predictability of the relevant processes. The strongest signal on seasonal to interannual climate time-scales is given by the natural climate variation in the tropical-subtropical Pacific, known as ElNiño/Southern Oscillation (ENSO, see also Trenberth et al. 1998, Diaz and Markgraf 2000). For Europe, however, a direct impact of the equatorial Pacific is supposed to be limited (Fraedrich 1994, Merkel and Latif 2003). Other relevant processes such as the North Atlantic Oscillation (NAO) have substantial impact on the European seasonal climate.

The aim of this thesis is to explore the climate patterns primarily responsible for the seasonal predictability in the European-Atlantic region, using state-of-the art coupled ocean-atmosphere general circulation models (CGCM). Their impact on surface related quantities, such as near surface temperature is also analysed. The present chapter continues with a brief description of the basic properties of ENSO and the mid-latitude response to tropical forcing. Subsequently, the NAO is introduced in more detail. The methods and Seasonal Forecast Systems used in this thesis are illustrated in chapter 2. The forecast verification scheme is examined in chapter 3. In the following chapters the forecast skill on seasonal timescales is discussed on grid point scale (chapter 4) and in terms of the winter NAO and its impact on near surface temperature (chapter 5). Implications for the predictability of the European-Atlantic climate are also discussed. Conclusions follow in chapter 6. Finally a user specific application is prescribed (Appendix A).

1.2. The Role of the Tropics

El Niño/Southern Oscillation (ENSO)

The seasonal to interannual climate variability in the tropical region is strongly influenced by the atmosphere-ocean phenomenon known as ENSO. ENSO is defined as the coupled process of a see-saw in atmospheric pressure between Tahiti and Darwin namely the Southern Oscillation (SO, Walker 1924, Walker and Bliss 1930, Walker and Bliss 1932) and the occurrence of warm southward flowing water off the coast of Peru and Ecuador around the end of the year (El Niño).

A physical explanation of the coupling process between the SO and the equatorial Pacific has been investigated by Bjerknes (1969). He proposed the existence of a large-scale vertical and zonal circulation cell (Walker circulation) which is mutually linked to the gradient of relatively warm surface water in the western and cold water in the eastern equatorial Pacific. During normal (or negative) ENSO conditions warm surface waters in the western Pacific are associated with enhanced convection and heavy rainfall in the western Pacific. The strong convection causes an atmospheric flow convergence, which redirects air parcels along the equator and forms a basin-wide circulation. In the eastern Pacific air parcels descend and cause relatively dry surface conditions. Resulting easterly surface winds in turn produce an east-west sea level gradient with deep warm water in the West Pacific and shallow cold water in the East Pacific. During positive ENSO conditions the warm surface water and corresponding convection is shifted in the middle Pacific causing a change in the atmospheric circulation structure.

El Niño/Southern Oscillation: Impact

Such basin-wide variations appear with a frequency of 2-7y and have a remote impact on distinct regions around the globe (for review see Diaz and Markgraf 2000). Figure 1-2 shows the correlation of the winter Nino3 index with 2m temperature for stations around the globe (van Oldenborgh and Burgers 2004). Strong climate shifts are located in the central Pacific and adjacent continents where ENSO is most active. Warmer than normal winters which are associated with a positive phase of ENSO are found in eastern Australia, northern South America and southeast Africa.

But also remote regions in the extra-tropics are affected by ENSO. During positive ENSO phases there is a predominance of warmer than usual winter conditions in northwestern and eastern Canada. In the southeast of the United States and the northwest Pacific cool conditions are associated with the positive phase of ENSO. The European-Atlantic impact of ENSO, however, is rather weak and more uncertain. Observational studies

suggest that a weak response exists during boreal winter seasons (Fraedrich and Müller 1992, Fraedrich 1994). In particular, they found a correlation between strong positive phases of ENSO with negative pressure anomalies in central and western Europe on the one hand, and negative anomalies in northern Europe on the other hand. During cold periods these fields are shifted northwards. They argue that this shift corresponds to a shift in the dominating cyclone tracks. Recent modelling studies support this view of an El Niño-related weakening of the North Atlantic mean meridional pressure gradient and an associated shift of the storm tracks (Merkel and Latif 2002).

El Niño/Southern Oscillation: Predictability

The equatorial waves in the tropical Pacific are essential for the description and prediction of ENSO. The interplay of various types of ocean Kelvin and Rossby waves propagating and reflecting at coastal equatorial boundaries have been suggested to provide the memory of ENSO. An important theoretical development has been issued by the "delayed oscillation theory" described by Schopf and Suarez (1988), Suarez and Schopf (1988) and Battisti and Hirst (1989). Battisti and Hirst (1989) found that the propagation delay of oceanic Rossby waves originates in the mid-equatorial Pacific. Together with the reflection at the western boundary it yields the characteristic timescale of ENSO. Although this model is highly idealised and the oscillatory variations are

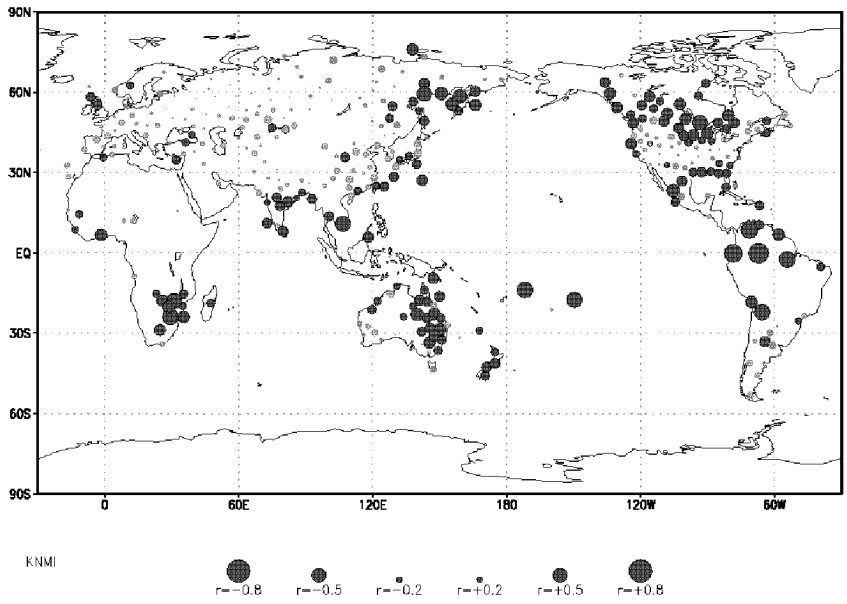


Figure 1-2: The correlation coefficients of winter 2m temperature with the Nino3 index [5°S-5°N, 150°W-90°W] at different GHCN stations. Only stations with at least 50 years of data are considered (from van Oldenborgh and Burgers 2004).

superposed by non-periodic elements, such as random weather noise (i.e. westerly wind burst) and the intrinsic chaos within the non-linear dynamics of longer lived components (for a review see Neelin et al. 1994, Neelin et al. 1998), it provides the basis for understanding the predictability of seasonal to interannual timescales in the tropics.

Coupled atmosphere-ocean models as well as statistical models yield an appropriate level of forecast skill up to a lead time of 12 months (for review see Latif et al. 1998). A forecast of the actual ENSO phase, however, can currently be made only a few months ahead. The predictions on seasonal to interannual timescales critically depend on the information on the phase at the beginning of such period. Especially, random weather noise can strongly affect the forecast. Random bursts of westerly winds that occur when El Niño is developing can amplify the event, while their occurrence after the peak of the event can prolong its duration (Fedorov 2002). Concerning the 1997/98 El Niño, it is argued that the strong amplitude could not have been attained without the unusual occurrence of westerly wind bursts (McPhadden and Yu 1999, Vialard et al. 2001, Boulanger 2001, Vitart et al. 2003).

In a comparative study of the forecast for the 1997/98 El Niño, Landsea and Knaff (2002) examined 12 deterministic statistical and dynamical models and found no skill beyond persistence on seasonal to interannual timescales for any model. However, if the uncertainty of random atmospheric noise is accounted for, an enhanced forecast skill is suggested (Fedorov et al. 2003, Coelho et al. 2003). In this thesis we shortly return to this discussion and show that, on the base of state-of–the-art CGCM or a newly designed probabilistic persistence model, there is reasonable forecast skill throughout the year, at least one season ahead (see chapter 4).

1.3. The Mid-Latitude Response to Tropical Forcing

The dominant seasonal to interannual climate variability in the mid latitudes are the Pacific North America pattern (PNA) and the North Atlantic Oscillation (NAO). These patterns are characterised by a statistical linkage over large distances. Therefore they are commonly referred to as “teleconnections”. These teleconnections, however, have also a close link to remote regions in the tropics (for review see Trenberth et al. 1998). The seasonal variability of the PNA for instance is significantly affected by tropical heating anomalies and climate variability such as ENSO (Straus and Shukla 2000).

The PNA is characterized by four centres of action which are spread in an arcade from the mid-Pacific to the mid-Atlantic (Figure 1-3). The PNA pattern has similar signs located south of the Aleutian Islands, over the southeastern United States and the sub-tropical Atlantic. Anomalies with signs opposite to the Aleutian center are located in the vicinity of Hawaii and over central Canada. The pattern is most pronounced in the

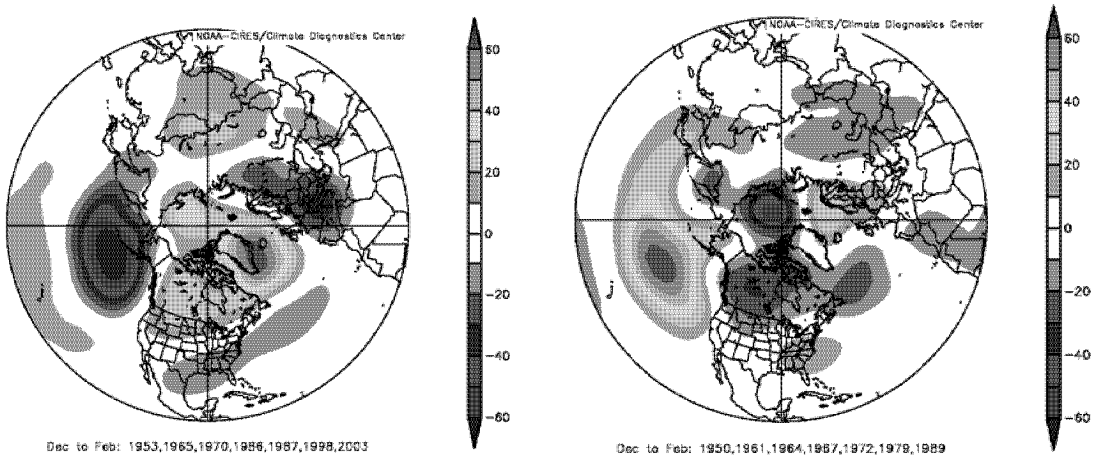


Figure 1-3: Composites of winter mean geopotential anomalies at 500hPa for seven strong positive (left panel) and seven strong negative (right panel) PNA events. The climatology is based on the period 1968-1996 (source: Climate Diagnostic Centre (CDC)).

northern hemispheric winter season and vanishes in the summer months. During winter the Aleutian center spans most of the northern latitudes of the north Pacific.

The positive phase (Figure 1-3a) of the PNA shows a deeper than normal trough over the Aleutian area and a higher than normal ridge over the Rocky Mountains. This phase is associated with an enhanced meridional upper-level flow over the US, which transports above normal temperatures to the western United States. The east to southeast of the United States is affected by an anomalous upper-level ridge and may experience intrusions of polar air masses and enhanced cyclonic activity. Since this is the region which provides the major initial sensitivity for cyclones over the North Atlantic, an anomalous flow field may also be reflected in the storm track statistics and thus the European climate (Merkel and Latif 2002). The negative phase (Figure 1-3b) produces a more zonal upper-level flow associated with less deep troughs over the Aleutian area and the southeastern United States, and a lowered ridge over the Rocky Mountains.

Predictability

In contrast to the tropics the internal atmospheric variability in the mid latitudes has almost the same amplitude as the variability generated by the "external" sources. The interannual variability, however, has associated life cycles of several days and thus limits the degree of seasonal predictability. The relationship of the variance of the internal atmospheric dynamics to the variance generated by external forcing gives a first estimate of the potential predictability on seasonal time-scales (Kumar and Hoerling 1995; Zwiers and Kharin 1998 and reference therein). Figure 1-4 shows the ratio of SST

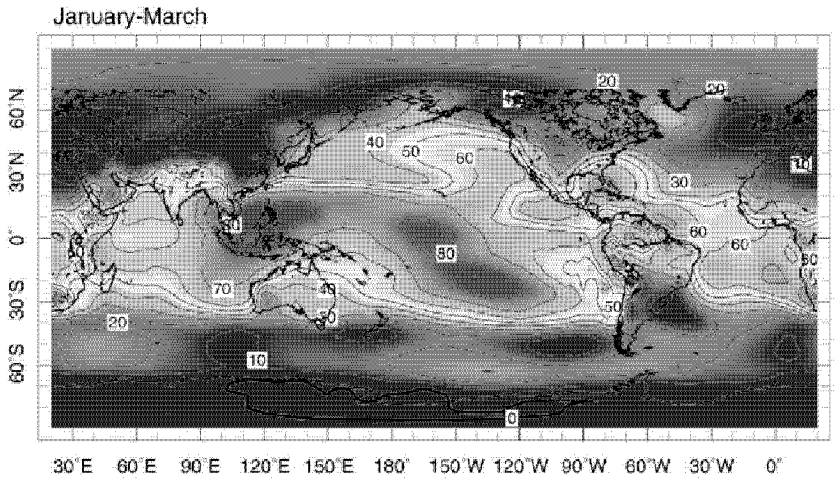


Figure 1-4: The potential predictability defined as the ratio of SST determined winter mean sea level pressure (SLP) variance to total SLP variance from an ensemble of 10 GCM integrations for the period 1950-1999 (for details see Kushnir et al. 2002). The contour intervals denote 10 % steps.

determined winter mean sea level pressure (SLP) variance to total SLP variance generated from an ensemble of GCM integrations forced with global SST.

In the tropical belt the SST forced variance reaches values ranging from 60 % to 80 %. In the mid-latitudes the externally forced variance is covering only about 20 % of the total variance. At single locations, in particular off the west coast of America, a percentage of about 60 % is found. For the European region only about 10 % of the variance is described by the SST forcing. This demonstrates that most of the interannual variability in the tropics is described by the SST-forced signal, whereas in the extra-tropics a large fraction is related to the internal atmospheric circulation, apart from distinct locations reminiscent to the teleconnections such as the PNA.

The forecast skill of seasonal climate mean anomalies in the extra-tropical Pacific region has been investigated in a multi-institutional dynamical seasonal prediction project (Shukla et al 2000). Several state-of-the-art atmospheric general circulation models with identical initial and SST condition were integrated to produce 5- to 10-member ensembles. Among others, they demonstrated that the forecast skill of winter mean mid-tropospheric geopotential forecasts, defined for the PNA region, highly correlates to strong ENSO phases. Figure 1-5 (from Shukla et al 2000) shows the anomaly correlation coefficients (ACC) for the resulting ensemble mean of all models, sorted as a function of the observed El Niño amplitude. It can be seen that the forecast skills in the PNA region are enhanced during phases with anomalous tropical SST situations. For strong El Niño years such as the winters 1982/83, and 1997/98 the forecast skills are positive for all models. Less consistent skills are found for ENSO events with small amplitude.

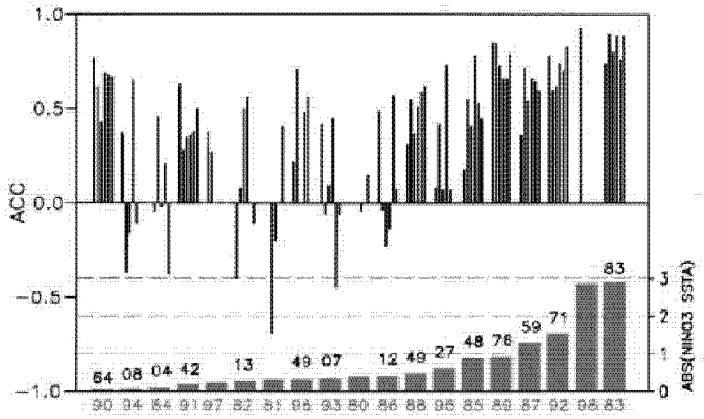


Figure 1-5: The anomaly correlation coefficients (ACC) of the ensemble mean JFM forecasts of the geopotential 500hPa sorted for the absolute value of the Nino-3 anomalies. The ACC are calculated over the PNA region for six different AGCM (from Shukla et al. 2000).

1.4. The North Atlantic Oscillation

Definition

The North Atlantic Oscillation (NAO) provides the major natural winter climate variation for the Atlantic-European region. It is characterised by a dipole-like structure with high pressure in the sub-tropical Atlantic and low pressure in the Arctic. The NAO is strongly associated with the large-scale atmospheric circulation. Changes in wind speed and direction as well as heat and moisture transport over the Atlantic and neighbouring continents occur for different phases of the NAO. In its positive phase for example, it is associated with a higher than normal pressure gradient in the North Atlantic and enhanced westerly wind flows over the middle Atlantic towards northern Europe (for a detailed review see Hurrell 2003). The NAO over the Alps is rather weak (Schär et al. 1998, Schmidli et al. 2002).

The NAO is typically described by an index representing a normalised sea level pressure difference between the Azores and Island (Rogers 1984, Jones et al. 2003). Yet, there is no universally accepted method to describe the NAO index. Most common indices are deriving from instrumental records of individual stations or from statistical variance analysis of regional meteorological fields, such as empirical-orthogonal-function (EOF) analysis. The major advantage of NAO indices based on instrumental records is their extension to several centuries in the past (Appenzeller et al. 1998, Cook 2003). Nevertheless, they are limited to measurements at a few locations and hence are affected by transient small-scale phenomena not related to the NAO. NAO indices based on EOF analysis of gridded fields provide a more robust representation of the full spatial pattern,

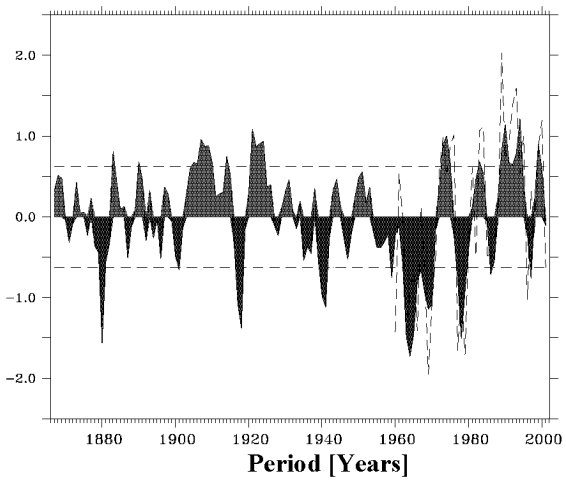


Figure 1-6: The NAO Index for the winter (DJF) mean based on the difference of normalised SLP between Lisbon and Stykkisholmur/ Reykjavik (colours) and the first principal components (PC) of an EOF analysis of the European-Atlantic geopotential at 500hPa (dashed). The dashed horizontal lines denote +/- standard deviation of the normalised SLP.

but their data are available for the second part of the 20th century solely. In Figure 1-6 two NAO indices are shown: the standard station based index and the EOF-analysis based index used in this thesis.

An EOF-analysis examines the major patterns of variability ranked by their variance. Figure 1-7 illustrates the four dominating patterns of variability resulting from an EOF-analysis of the 500hPa fields for the period 1959-2001. EOF1 (a) is representative of the NAO pattern which explains 39 % of the total variance. The variability of the NAO is evident throughout the year but is strongest during winter months and accounts for about one third of the total variance (Barnston and Livezey 1987). In summer the variance reaches its minimum and the NAO is no longer the primary mode. The spatial pattern is limited to a smaller region in the Atlantic region. EOF2 (b) has strong similarities to the Eastern Atlantic pattern described by Wallace and Gutzler (1981). The pattern of EOF3 (c) is the spatially most confined one and bears strong resemblance to the composite maps of the European-Atlantic blocking described by Tibaldi and Molteni (1990) and Tibaldi et al. (1994). In our analysis EOF4 (d) is similar to the Eurasian type I pattern of Barnston and Livezey (1987). But it also shares features with the Scandinavian pattern described by Rogers (1990).

Using such statistical techniques for the northern hemispheric fields Thompson and Wallace (1998) suggest that the NAO is a component of the more hemispheric pattern referred to as the Northern Hemispheric Annual Mode (NAM) or Arctic Oscillation (AO). In the AO a second, smaller high pressure centre in the mid-latitude Pacific is added to the NAO. These ideas also refer to the role of the stratosphere within the

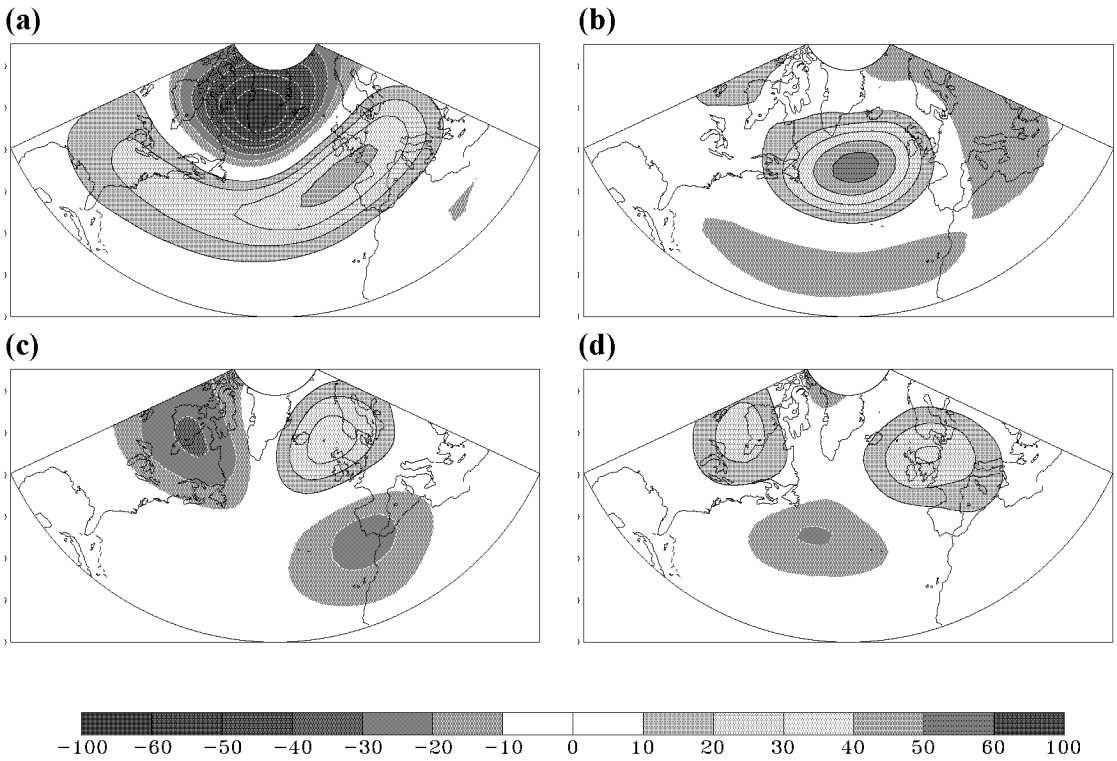


Figure 1-7: The first four patterns of an EOF analysis of the winter (DJF) mean geopotential at 500hPa for the European-Atlantic domain. The analysis is based on ERA40 data for the period 1959-2001. Shown are the (a) EOF1, (b) EOF2, (c) EOF3 and (d) EOF4. The contour intervals are 10gpm. The patterns describe 39 % (EOF1), 19 % (EOF2), 13 % (EOF3) and 9 % (EOF4) of the total variance.

climate system. (Perlwitz and Graf 1995, Thompson and Wallace 2000, Baldwin and Dunkerton 2001, Ambaum and Hoskins 2002). Baldwin and Dunkerton (2001) for instance suggest that at low-frequencies the annular mode is strongly coupled in the vertical with the zonal symmetric fluctuation in the polar vortex and that the phase of the patterns tends to propagate downwards. However, at the time of writing the discussion is ongoing, whether the NAO and AO are the same phenomenon (e.g. Wallace 2000, Deser 2000, Ambaum et al. 2001, Wallace and Thompson 2002). Caution is also demanded when interpreting the physics of teleconnection patterns deriving from EOF analysis (e.g. Dommenget and Latif 2002).

NAO-related Impact

The NAO is accompanied by various impacts on the climatological, ecological and economic systems. During extreme phases of the NAO with either enhanced, or reduced westerly flow, respectively, it is associated with the modulation of surface temperature, location and strength in storminess as well as the transport of moist air on large-scale

over the Atlantic-European area. The paths of synoptic storm systems for example which are located in a well defined region, the so-called storm-track, are shifted north-easterly during the positive NAO phase (Rogers 1990, Rogers 1997). More intense and frequent storms are found in the vicinity of Iceland and the Norwegian Sea (Serreze et al. 1997). With an increased storminess the ocean surface in turn responds with enhanced wave heights (Bacon and Carter 1993, Kushnir et al. 1997). An increase in wave height has major consequences for weather parameter dependent systems such as coast protection, offshore industries or shipping.

Changes in synoptic eddy activity inevitably modulate the transport of atmospheric moisture and thus are closely related to variations in regional winter evaporation and precipitation (Hurrell 1995, Dickson et al. 2000). Less precipitation and dryer conditions are found over Greenland and northern Canada as well as in the Mediterranean region during a positive phase of the NAO (Appenzeller et al. 1998) whereas wet conditions occur more frequently over Scandinavia and bordering regions.

Additionally, variations in the near surface winter temperature are related to a large extent to the variation of the NAO. During a positive phase warm air is advected over northern Europe and further downstream. Increased temperatures are found over the southeast of North America and the bordering ocean. Northerly winds with strong cold air advection, however, are found in Northeast Canada and Greenland, and further cooling is observed in north Africa and the Middle East (Appenzeller et al. 2000).

Further this temperature pattern is of special importance for the description of climate change. Since the heat storage over land is smaller than in the ocean, this pattern strongly contributes to the averaged northern hemispheric temperature variability (Hurrell 2003). Furthermore, much of the global temperature increase is found to have occurred over the northern continents during winter and spring (Folland and co-authors 2001), where the pattern is characterised by a 1-2°C warmer-than-average over the continents, while the surface temperature of the northern oceans are found to be colder-than-average. Such pattern in turn is strongly associated with changes in the atmospheric circulation. The latter are reflected by the positive phase of the NAO index in the Atlantic sector and an intensification of the Aleutian low pressure system which forms part of the PNA (Hurrell 2003).

Several authors provide modelling and observational evidence for a close linkage of climate change and the NAO (for details see Gillett et al. 2003). Although the increase in anthropogenic greenhouse gases represents the largest man-made forcing on the climate system, its influence on the NAO is still unclear (see Hartmann et al. 2000 for details). There is also evidence that tropical SST is of special relevance for the recent trends of the NAO (Hoerling et al. 2001). A possible mechanism is ascribed to an increase in the tropical rainfall over the Indio-Pacific and corresponding dryness over the equatorial

Atlantic and South-America. In a modelling study Bader and Latif (2003) found that the warming of the tropical Indian Ocean may contribute to the strengthening of the NAO during the recent decades.

NAO Predictability: The Climate Noise Scenario

The NAO index shows fluctuations on a wide range of timescales (Appenzeller et al. 1998). Under certain circumstances the NAO shows persistence for several winters. For example Figure 1-6 shows several decadal cycles with positive values from the beginning of the 20th century to the early 20ties and from the early 20ties to the mid 40ties. This is followed by a cycle of a strong negative phase and severe winters in Europe from the early 50ties till the late 60ties. In the late 20th century the NAO index has undergone a steady increase in amplitude. The significant frequency bands of the NAO are identified by a spectral analysis (Appenzeller et al. 1998, Greatbach 2000) with strong enhancement of the signal in quasi-biennial and decadal timescales (Barsugli and Battisti 1998, Deser and Blackmon 1993). The spectral analysis indicates a relative increase of power with decreasing frequency suggesting that the NAO behaves like an auto-regressive process (red noise process) distinctly different from a purely random process.

However, the NAO index also provides considerable variability even within a given season. Lately, it has been argued that the energetic weekly to monthly fluctuations in the mid latitudes are essential for the subsequent variability on interannual and longer timescales. In this so-called "climate noise paradigm" (Leith 1973, Madden 1976, Hasselmann 1976, Madden and Shea 1978, Madden 1981) the atmospheric long-term fluctuations are suggested to be entirely explained by processes intrinsic to the atmosphere. Intrinsic processes are defined as the daily variations of any state of the atmosphere. They are regarded as the unpredictable component, whereas the long-term fluctuations result from finite time-averages of daily fields which represent the predictable component.

Since the sample distributions of time-averaged atmospheric processes are not always independent, nearby values are correlated. This leads to a conditional estimate of sample averages, which is different from an estimation of an independent sample. The estimation of the variance of such time series inevitably leads to an underestimation of the variance of a sample distribution. In practice the variance of a N-day time average can be obtained by including a variance inflator factor Q, as follows:

$$\sigma^2{}_N = Q\frac{\sigma^2}{N}. \qquad (1.1)$$

Here σ^2 is the daily variance, σ_N^2 is the variance of an N-day time average and N the sample size (Wilks, 1995). The factor Q can be estimated, for example by a Markov process, assuming that the time series is represented by an autoregressive process AR(p), where p denotes the order. By using maximum likelihood estimators the variance of time averages of stationary time series can be estimated (for details see Jones 1975). The variance for a specific time average is given by

$$\sigma^2{}_N = \frac{\sigma^2}{N} Q = \frac{\sigma^2}{N}\left[1 + 2\left(1 - \frac{1}{N}\right)\alpha + 2\left(1 - \frac{2}{N}\right)\alpha^2 + \ldots + \frac{2}{N}\alpha^{N-1}\right], \qquad (1.2)$$

where α is the lag 1-day auto correlation. The ratio σ_N^2 / σ^2 reveals the percentage of variance that could be explained by an averaged time-series. Figure 1-8 illustrates the results of (1.2) for different time averages N. For example, in the case of monthly mean time series an autocorrelation value of $\alpha = 0.8$ (corresponding to a de-correlation time of nine days) indicates that about 25 % of variance is kept for longer periods. Feldstein (2000) investigates the importance of the AR(1) model and the intrinsic variability of the atmosphere to the NAO. Within the daily fields of 300hPa geopotential he fits an AR(1) process to the observed intra-seasonal spectrum of the NAO with a de-correlation time of about 19 days. Hence, for longer periods such as monthly and seasonal mean time series he finds a remaining variance of about 45 % and 20 %, respectively.

NAO-SST Feedback: Intraseasonal to Interannual Timescales

Coupled variations of the atmosphere and the ocean have early been suggested to be influential in the North Atlantic climate system (Bjerknes 1964). Components such as the heat capacity of the ocean, the SST variability and the ocean circulation are important factors for a complete description of atmospheric interannual circulation anomalies. Many observation and atmospheric modelling studies support the potential role of the sea surface temperature (SST) anomalies in the Atlantic for the mid-latitude atmospheric circulation (see Frankignoul 1985 and Kushnir et al. 2002 as review). Frankignoul (1985) and Palmer and Sun (1985), for instance relate the impact of the oceans to the mean seasonal mid-tropospheric geopotential height of about 20-80m and 10-20m, respectively (standard deviation is considered about 40-110m).

In a conceptual approach, Barsugli and Battisti (1998) confirm that the power spectrum of atmospheric temperature anomalies is to a large extent described by the rate of heat exchange with the ocean (reduced thermal damping paradigm). The adjustment time-scale of temperature anomalies in the ocean mixed-layer is approximately 1 month. Behind that the exchange of atmospheric and ocean heat anomalies is reduced. This

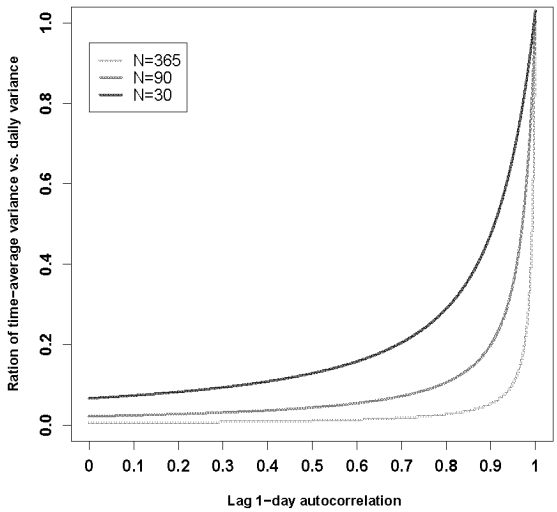


Figure 1-8: The ratio of time-average variance versus daily variance for monthly (N=30), seasonal (N=90) and yearly (N=365) mean (see text for details).

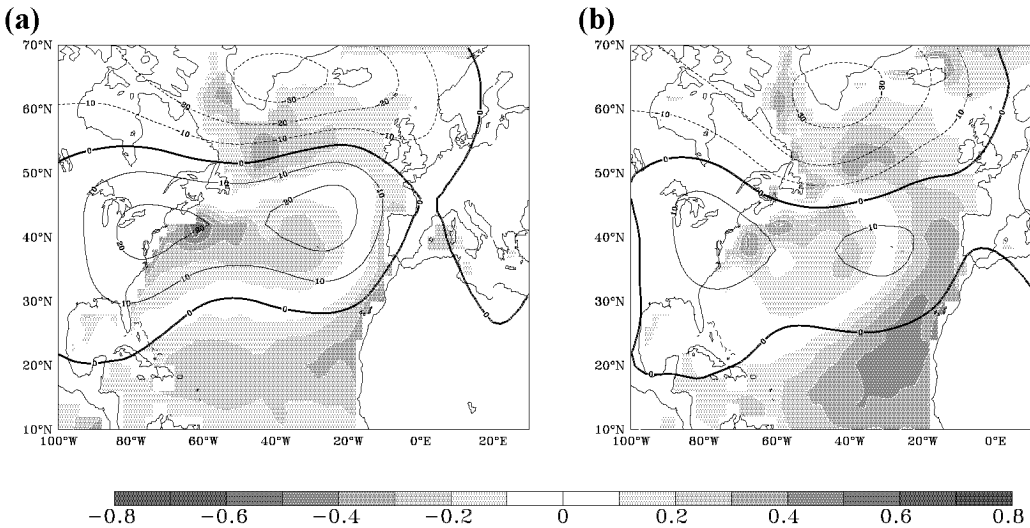


Figure 1-9: The first pair of eigenvectors of a SVD analysis of the NDJ geopotential height at 500hPa (contours 10hPa) with (a) subsequent DJF and (b) previous MJJ SST (shadings 0.1°C). The climatology is based on ERA40 reanalysis for the period 1959-2001.

would lead to a decrease of the thermal damping rate, which in turn is associated with deviations from the climate noise scenario. They assume this reduction to be responsible for the enhancement of the energy level in the spectrum of the atmospheric anomalies. With realistic parameters they could show that the reddening of the spectrum takes place on interannual and longer timescales and the energy level is about 75 % larger than in the climate noise scenario. These changes in the shape of the atmospheric temperature

spectrum indicate a certain degree of enhanced potential predictability on the corresponding time-scales.

The leading pattern of SST variability in the Atlantic appears as a tripole which reaches from the high latitudes to the sub-tropical Atlantic. Figure 1-9 shows the dominating pattern of co-variability between seasonal SST and geopotential 500hPa calculated with a singular value decomposition (SVD) analysis. The SST tripole is most active in late winter lagging the atmosphere by roughly one month (Figure 1-9a – see also Davis 1976, Davis 1978, Frankignoul 1985). This indicates the direct and stochastic response of the ocean to anomalous air-sea fluxes (Hasselmann 1976, Frankignoul and Hasselmann 1977, Cayan 1992).

However, there is increasing evidence of a positive feedback between the NAO and the underlying ocean on intra-seasonal to interannual timescales. In a lagged singular value decomposition (LSVD) analysis Czaja and Frankignoul (1999) and Frankignoul and Czaja (2002) examine the correspondence between the Atlantic SST and the geopotential at 500hPa for different time lags. They find statistically significant co-variance for a summer/autumn North Atlantic SST and winter NAO. When SST leads up to seven months a large-scale SST pattern is found with alternating signs southeast of Newfoundland and along the eastern North Atlantic, denoted as the North Atlantic Horseshoe pattern (NAH) (Figure 1-9b). The NAH is in such a way that an anomalous warm and cold SST off Newfoundland and the eastern Atlantic, respectively, are associated with a positive phase of the NAO index (Czaja and Frankignoul 1999, Drévillion et al. 2001). Czaja et al. (2003) argue that the NAH generates a NAO response, which in turn generates a response in the anomalous SST tripole. As pointed out by Peng et al. (2003) changes in baroclinicity and transient eddy activity are associated with changes in the meridional SST gradient of the NAH and hence reflect a direct interaction of the anomalous SST and the storm tracks. Furthermore, Peng et al. (2003) show that a NAO-like response is primarily maintained by an anomalous eddy forcing which results from the interaction of the heating-forced flow and the Atlantic storm-tracks.

Recent modelling studies (Rodwell et al. 1999, Rodwell and Folland 2002) support the idea of a forced NAO with proceeding SST. In coupled ocean-atmosphere GCMs they provided evidence of an ensemble-mean NAO pattern forced by summer SSTs. They suggest a low but significant correlation skill of the winter NAO based on the previous May SST of about 0.45. They also find a seasonal forecast skill throughout the annual cycle. Yet, some authors explain this skill as a result of the ensemble-mean statistic rather than of a physically based mechanism (Bretherton and Battisti 2000).

Finally, the dominant pattern of the SST tripole shows also a pronounced branch over the sub-tropical Atlantic. It should therefore be noted that there is also a close link of the

NAO to tropical climate variability in the Atlantic (for review see Marshall et al. 2001, Halliwell and Mayer 1996, Czaja et al. 2002).

NAO-SST feedback: Decadal to Inter-decadal Timescales

On decadal to inter-decadal timescales, there are also preferred modes of ocean heat transport variability which provide certain degrees of atmospheric predictability (Latif 1998, Grötzner et al. 1999). These particular modes involve either wind-driven ocean gyres (Latif and Barnett 1994, Grötzner et al. 1998), or thermohaline circulation (THC) (Delworth et al. 1993 ,Timmermann et al. 1998).

The mid-latitude decadal coupled mechanism involving the wind-driven ocean gyres for instance, is introduced by Latif and Barnett (1994), Latif and Barnett (1996) and Grötzner et al. (1998) for the North Pacific and Atlantic ocean, respectively. The decadal coupling describes the coexistence of SST anomalies in the sub-tropics and mid latitudes with large-scale atmospheric circulations such as the PNA and the NAO. The spatial pattern for the North Atlantic is similar to Figure 1-9a. The evolution of the SST shows a clockwise rotation of the anomalies around the sub-tropical gyre. The characteristic time-scale of this damped decadal oscillation amounts from 12y to 17y, as obtained from observational (Deser and Blackmon 1993) and modelling studies (Latif and Barnett (1996), Grötzner et al. 1998). Such gyre modes are inherently wind driven but the memory resides in the ocean (Latif and Barnett 1998). It has been suggested by many studies that the enhanced variability displayed by the SST tripole at the decadal band might reflect the oceanic impact (Bjerknes 1964, Battisti et al. 1995). It is found that the variations in the extension and intensity of the Gulf Stream are essential for the coupled oscillation (Latif and Barnett 1998, Czaja and Marshall 2001).

Inter-decadal coupled mechanisms typically involve the THC. On the one hand they include ocean-only modes which do not depend on a feedback with the atmosphere (e.g. Delworth et al. 1993, Griffies and Tziperman 1995, Delworth and Greatbach 2000). On the other hand Timmermann et al. (1998) found a coupled atmosphere-ocean mode with a close relationship between the THC and the NAO variability. They suggest that a strong North Atlantic THC leads to an increase in SST. An increase in SST corresponds to a strengthening of the NAO, which in turn produces a weaker than normal evaporation off Newfoundland. Less evaporation, however, is linked with negative anomalies of sea surface salinity and a weakening of the deep convection in the sinking region. This leads to reduced THC, which again decreases the poleward heat transport. The characteristic timescale of this oscillation is found to be about 35y.

Chapter 2:

Seasonal Climate Forecasting

Long-range forecasting with seasonal outlook is defined as the "description of averaged weather parameters expressed as a departure from climate values for a specific season"[2], where a season is defined as a period which encompasses several months. A prediction of weather parameters and their averaged quantity for such a period, however, always contains uncertainties, which are related to the intrinsic chaotic nature of the system and to the model formulation. Intrinsic uncertainties reduce the prediction of a system. Forecast skill can be enhanced by using ensemble forecasts and forecasts of multiple models, respectively. These forecasts are probabilistic in nature and require probabilistic verification strategies. Hence in the first section (2.1) of this chapter the concept of probabilistic prediction is introduced including a brief description of the sources of uncertainty and the probabilistic forecast verification.

The remainder of the chapter is reserved for the description of the forecast systems. In this study mainly dynamically based forecast systems are used, namely the ECMWF Seasonal Forecast System 2 (2.2) and the joint European Multi-Model DEMETER system (2.3).

2.1. Probabilistic Prediction

Sensitivity to Initial Conditions

Predictions within complex dynamic systems such as the ocean-atmosphere are inherently related to the sensitivity to its initial conditions. It is known from the dynamical system theory that the growth of errors, introduced by observations, are usually linked to enhanced instability in phase space (Lorenz 1963, Guckenheimer and Holmes 1983, Argyis et al. 1995). Two close initial conditions may have similar solutions after a short time of integration. The passage through an unstable region in phase space forces them to disperse, so that their final states indicate totally different solutions and hence no prediction can be made. Therefore a single solution produces not necessarily a useful forecast.

[2] http://www.wmo.ch/web/www/DPS/GDPS-Supplement5-AppI-4.html

An appropriate approach which accounts for the uncertainty due to initial conditions is the stochastic-dynamical approach introduced by Epstein (1969). He derived an equation for the evolution of a probability density function (PDF) governed by a simple dynamical model. The stochastic-dynamical approach which was introduced for a low-order model with an infinite set of ensemble members was found to be impractical for models with high degrees of freedom. Monte-Carlo methods provide a more practical method. Rather than solving the equations for the whole distribution of initial conditions, the forecast is based on the behaviour of a finite size of trajectories. Each trajectory is defined as a single solution of the system. This type of method is called ensemble forecast and was first introduced by Leith (1974).

A forecast is finally attained by a statistical post-processing of the ensemble. The ensemble mean for instance is commonly used because it removes the unpredictable noise. A large or small ensemble spread further indicates whether the forecast uncertainty is high or low (Scherrer et al. 2004). And finally the full distribution of the ensemble can be used to give a quantitative probabilistic forecast.

Sensitivity to Boundary Conditions

But reliable predictions on seasonal timescales are not only possible by using ensemble forecasts alone. The variability of synoptic disturbances for instance is able to obscure the signal on seasonal timescales. A seasonal climate outlook becomes only feasible if the long-term variability associated with e.g. sea-surface temperature or land-surface characteristics forces the atmospheric system in such a way, that significant shifts in the probabilities of low-frequency variability occur (Palmer 1993).

The response of a nonlinear system such as the atmosphere to an external forcing (i.e. the ocean) is illustrated by the experiment shown in Figure 2-1. The resulting probability density function of the system is described by the probabilities of finding balls in the corresponding cups. With no external forcing, the balls end up with a 50 % chance in either the left or right cup (Figure 2-1a). This is what we expect if we consider the internal atmospheric variability alone. If we further assume a Gaussian distributed climatology, any prediction based on such a model solution would have no further benefit compared to a climatological based prediction. This situation is different if a weak external forcing (the airflow does only influence the falling balls) affects the atmospheric system (Figure 2-1b). This corresponds, for example, to an event such as El Niño. Now, the balls are preferably (say with a chance of 70 %) placed in the left cup, which lies in the forcing direction. The probability distribution of the system is somewhat skewed and differs from the climatological distribution. Such a forecast provides better results than a forecast based in climatology.

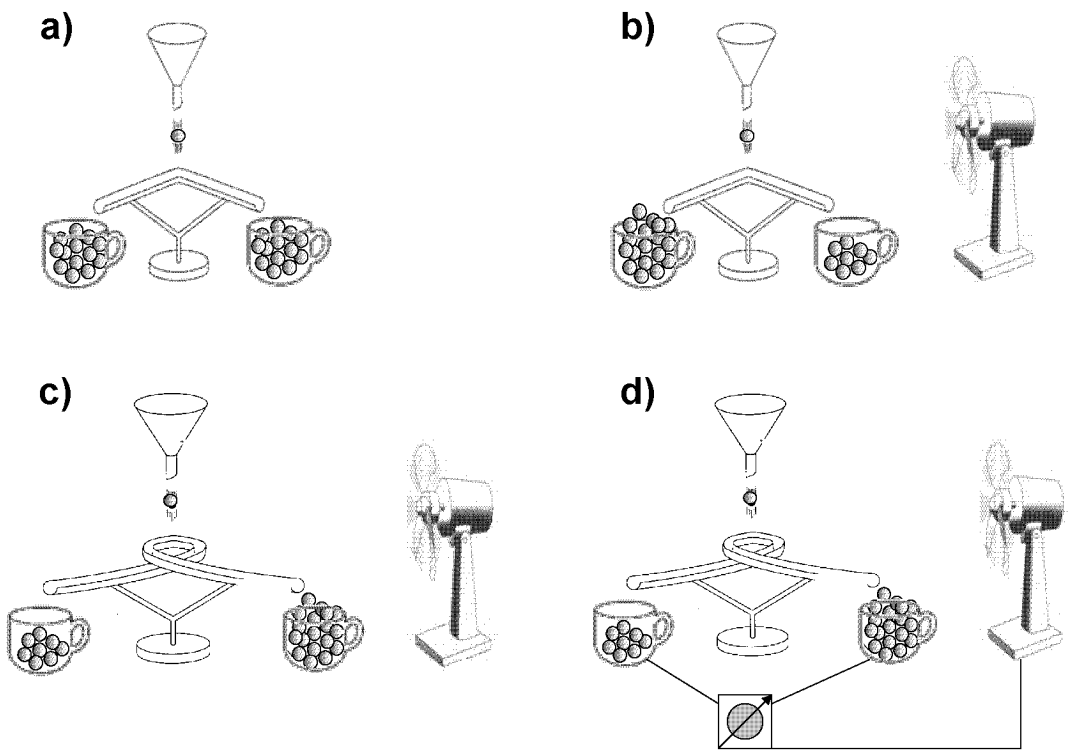


Figure 2-1: An experimental set up of a coupled atmosphere-ocean toy system. The atmospheric system and its possible solutions are described by a ridge and underlying cups, respectively. Initially disturbed balls mark possible synoptic weather systems. The ocean is described by a fan and the coupling is demonstrated by a switch connected to the fan. The figure shows (a) the atmosphere only, (b) atmosphere and ocean uncoupled, (c) nonlinear atmosphere and ocean uncoupled and (d) nonlinear atmosphere and ocean coupled (adapted from Palmer, 1993).

In a highly nonlinear system the solution trajectories become more complicated. This is shown in (Figure 2-1c) with the channels reversed. Again the dropping balls are influenced by the airflow, but here there is a higher probability to find a ball in the upstream cup. Furthermore, in a system such as the atmosphere-ocean, forcing and response processes may be coupled in either a positive or negative way (Figure 2-1d). In a positive feedback loop for example the right cup would send a signal that decreases the amplitude of the forcing instrument. In turn this forces the balls to be further placed in the left cup.

As pointed out by Palmer (1993) the response of a dynamical system such as the ocean-atmosphere to an external forcing is primarily manifested in the relative frequency of the associated regimes. He argues that the location of distinct quasi-stationary regimes is insensitive to the forcing, and that the influence due to the forcing is localized in space

and time. This is of particular interest for seasonal to interannual flow regimes such as the PNA and the NAO (see section 1.3 and 1.4).

Probabilistic Forecast Verification

In a general sense an objective evaluation of the forecast quality makes use of the joint distribution of the probability function of forecasted values y_i and the observed values o_i. For practical use the forecast quality is often summarised by scalar measures. For meteorological applications the binary Brier Score (BS) and the multi-categorical Ranked Probability Score (RPS) are the most common used measures. The BS is defined as follows:

$$BS = \frac{1}{N}\sum_{k=1}^{N}\left(y_k - o_k\right)^2 . \tag{1.3}$$

The RPS as:

$$RPS = \frac{1}{N}\sum_{k=1}^{N} RPS_k = \frac{1}{N}\sum_{k=1}^{N}\sum_{m=1}^{J}\left[\left(\sum_{i=1}^{m} y_{i,k}\right) - \left(\sum_{i=1}^{m} o_{i,k}\right)\right]^2 . \tag{1.4}$$

Here N denotes the number of forecast/observation pairs and J denotes the number of the equiprobable classes. The BS averages the squared differences between pairs of forecast and observation probabilities. The RPS averages the squared difference between the cumulative probabilities of a forecast and an observation over a number of equiprobable classes J. If the observation lies in class i, the cumulative probability of the observation is unity in such case and subsequent cases. For classes prior to the class i the cumulative probabilities are zero. For two equiprobable classes the RPS is similar to the BS. For a perfect forecasting system the BS and the RPS are zero. The scores are defined with the quadratic norm and are comparable to the mean squared error (MSE). However, they are taken in probability space and not in physical space, as is the case for the MSE.

The forecast performance is usually quantified with a skill score (SS). Formally, the skill score describes the benefit of a forecast score S over the reference forecast score S_{ref} relative to a perfect forecast score S_{perf}, i.e.:

$$SS = \frac{S - S_{ref}}{S_{perf} - S_{ref}} \times 100 \tag{1.5}$$

Usually the reference forecast is a climatological, persistence or random forecast. A positive skill score indicates an improvement with respect to the reference forecast.

The BS and the RPS can be decomposed into scalar attributes as follows (a full description of the decomposition is given in Murphy 1973 or Stephenson and Jolliffe 2003):

$$BS = \underbrace{\frac{1}{N}\sum_{k=1}^{J} N_k (y_k - \overline{o}_k)^2}_{\text{reliability}} - \underbrace{\frac{1}{n}\sum_{k=1}^{n} N_k (\overline{o}_k - \overline{o})^2}_{\text{resolution}} + \underbrace{\overline{o}(1-\overline{o})}_{\text{uncertainty}}, \tag{1.6}$$

where $\overline{o}$ is the relative frequency of the observations, N_k the number of times each forecast is used in the collection of forecast being verified and $\overline{o}_k$ is the conditional average observation.

The *reliability* (REL) term is defined as the squared difference between the forecast probability and the conditional distribution in the different probability categories. It describes the relationship between the forecasts and the average observation of specific forecasts (i.e. is it usually 20°C when the temperature forecast is 20°C?).
The *resolution* (RES) term is defined as the average square difference between the conditional distribution in each probability category and the relative frequency observed in the whole sample. It compares the average of observations for one set of forecasts with another (i.e. is it usually warmer when the forecast is 25°C compared to when the forecast is 20°C?).
The *uncertainty* (UNC) term is independent of the forecast and is only related to the relative frequency of the observations. Hence it is the reliability and the resolution term that determine the forecast performance.

The Brier Score for a climatological forecast is equal to the uncertainty term, since the resolution term and the reliability term are both zero (Wilks 1995). Hence using a climatological forecast as a reference forecast in the calculation of the Brier Skill Score (BSS) can be written as:

$$BSS = \frac{RES - REL}{UNC}. \tag{1.7}$$

From this equation follows, that positive skill scores are achieved if the RES term > REL term (the UNC term is always positive). A similar equation holds for the Ranked Probability Skill Score (RPSS, see Hersbach 2000). This equation will be used in chapter 3 where the sensitivity of the BSS and RPSS regarding the ensemble size is discussed.

Finally there are desirable characteristics of a skill score which are related to the consistency of a forecast verification statistic. For example a *strictly proper* score is necessary to avoid that better scores can be achieved by artificially forecasting always the same probability values. *Equitability* of a scoring rule is also important. It requires random forecasts or constant forecasts to have the same expected no-skill value (Murphy

and Daan 1985). The BS and the RPS for example are strictly proper, but are not equitable as will be shown in chapter 3.

In this thesis the forecasts are verified using the RPSS. In chapter 3 we turn to the consistency problem linked to the RPSS and present a probabilistic verification statistic which is equitable and strictly proper.

2.2. The ECMWF Seasonal Forecast System 2

Ocean and Atmosphere Model

The ECMWF Seasonal Forecast System 2 is an ensemble prediction system (EPS) which consists of an ocean and an atmospheric model as well as a coupling strategy which interpolates between the two components. The prediction period of System 2 is up to 6 months. A complete description is found in Anderson et al. (2003) and on the ECMWF webpage[3].

The ocean model is the Hamburg Ocean Primitive Equation model (HOPE) version 2 (Latif et al. 1994, Wolff et al. 1997). The ocean model is run on an Arakawa E grid with a zonal resolution of 1.4° degrees and a meridional resolution of 0.3 degrees to 1.4 degrees grid spacing. The model has a total of 29 vertical levels. Since the model is run on a sparse grid, several physical processes are not resolved explicitly but parameterized (for details see Anderson et al. 2003). Furthermore a pseudo ice-model is used over polar regions. In the forecast mode the sea-ice is embedded with a 60 day relaxation time to climatology. Monthly climatological river runoffs are included as well.

The atmospheric model is the ECMWF Integrated Forecast System (IFS) version 23r4[4] with 40 vertical levels (see also Gregory et al. 2000). The horizontal representation employs a spectral grid resolution (T95) for the dynamical part and a Gaussian grid of 1.875 degrees for model parameterizations. A two time-level semi-Lagrangian scheme with a 1-hour step is applied. Physical processes are treated in the state-of-the-art fashion (Anderson et al. 2003). Furthermore, four sub-surface soil levels are included with prognostic moisture and temperature as well as a vegetation canopy.

The ocean and the atmosphere model are associated with different grids and coastlines. An Ocean Atmosphere Sea Ice Soil (OASIS) coupler is applied to interpolate between these two grids. The atmosphere passes surface fluxes of heat, momentum, precipitation and moisture once a day. The ocean returns its instantaneous SST.

[3] http://www.ecmwf.int/products/forecasts/seasonal/documentation

[4] http://www.ecmwf.int/research/ifsdocs/CY23r4

Finally, a sub-surface ocean data assimilation scheme is introduced in the model. The assimilation scheme is a univariate temperature optimum interpolation (Smith et al. 1991), which is evaluated on sub-domains on the horizontal model grid. The interpolation is used for each level in the first 1000m, with exception of the top layer. The impact of sub-surface ocean data assimilation has shown to improve dynamical forecasting such as ENSO (Alves et al. 2003).

Ocean and Atmosphere Initial Conditions

The initial conditions of the ensemble members in the ocean model are generated by a perturbation technique that accounts for wind stress and SST anomalies. These perturbations include uncertainties deriving from the assimilation, observations and time interpolation schemes. Unlike earlier systems, the wind stress estimates use analysed winds instead of forecasted stresses from the atmospheric numerical weather prediction system. The ocean model provides an ensemble of five ocean analyses, which include a measure of uncertainty imported by the surface winds. The forcing fields are contributed by the ERA15 Re-analysis and operational analysis from 1994 onwards.

The atmospheric initial conditions are also introduced by ERA15 and operational analysis from 1994 onwards. A total of five ensemble members were generated for a 15 years hindcast period (1987-2001), giving a total climatology consisting of 75 ensemble members. For forecasts starting in May and November 40 ensemble members were generated. This holds also for all the operational runs from August 2001 onwards. Unlike the Seasonal Forecast System 1, all ensemble members are generated at the first day of each month. The ensemble members are constructed by combining the five ensemble members from the ocean run with different SST perturbation and a stochastic forcing. This takes into account a greater and thus more realistic spread in the initialization.

Bias and Drift

Model deficiencies cause the climate model to generate systematic errors known as bias and drift (Arpe and Klinker 1986, Tibaldi et al. 1990). Bias is defined as the difference between the model and the observed climatology for a given lead time, whereas the drift is determined as the time dependent difference of the bias for different lead times. Different strategies exist to assess systematic errors in climate models (e.g. Molteni and Buizza 1999, Vidale et al. 2003). Vidale et al. (2003) for example use the interannual variability to evaluate a high-resolution climate model (HCRM, Schär et al. 1999).

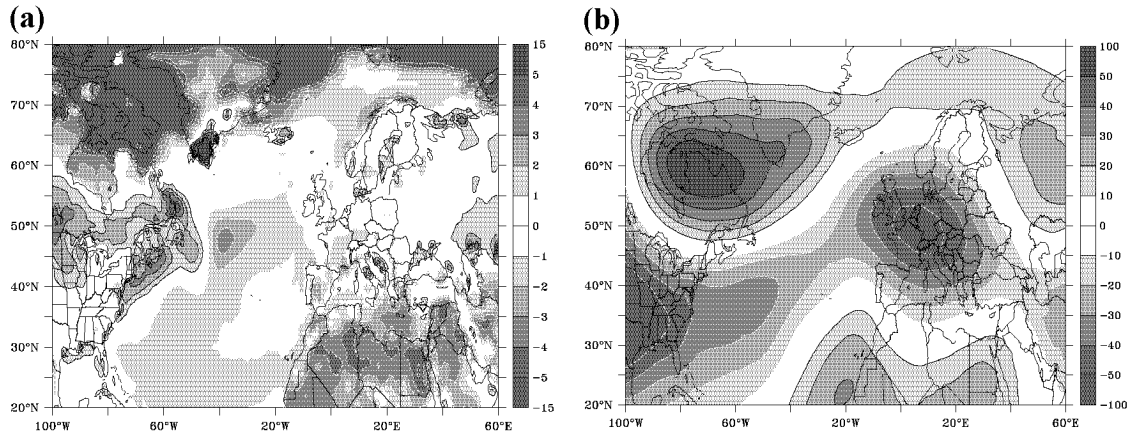


Figure 2-2: The ECMWF system 2 bias of (a) the winter mean 2m temperature (shadings 1°C interval) and (b) geopotential 500hPa (shadings 10gpm interval) for lead time 3 months. The ERA40 Re-analysis was used as observations.

The two basic error properties of the ECMWF system 2 are examined briefly. Figure 2-2 shows the bias for the winter mean 2m temperature (Figure 2-2a) and the geopotential at 500hPa (Figure 2-2b) against ERA40 for a lead time of three months. The pictures reveal strong bias in particular for the 2m temperature. In the central Atlantic there is a large-scale bias of -1°C to -3°C. The large-scale bias suggests a systematic failure of the mid-Atlantic ocean-circulation, in particular regarding the Gulf-Stream location. Further strong negative biases up to -3°C are found over the northern African continent. The negative bias exceeding -5°C in the polar region, suggesting a major error with the sea ice model. A strong positive bias is found over the northern American continent with values ranging from 1°C to 3°C, whereas over Central Europe the temperature bias is mostly around zero. Since the error amplitude is around the magnitude of the variability on seasonal time-scales, the bias is a major drawback for real applications and needs to be corrected. The 500hPa geopotential (Figure 2-2b) shows an enhanced negative bias over central Europe and over continental North America with magnitudes up to 60hPa. A similar positive bias is found over Newfoundland.

The model drift for the winter mean near surface temperature and mid-level geopotential is shown in Figure 2-3. Here the drift is determined as the difference between the bias for lead time 3 months and the bias for lead time 0 months. There is a positive drift of the 2m temperature (Figure 2-3a) over the Eurasian and north American continents with a value of about 1-2°C. A massive negative drift with magnitude >2.5°C is found over the polar oceans. The drift for the mid-tropospheric circulation (Figure 2-3b) is dominated by a large positive signal over Newfoundland and the Middle East. The magnitudes are around 10-20hPa.

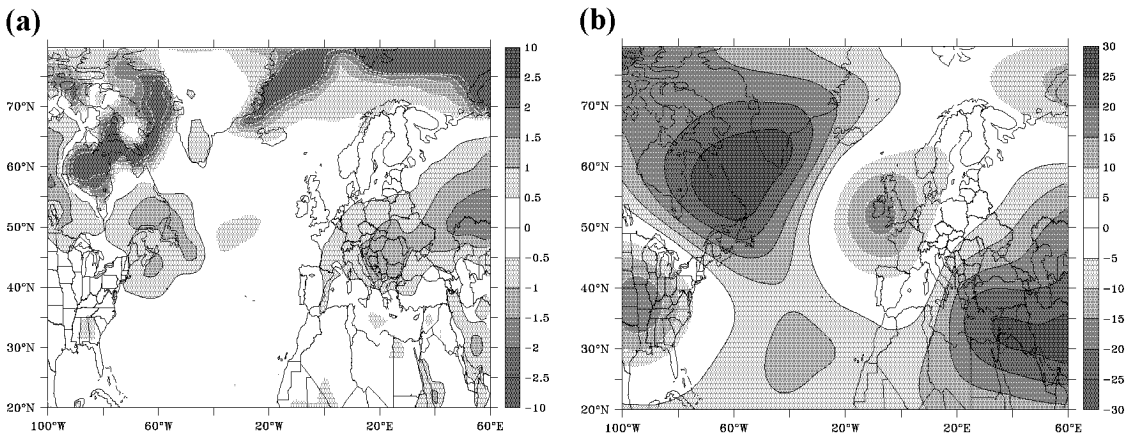


Figure 2-3: The drift of the ECMWF System 2 of (a) the winter mean 2m temperature (shadings 0.25°C interval) and (b) geopotential 500hPa (shading 10gpm interval). Shown are the differences between lead time 3 months and lead time 0 months.

Climate Mean Replacement & Probabilistic Correction

Observed seasonal climate anomalies are typical of similar magnitude as the model bias (for example, temperature anomalies amount to about 1°-2°C). Post-processing is needed to remove the model bias. In this thesis two simple approaches are used to correct the bias.

A first order correction is applied to the climatological mean and is termed climate mean replacement (CMR). Here the model anomalies are determined with respect to the model climatology at every lead time and for all start months. This is necessary in order to correct the time dependent drift. The CMR can be used for applications requiring absolute values by adding the observed climatology to these model anomalies. Figure 2-4 illustrates the concept of the CMR.

Second, a probabilistic correction is applied to the forecasts. Here the model and the observation anomalies are compared in terms of equal probabilities. If the occurrence of a forecasted anomaly of 2°C has a 70 % chance, for example, then the forecasted anomaly is compared to an observation anomaly having a 70 % chance of occurrence. The probabilities of the observations and forecasts are defined by using their individual distributions. If the observation and model anomalies are Gaussian distributed then the probabilistic correction can be applied by standardisation of the observation and model distribution.

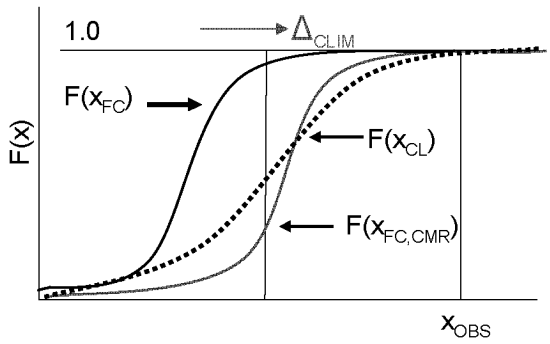


Figure 2-4: The climate mean replacement (CMR) illustrated for a single forecast $F(x_{FC})$ and a given climatological distribution $F(x_{CL})$. The CMR shifts the cumulative forecast distribution by Δ_{CLIM} closer to the observation x_{OBS}. The replaced forecast distribution is given by $F(x_{FC,CMR})$.

2.3. The Multi-Model System (DEMETER)

In the last years several projects have assessed the benefit of multi-model configurations. In the European project PROVOST (Prediction of Climate Variations on Seasonal to Interannual Timescales) a multi-atmospheric-model-ensemble based on prescribed observed SST was created to quantify seasonal predictability (Palmer and Shukla, 2000). PROVOST was launched from 1997-1999 to develop the scientific basis and the means for seasonal forecasting. The two major results of PROVOST can be summarised as: (I) Despite using identical SST boundary conditions, the models have shown a large variation in the internal and external forced variability. This has also been demonstrated in the Dynamical Seasonal Prediction (DSP) project undertaken at the same time in the United States (Shukla 2000). (II) The multi-model approach has proven to have a higher forecast skill than the single–model approach (Doblas-Reyes et al. 2000).

The DEMETER (Development of a European Multi-Model Ensemble Prediction System for Seasonal to Interannual Prediction) project has been developed as an extension to PROVOST (Palmer et. al, 2004, Hagedorn et al. 2004, Doblas-Reyes et al. 2004). DEMETER was designed to produce and combine state-of-the-art coupled atmosphere-ocean models to a series of multi-model hindcast ensembles. The focus was on seasonal climate prediction but there have been selected applications in the area of agronomy and tropical diseases. One of the major results has been the enhanced probabilistic skill of seasonal malaria incidence, obtained by a dynamical malaria model initialised with forecasts of the DEMETER system (Morse et al. 2003). Further efforts have been made in probabilistic downscaling of seasonal forecasts to produce crop yield estimations. A complete description of the DEMETER system can be found in Palmer et al. (2004) and on the ECMWF web-page[5].

[5] http://www.ecmwf.int/research/demeter

With twelve partners contributing to the project, seven were extensively developing the comprehensive CGCMs. The model-developing partners are summarised in Table 2-1. The DEMETER system consists of a total of seven combinations of coupled atmosphere-ocean models running in full ensemble mode. The integrations start four times a year (Feb, May, Aug, and Nov) with a common hindcast production period from 1980-2001. Each of the models contains a set of nine ensemble members with 6-monthly hindcast integrations. Prior to conducting the forecast runs, three ocean analyses are initiated with ERA40 flux forcing. The resulting SST is finally perturbed by superposed symmetric anomalies giving nine ensemble members. The atmospheric initial conditions come from the ERA40 (details see Table 2-1). In this thesis combined DEMETER data for the period 1987-2001 are used for forecasts starting in May and November. In chapter 5 the long hindcast period 1959-2001 of the ECMWF model is analysed for forecasts starting in November.

	Météo-France CNRM	Max-Planck Institute SMPI	MetOffice UKMO	ECMWF SCWF	INGV SCNR	LODYC	CERFACS CRFC
Atmospheric Component	ARPEGE	ECHAM-5	HadAM3	IFS	ECHAM-4	IFS	ARPEGE
Resolution	T63 31 Levels	T42 19 Levels	2.5°-3.75° 19 Levels	T95 40 Levels	T42 19 Levels	T95 40 Levels	T63 31 Levels
Atm. Initial Conditions	ERA40	Coupled run relaxed to observed SST	ERA40	ERA40		ERA40	ERA40
References	Deque 2001	Röckner 1996	Pope et al. 2000	Gregory et al. 2000	Röckner 1996	Gregory et al. 2000	Deque 2001
Ocean Component	OPA 8.0	MPI-OMI	GloSea	HOPE-E	OPA 8.1	OPA 8.2	OPA 8.2
Resolution	182GPx152GP 31 Levels	2.5°x0.5°-2.5° 23 Levels	1.25°x0.3°-1.2° 40 Levels	1.4°x0.3°-1.4° 29 Levels	2.0°x0.5°-1.5° 31 Levels	2.0°x2.0° 31 Levels	2.0°°x2.0° 31 Levels
Oc. Initial Conditions	ERA40	Coupled run relaxed to observed SST	ERA40	ERA40	ERA40	ERA40	ERA40
References	Madec et al. 1997	Marsland et al. 2002	Gordon et al. 2000	Wolff et al. 1997	Madec et al. 1998	Delecluse and Madec 1999	Delecluse and Madec 1999
Ensemble Generation	wind stress and SST	lagged method	wind stress and SST	wind stress and SST	wind stress and SST	wind stress and SST	wind stress and SST
Hindcasts	1958-2001	1969-2001	1959-2001	1958-2001	1973-2001	1974-2001	1980-2001

Table 2-1: Models used in the DEMETER system with information about their atmospheric and ocean components, initial conditions and ensemble generation. For details see Palmer et al. (2003).

Chapter 3:

A Debiased Ranked Probability Skill Score to Evaluate Probabilistic Ensemble Forecasts with Small Ensemble Sizes [6]

Müller[7], W. A., C. Appenzeller[8], F. J. Doblas-Reyes[8] and M. A. Liniger[8]

Abstract:

The ranked probability skill score (RPSS) is a widely used measure to quantify the skill of ensemble forecasts. The underlying score is defined by the quadratic norm and is comparable to the mean squared error (MSE), but is applied in probability space. It is sensitive to the shape and shift of predicted probability distributions. However, the RPSS shows a negative bias for ensemble systems with small ensemble size as recently shown. Here, two strategies are explored to tackle this flaw of the RPSS. First the RPSS is examined for different norms L ($RPSS_L$). It is shown that the $RPSS_{L=1}$ based on the absolute rather than the squared difference between forecasted and observed cumulative probability distribution is unbiased. $RPSS_L$ defined with higher order norms shows a negative bias. However, the $RPSS_{L=1}$ is not strictly proper in a statistical sense. A second approach is then investigated, which is based on the quadratic norm, but with the reference forecast reduced to sub-samples of the same size as the forecast ensemble size. This technique results in an unbiased and proper skill score which is denoted as the de-biased ranked probability skill score ($RPSS_D$) hereafter. Both newly defined skill scores are independent of the ensemble size whereas the associated confidence intervals are a function of the ensemble size and the number of forecasts.
The $RPSS_{L=1}$ and the $RPSS_D$ are then applied to the winter mean (DJF) near surface temperature predictions of the ECMWF Seasonal Forecast System 2. The overall structure of the $RPSS_{L=1}$ and the $RPSS_D$ are more consistent and largely independent of the ensemble size, unlike the $RPSS_{L=2}$. Further the minimum ensemble size required to predict a climate anomaly, given a known signal-to-noise ratio is determined by employing the new skill scores. For a hypothetical set up comparable to the ECMWF hindcast system (40 members, 15 hindcast years) statistically significant skill scores were only found for a signal-to-noise ratio larger than ~0.3.

---

[6] Accepted by J. Climate, revised version 13 April 2004

[7] Swiss Federal Office of Meteorology and Climatology (MeteoSwiss), Zürich, Switzerland.

[8] European Centre for Medium-Range Weather Forecast (ECMWF), Reading, UK

3.1. Introduction

In recent years probabilistic ensemble forecast systems have been established in a wide area of applications. The probabilistic nature of these forecasts requires verification techniques based on probabilistic skill measures. However, a general consent on the best skill score does not exist. The choice depends on the particular application considered or the forecast system being used. Examples are Brier Scores or the Relative Operating Characteristics (for details see Swets 1973; Mason 1982; Stephenson and Jolliffe 2003; Wilks 1995). The Brier Score, for instance, is essentially the mean squared error of the probability forecast of a dichotomous event. An example is a probability forecast of the winter mean temperature to be above or below the climatological mean (Palmer et al. 2000). For a range of applications, such a categorical score gives an incomplete picture since not the entire shape of the probability function is considered. A multi category score that measures the shape as well as the central tendency of the whole probability density function (PDF) is more eligible. An often used score for such applications is the ranked probability score (RPS) (Epstein 1969; Murphy 1969, 1971).

The RPS is based on the cumulative density function (CDF) and classically defined by the quadratic norm, hereafter denoted as $RPS_{L=2}$. The score is the integrated squared difference between the forecasted and the observed CDF. It can be seen as the probabilistic extension of the mean squared error (MSE). However, the $RPS_{L=2}$ is applied in the cumulative probability space and not in the physical space, i.e. the integration is taken over categories. It can be interpreted as an extension of the Brier Score for finite ordered categories. The extension to an infinite number of classes results in the continuous RPS (Unger 1985; Hersbach 2000)

The current ensemble prediction systems (EPS) for medium range forecasts (3 to 10 days) use ensemble sizes varying from 17 (NCEP) to 50 (ECMWF) members to construct the probability density function (Toth and Kalnay 1993, Tracton and Toth 1993, Buizza et al. 1998). For long-range forecasts, with prediction times of months to years, the ensemble size is usually smaller. The reason for having small ensemble sizes lies primarily within computational costs. This is particularly true for hindcast experiments that are used for verification and calibration. For example, the 15 years hindcast database of the ECMWF Seasonal Forecast System 2 mostly consists of 5 ensemble members.

Obviously, such a small ensemble size (and number of forecasts) is a potential statistical problem. In the frame of numerical weather prediction Buizza and Palmer (1998) have shown that the skill can be improved by increasing the ensemble size, but the extent to which improvement occurs depends on the measure used. For the $RPSS_{L=2}$ they found major skill improvement for at least up to eight members. In a perfect seasonal forecast approach Dequé (1997) determined the ensemble size required for skill score saturation

for various parameters. An ensemble size of 40 was suggested for European temperature forecasts. Kumar et al. (2001) have also explored the influence of the ensemble size on the $RPSS_{L=2}$ and noted that the $RPSS_{L=2}$ is strongly negative biased for a small ensemble size. Based on this bias the minimum ensemble size which is required to predict a climate signal given a known signal-to-noise ratio was derived (mean shift of the anomaly distribution in standardised units).
Here we show that the substantial negative bias of the $RPSS_{L=2}$ for small ensemble sizes is primarily a consequence of the discretization and squaring measure in its formulation. Two strategies are introduced that overcome these deficiencies (section 3.2). The characteristics of the suggested modifications of the $RPSS_{L=2}$ are examined with a synthetic example (section 3.3). The new techniques are then applied to a real seasonal winter temperature forecast based on the ECMWF Seasonal Forecast System 2 for the years 1987-2001 (section 3.4) and to the climate signal-to-noise detection problem (section 3.5). Finally, a conclusion and discussion are given in the final section.

3.2. Definition of $RPSS_L$

Similar to the mean squared error (MSE) the $RPS_{L=2}$ is a quadratic measure and thus larger deviations from the actual probability are penalised much stronger than smaller ones. The RPS_L for any norm L is defined as $RPS_L = 1/N \sum_{k=1}^{N} RPS_{L,k}$ where N is the number of forecasts, k is the forecast index and:

$$RPS_{L,k} = \sum_{j=1}^{J} \left| Y_j - O_j \right|^L . \tag{3.1}$$

Here J denotes the number of the equiprobable classes, L is the norm. The cumulative probabilities of the forecasts Y_j and the observations O_j are defined as $Y_j = \sum_{i=1}^{j} y_i$ and $O_j = \sum_{i=1}^{j} o_i$, where y_i and o_i are the probability of the forecast and observation, respectively, for the class i. The RPS_L is zero in case of a perfect forecast and positive otherwise. The calculation of the skill score is based on the comparison of the forecast score ($RPS_{L,FC}$) to a reference forecast score ($RPS_{L,CL}$) relative to a perfect forecast. Thus the $RPSS_L$ becomes:

$$RPSS_L = 1 - \frac{\sum_{k=1}^{N} RPS_{L,FC,k}}{\sum_{k=1}^{N} RPS_{L,CL,k}} . \tag{3.2}$$

Any positive value of the $RPSS_L$ indicates a forecast benefit compared to the reference forecast. For L=2 the skill score becomes the standard squared definition (Wilks 1995), whereas L=1 gives the absolute skill score $RPSS_{L=1}$. In this study the distribution of the reference forecast is defined by a Gaussian fit to the observations. The equiprobable classes are obtained from the distribution of the reference forecasts.

3.3. A Synthetic Example

In the following a white noise climate system is employed in order to explore the sensitivity of the $RPSS_{L=2}$ to the ensemble size. Forecasts and observations are chosen to be random Gaussian time series. In such a system there is nothing to predict by definition. Whatever skill score is used, the expected outcome should be a value that gives no benefit compared to a reference forecast. The white noise climate consists of a sample of 100.000 cases. For the skill scores, 300 sub-sets are randomly chosen and verified against randomly chosen observations. This provides for a robust estimate of the skill score. This procedure is repeated 100 times for 100 chosen observations. This gives 100 skill scores from which the mean and the 95 % confidence intervals are calculated. These confidence intervals are chosen as guidance for further testing of the real seasonal forecasts against random time series. Other significance tests exist and a more complete discussion is found in Wilks (1995) and Nichollis (2001). In the light of a potential application to the ECMWF forecast system, the skill scores are calculated with up to 40 ensembles members and a number of 15 forecasts. Three equiprobable classes (above, normal, below) are used.

Figure 3.1a shows the dependence of the $RPSS_{L=2}$ on the ensemble size for white noise climate forecasts. The mean of the $RPSS_{L=2}$ exhibits negative skills ranging from -0.20 up to -0.02 for an ensemble system of size 5 and 40, respectively. These values are below the expected value of zero benefit. With a larger ensemble size the bias decreases slowly towards zero. The 95 % confidence intervals (thin lines) are asymmetric and vary from about -0.45/0.05 for a 5-member system, to about -0.10/0.05 for a size of 40. The confidence intervals as a function of the number of equiprobable classes are shown in Fig. 3.2a. The confidence intervals of a 40-member forecast system (black) are much closer to the mean values than for a 5-member system (grey). A strong asymmetry around zero is visible for a small ensemble size. The magnitudes of the confidence intervals are largest for 2 classes and become independent of the number of classes from 5 onwards. Finally, the bias and confidence intervals are evaluated as a function of the number of forecasts (Fig. 3.3). The confidence intervals are closer to the mean for a higher number of forecasts. However, the bias remains unchanged indicating that it is related to the ensemble size.

The corresponding results for the $RPSS_{L=1}$ are shown in Fig. 3.1b and Fig. 3.2b. The mean of the $RPSS_{L=1}$ (Fig. 3.1b) is located close to zero even for the smallest ensemble size, e.g. for a 5-member system. The 95 % confidence intervals are spread symmetrically around the zero mean with a magnitude of about ± 0.14. For a large ensemble size (40) the spread is about ± 0.05. The confidence intervals are now symmetric around zero for all classes (Fig 3.2b) and only depend on the ensemble size, with reduced spread for larger samples. Obviously, the $RPSS_{L=1}$ has no bias for small ensemble sizes.

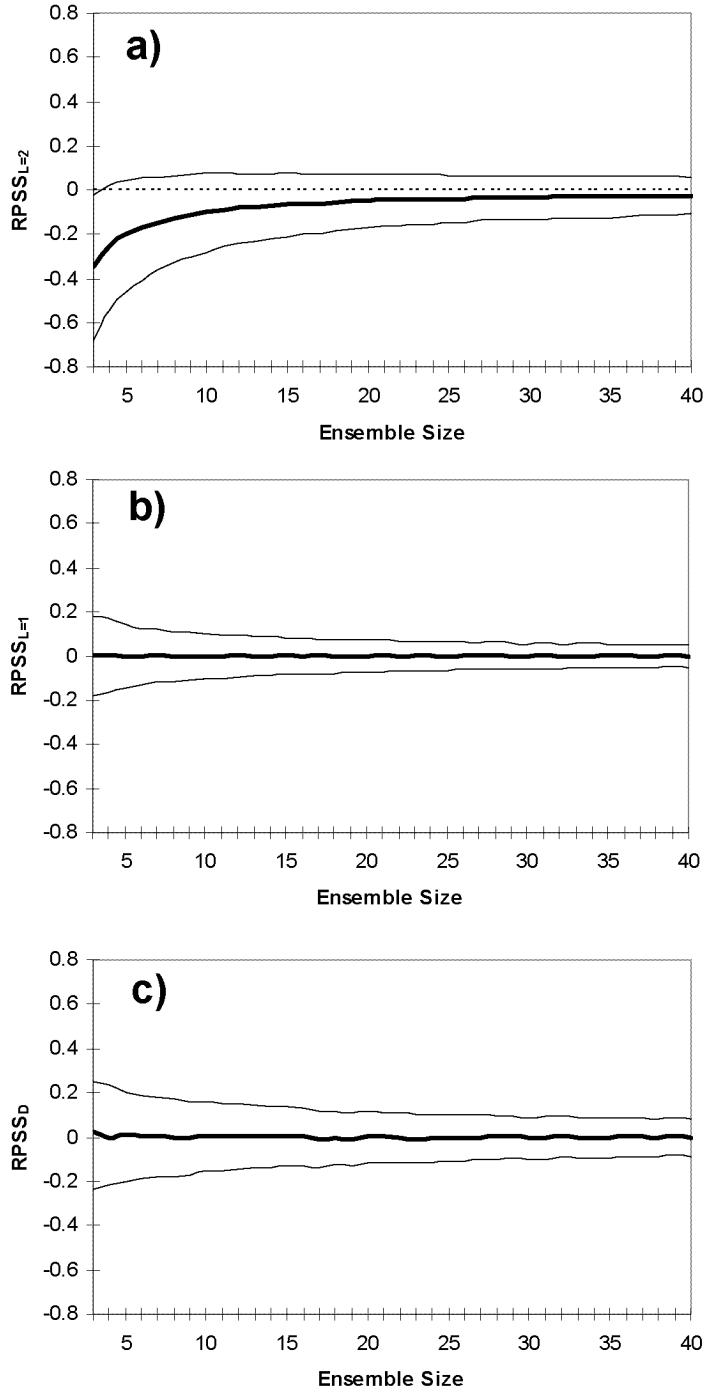


Figure 3.1: The (a) $RPSS_{L=2}$, (b) $RPSS_{L=1}$ and (c) $RPSS_D$ for white noise climate forecasts as a function of the ensemble size. The thin lines denote upper and lower 95 % confidence intervals for a 15 year sample. Thick lines show the mean values. Three equiprobable classes are used.

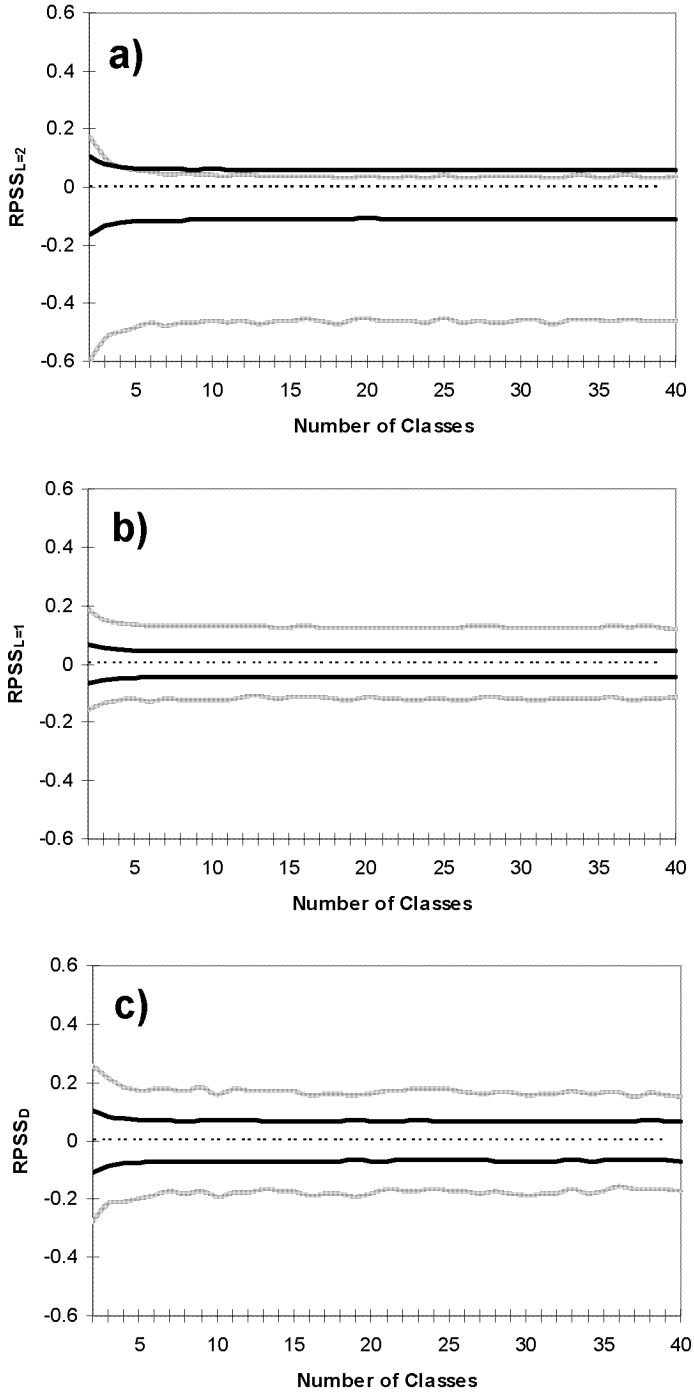


Figure 3.2: Same as Figure 3.1 but upper and lower confidence interval (95 %) as a function of equiprobable classes. Black (grey) lines illustrate sets of 40 (5) ensemble members.

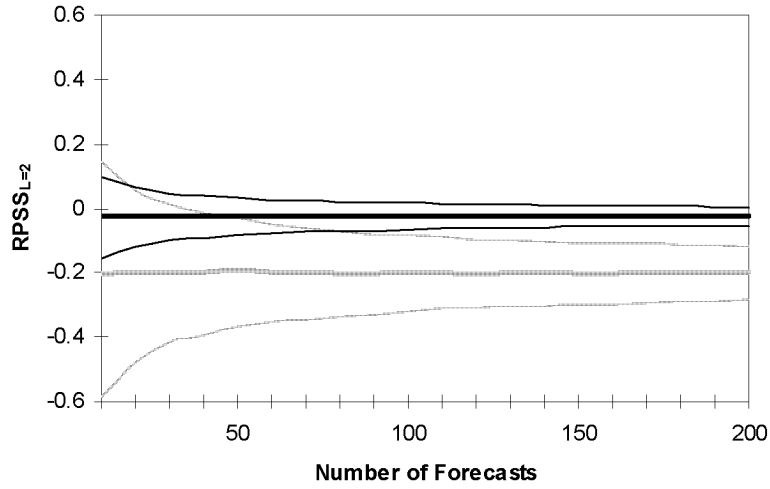


Figure 3.3: The $RPSS_{L=2}$ for 5 (grey) and 40 (black) ensemble members as a function of the number of forecasts. The thin lines denote upper and lower 95 % confidence intervals for a 15-year sample. Thick lines show the mean values.

To clarify the origin of these differences, a step by step calculation is performed for the $RPS_{L=2}$ and $RPS_{L=1}$ with white noise climate forecast. Suppose a forecast for two equiprobable classes (in this case the RPS_L reduces to the Brier Score, BS). Since the cumulative probabilities for the forecast and observation in the second class are always both equal to one, the $RPS_{L=2,k}$ value reduces to:

$$RPS_{L=2,k} = \left(Y_1 - O_1\right)^2 . \tag{3.3}$$

Similarly the $RPS_{L=1}$ reduces to:

$$RPS_{L=1,k} = \left|Y_1 - O_1\right| . \tag{3.4}$$

A climatological reference forecast predicts a probability of 1/2 for the event to be below the mean value. For such a forecast the difference to the observations in the cumulative probability space is always 1/2, independent of whether the observations were above or below the mean value. Hence the $RPS_{L=2,CL}$ is 1/4, whereas the $RPS_{L=1,CL}$ is 1/2.

In practice an ensemble system has only a finite ensemble size. Thus, the predicted cumulative probabilities y_1 can only take on a set of discrete values. These probabilities are (0, 1/m, 2/m, …, (m-1)/m, 1) for an ensemble system with m members. For a 3-member ensemble system these values are explicitly (0, 1/3, 2/3, 1). The mean $RPS_{L=2,FC}$ is then $(1/8)1 + (3/8)(2/3)^2 + (3/8)(1/3)^2 + (1/8)0 = (1/3)$, where the forecast probabilities are weighted by the relative frequency of the occurrence. Evidently the squaring of the cumulative measures gives an $RPS_{L=2,FC}$ that is larger than the expected reference value

$RPS_{L=2,\,CL}$ of 1/4. As a consequence, the $RPSS_{L=2}$ takes a negative value. In a general formulation of the bias of a white noise climate forecast, each possible cumulative probability value is weighted by the probability of occurrence that is given by the binominal distribution. Thus, for a two class system the scores of the unskilled ensemble forecast and the reference forecast are given by

$$RPS_{L,FC} = \frac{1}{2^m}\sum_{k=0}^{m}\binom{m}{k}\left(\frac{k}{m}\right)^L = \frac{1}{2^m}\sum_{k=1}^{m}\binom{m-1}{k-1}\left(\frac{k}{m}\right)^{L-1}, \tag{3.5}$$

whereas the score for the reference forecast is:

$$RPS_{L,CL} = \frac{1}{J^L}. \tag{3.6}$$

For L=2 the $RPS_{L=2,FC}$ is always larger than $RPS_{L=2,CL}$. Figure 3.4 shows the bias of the $RPSS_L$ for different ensemble sizes and norms. The analytical bias is almost equal to the bias in the white noise climate forecast (Fig. 3.1a). Finally note that for more than two classes, the relative frequency of occurrence does not linearly increase, and equation (3.5) and (3.6) increase in complexity. But the bias remains also for forecast systems which have ensemble sizes that are an integer multiple of the categories (not shown).

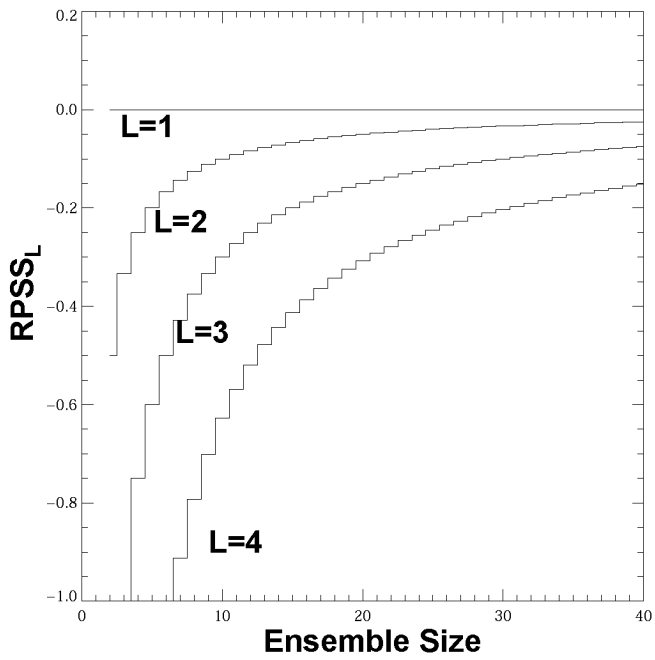


Figure 3.4: The $RPSS_L$, as a function of the ensemble size and different norms L for white noise forecasts of two classes.

For L=1 the $RPS_{L=1,FC}$ of a 3-member ensemble system gives (1/8)1 + (3/8)(2/3) + (3/8)(1/3) + (1/8)0 = 1/2. The outcome of $RPS_{L=1,FC}$ is exactly equal to the expected reference value $RPS_{L=1,CL}$ (1/2) and the $RPSS_{L=1}$ is equal to zero. The analytical bias (Fig. 3.4) is zero, as expected. For higher orders of the norm the bias is even stronger. The example illustrates that the negative bias in the standard definition of the $RPS_{L=2}$ is primarily a consequence of the squared measure used to quantify the forecast error in the cumulated probability space. This discretization-cumulative-squaring error also occurs for large ensemble sizes, but the negative bias is relatively small. However, for systems with small ensemble sizes, the bias can reach values comparable to the skill of the system and the score becomes meaningless.

Although the $RPSS_{L=1}$ is a skill score that can be used for systems with small ensemble size it is handicapped by the matter that it is not strictly proper (i.e. the forecasted probability values can be hedged towards values that are likely to give higher or equal scores). Strictly proper scores discourage forecasters from hedging their forecasted probabilities towards probabilities that are likely to score higher (Stephenson and Jolliffe 2003). To prove that the RPS_L is not strictly proper, we calculate the expected score a forecast y_i would receive and find what forecast f_i yields the best score. Let F_j be the cumulative probability for the event being forecasted defined as $F_j = \sum_{i=1}^{j} f_i$. The expected RPS_L becomes (e.g. see Wilks, 1995):

$$E(RPS_L) = \sum_{j=1}^{J} \left[F_j \left(Y_j - 1 \right)^L + \left(1 - F_j \right) Y_j^L \right]. \tag{3.7}$$

If the score is strictly proper, then the score will be minimised if $y_i = f_i$. To find the value of y_i that minimises $E(RPS_L)$, we take the partial derivative with respect to $\mathbf{y}$ (we assume f_i to be constant) which gives:

$$\frac{\partial E(RPS_L)}{\partial \mathbf{y}} = \sum_{j=1}^{J} \left[L \left(F_j \left(Y_j - 1 \right)^{L-1} \right) + L \left(1 - F_j \right) Y_j^{L-1} \right]. \tag{3.8}$$

For L=1 (3.8) reduces to:

$$\frac{\partial E(RPS_{L=1})}{\partial \mathbf{y}} = \sum_{p=1}^{J} 1 \neq 0. \tag{3.9}$$

The partial derivative is never zero and the $RPS_{L=1}$ cannot be minimised. Thus the $RPS_{L=1}$ and hence the $RPSS_{L=1}$ are not strictly proper. For L=2 the following result is reached:

$$\frac{\partial E(RPS_{L=2})}{\partial \mathbf{y}} = \sum_{j=1}^{J} \left(2\left(Y_j - F_j \right) \right), \tag{3.10}$$

which is zero for $y_i = f_i$ and hence the $RPS_{L=2}$ is strictly proper. For two classes, the $RPS_{L=2}$ is a special case of the Brier Score and has been shown to be strictly proper (for example see Mason 2004).

In order to maintain the strictly proper characteristic of the $RPSS_{L=2}$, the negative bias needs to be removed without changing the norm of the skill. A straight forward way is to introduce a discretization and squaring error in the reference forecast artificially. To do so, we calculate the score of the reference forecast in (3.2) with random re-sampling of the climatology, where the re-sample size is equal to the ensemble size. In this case the de-biased $RPSS_{L=2}$ becomes

$$RPSS_D = 1 - \frac{q \sum_{k=1}^{N} RPS_{L=2,FC,k}}{\sum_{k=1}^{N} \sum_{l=1}^{q} RPS_{L=2,CL,k,l}}, \tag{3.11}$$

where q is the number of discrete re-samples of the reference forecasts. To ensure that the climatology is fully represented, q must be chosen large enough. Since the reference forecast now consists of an ensemble of the same size as the forecast ensemble, the possible probabilities and the relative frequency of occurrence take on the same discrete values as for the forecast system. This newly defined skill score ($RPSS_D$) is zero for any ensemble size and number of classes (Fig 3.1c, 3.2c). Furthermore, since the $RPS_{L=2}$ is strictly proper, the set of sub-samples in the denominator of (3.11) are strictly proper, too. Therefore, the $RPSS_D$ is also proper and provides an adequate strategy to compare ensemble systems with low numbers of ensemble size.

3.4. Application to Seasonal Forecasts

In order to see the benefit of the de-biased $RPSS_D$ of a real application the skill scores of the ECWMF Seasonal Forecast System 2 are calculated. This system is an operational, fully coupled Atmosphere Ocean GCM and is described in detail by Anderson et al. (2003). The hindcast data analysed here consists of forecasts with 1-month lead time (months 2 to 4) of the winter mean (DJF) 2 metre temperature (T2). A set of 40 ensemble members is available for each year of the hindcast period 1987-2001. For further analysis the forecasts are post-processed by removing a lead time dependent mean model drift based on 15 years of hindcast climatology. The reference forecast is based on a Gaussian fit to the ECMWF ERA40 Re-Analysis for the same period 1987-2001. This data set is also used to define the edges of the three probability classes used.

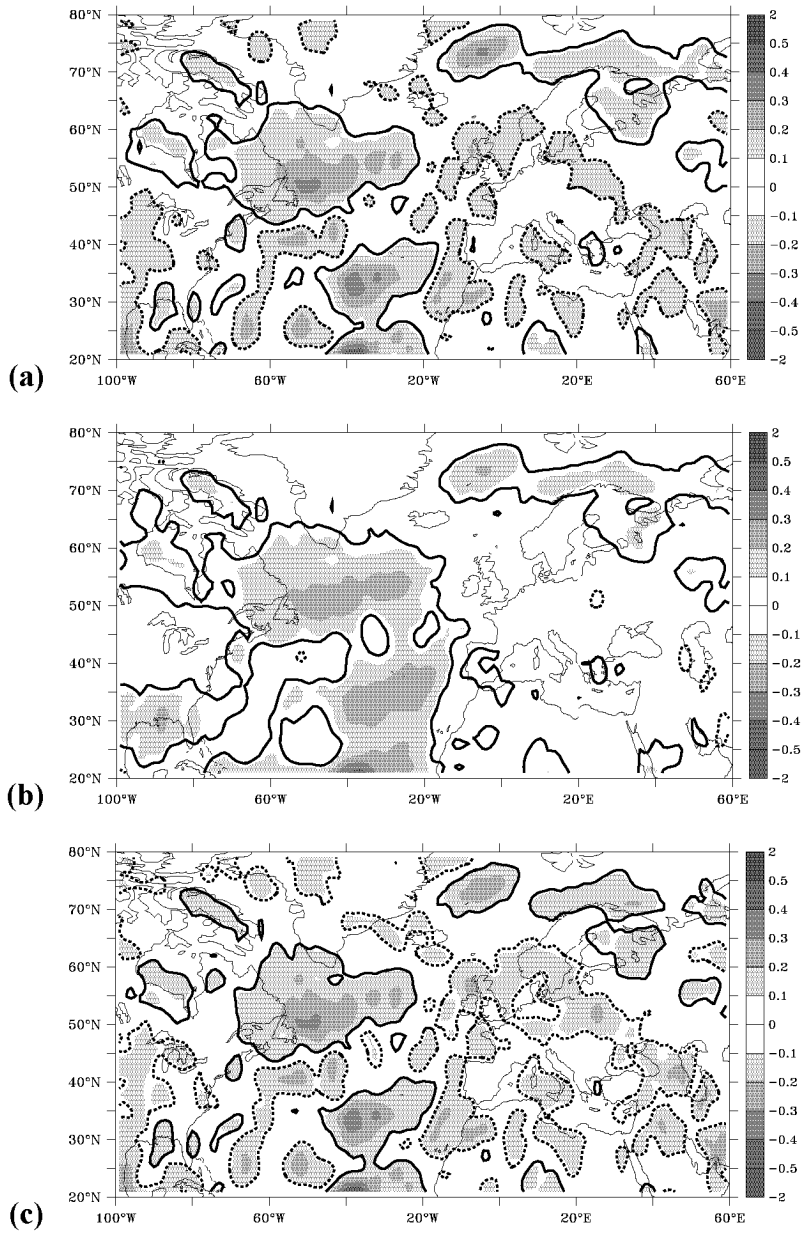


Figure 3.5: Grid point based (a) $RPSS_{L=2}$,(b) $RPSS_{L=1}$ and (c) $RPSS_D$ for winter mean (DJF) near surface temperature forecasts based on the ECMWF Seasonal Forecast System 2. The forecasts are based on November initialisation and cover the winters 1987/88 - 2001/2002. An ensemble system with 40 members is used. The ±95 % confidence intervals are denoted as thick plain (dotted) contours.

In Fig. 3.5a the grid point based $RPSS_{L=2}$ is shown for the full set of 40 ensemble members for T2. The overall picture of the $RPSS_{L=2}$ is dominated by alternate patterns of strong positive and negative skill scores. Negative skill scores are found in large areas over

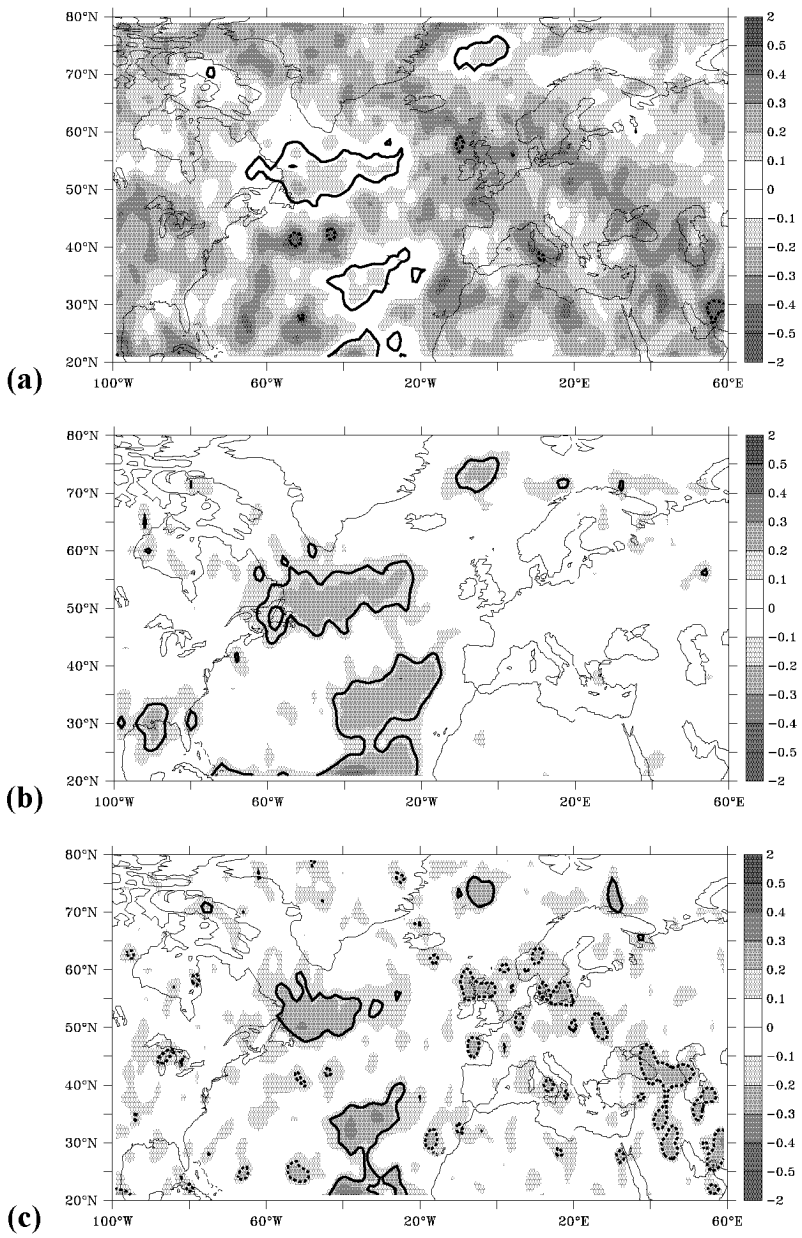


Figure 3.6: Same as Figure 3.5 but for an average of 30 simulations based on a random subset of 5 ensemble members.

Europe and positive skill scores over the Northern Atlantic Ocean, in the Northern part of Scandinavia and at the East coast of America. In Fig. 3.5b the skill scores are shown for the $RPSS_{L=1}$. The figure shows the same regions of positive skill score over the Atlantic and Northern Scandinavia as the $RPSS_{L=2}$. However, regions of localised negative skill score disappear over the continents. The $RPSS_D$ (Fig. 3.5c) shows similar results as the $RPSS_{L=2}$, which reflects that the $RPSS_D$ is equivalent for a large ensemble size.

The sensitivity to small numbers of ensemble members is shown in Fig. 3.6. Here 30 sets of forecasts are averaged, with each set consisting of five members, randomly re-sampled from the 40 members ensemble. Due to the smaller ensemble size the confidence intervals are wider than for 40 ensemble members. The resulting mean of the sub set of the $RPSS_{L=2}$ (Fig. 3.6a) shows distinct regions of positive skill scores. However, strong negative skill scores cover most of the entire region. Although these negative areas are not statistically significant, their magnitudes are comparable to the significant positive values. The $RPSS_{L=1}$ (Fig. 3.6b) does not show these regional negative areas in the $RPSS_{L=2}$. The overall picture exhibits mainly significant positive areas in the Atlantic. The values of the $RPSS_{L=1}$ based on 5 members are of the same order of magnitude as the values based on 40 members (Fig. 3.5b) but the confidence intervals are wider. The $RPSS_D$ for this set of ensemble members is illustrated in Fig. 3.6c. The $RPSS_D$ is generally higher than the $RPSS_{L=2}$ (Fig. 3.6a). The European and American continent, for which strong negative skill scores are found for the $RPSS_{L=2}$, are now mostly covered by skill scores in the range of ± 0.10. These are indications of the strong negative bias (see section 3.3). However, single localised areas with strong negative skill scores still remain.

3.5. Signal-to-Noise Detection Problem

The examined strategies to calculate the ranked probability skill score allow to readdress the question of how many ensemble members are required for a forecast system to detect a climate anomaly of a known signal-to-noise ratio. In the white noise climate system the signal-to-noise ratio is described by the mean shift in standardised units of a hypothetical distribution of climate anomalies to climatological distribution. The procedure corresponds to the one introduced by Kumar et al. (2001).

Fig. 3.7a) and c) show the skill scores ($RPSS_{L=2}$ and $RPSS_D$) as a function of the signal-to-noise ratio. Skill scores for an ensemble system with 2, 5, 40 and 100 ensemble members are shown. For large climate mean shifts, the skill scores are independent of ensemble sizes. A clear difference is found for forecasts with small ensemble sizes and small climate shifts. As expected from the discussion above, the $RPSS_{L=2}$ has a negative bias for small ensemble sizes. The $RPSS_D$ shows no negative skill for weak anomalies and small ensemble sizes. The $RPSS_D$ is positive and irrespective of the ensemble size. This is also found for the $RPSS_{L=1}$ (Fig.3.7b).

An ensemble forecast system with a given ensemble size and a given number of forecasts has still a signal-to-noise detection limit. In our two strategies the signal-to-noise detection limit can be described by the confidence intervals discussed above. In Fig. 3.7b), c) the vertical dashed lines indicate the 95 % confidence interval for the $RPSS_{L=1}$ and the $RPSS_D$ for ensemble sizes 5 and 40, and a number of 15 forecasts. For this hypothetical set up, comparable to the ECMWF hindcast system, a statistically significant skill score can only be estimated for a signal-to-noise ratio larger than ~0.3 (40 members)

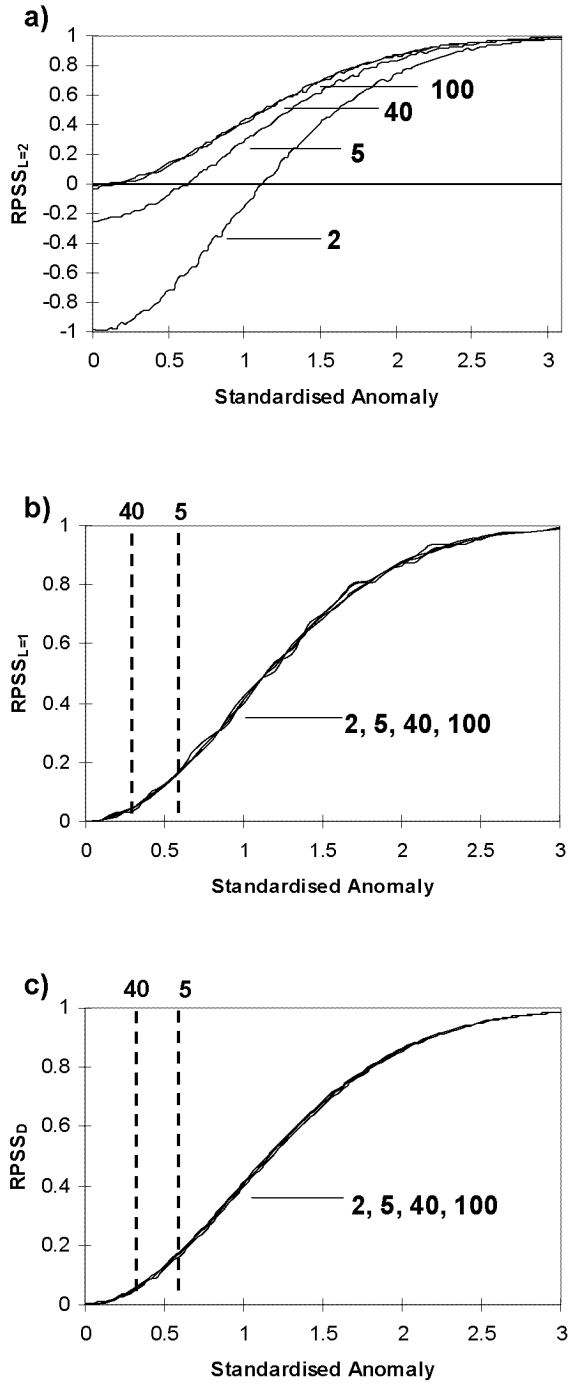


Figure 3.7: The (a) $RPSS_{L=2}$, (b) $RPSS_{L=1}$ and (c) $RPSS_D$ as a function of a standardised anomaly (signal-to-noise ratios, see text for details). Numbers denote ensemble size 2, 5, 40 and 100. The scores are based on three equiprobable classes. The vertical dashed lines in panel (b, c) indicate the 95 % confidence intervals for 40 and 5 members, respectively.

~0.6 (5members), respectively. This value can be interpreted as the minimum ensemble size required to achieve a positive skill score for a given signal-to-noise ratio (Kumar et al. 2001).

To further investigate the origin of the negative bias, a decomposition of the $RPSS_{L=2}$ and $RPSS_D$ is carried out. The RPS can be decomposed into a reliability term and a resolution term (for details see Hersbach 2000; Wilks 1995):

$$RPS = \frac{1}{N}\sum_{k=1}^{J} N_k (y_k - \overline{o}_k)^2 - \frac{1}{N}\sum_{k=1}^{J} N_k (\overline{o}_k - \overline{o})^2 + \overline{o}(1-\overline{o}) \tag{3.12}$$

where $\overline{o}$ is the relative frequency of the observations, $\overline{o}_k$ is the sub-sample relative frequency and N_k the number of times each forecast is used in the collection of the forecast being verified. The reliability score (first term) is a function of the squared difference between the forecast probability and the observed frequency in the different probability categories, while the resolution score (second term) is the average square difference between the observed frequency in each probability category and the mean frequency observed in the whole sample.

Since for the $RPS_{L=2,CL}$ the resolution and the reliability terms are zero, the skill score is only described by the uncertainty term. From (3.12) the skill score can be written as:

$$RPSS(A) = 1 - \frac{UNC + REL - RES}{UNC + A}. \tag{3.13}$$

Here A describes the difference between the $RPS_{L=2,CL}$ evaluated with the probabilities of the re-samples and those defined by the probabilities of the full reference distribution. Splitting up (3.13) a reliability skill score (RELSS) can be written as

$$RELSS = 1 - \frac{REL}{UNC + A}, \tag{3.14}$$

and a resolution skill score (RESSS) and uncertainty skill score (UNCSS) as:

$$RESSS = \frac{RES}{UNC + A} \tag{3.15}$$

$$UNCSS = -\frac{UNC}{UNC + A} \tag{3.16}$$

In Fig. 3.8a) and d) the $RELSS_{L=2}$ and $RELSS_D$, respectively are plotted as a function of the signal-to-noise ratio. The $RELSS_{L=2}$ shows a strong dependence on the ensemble size.

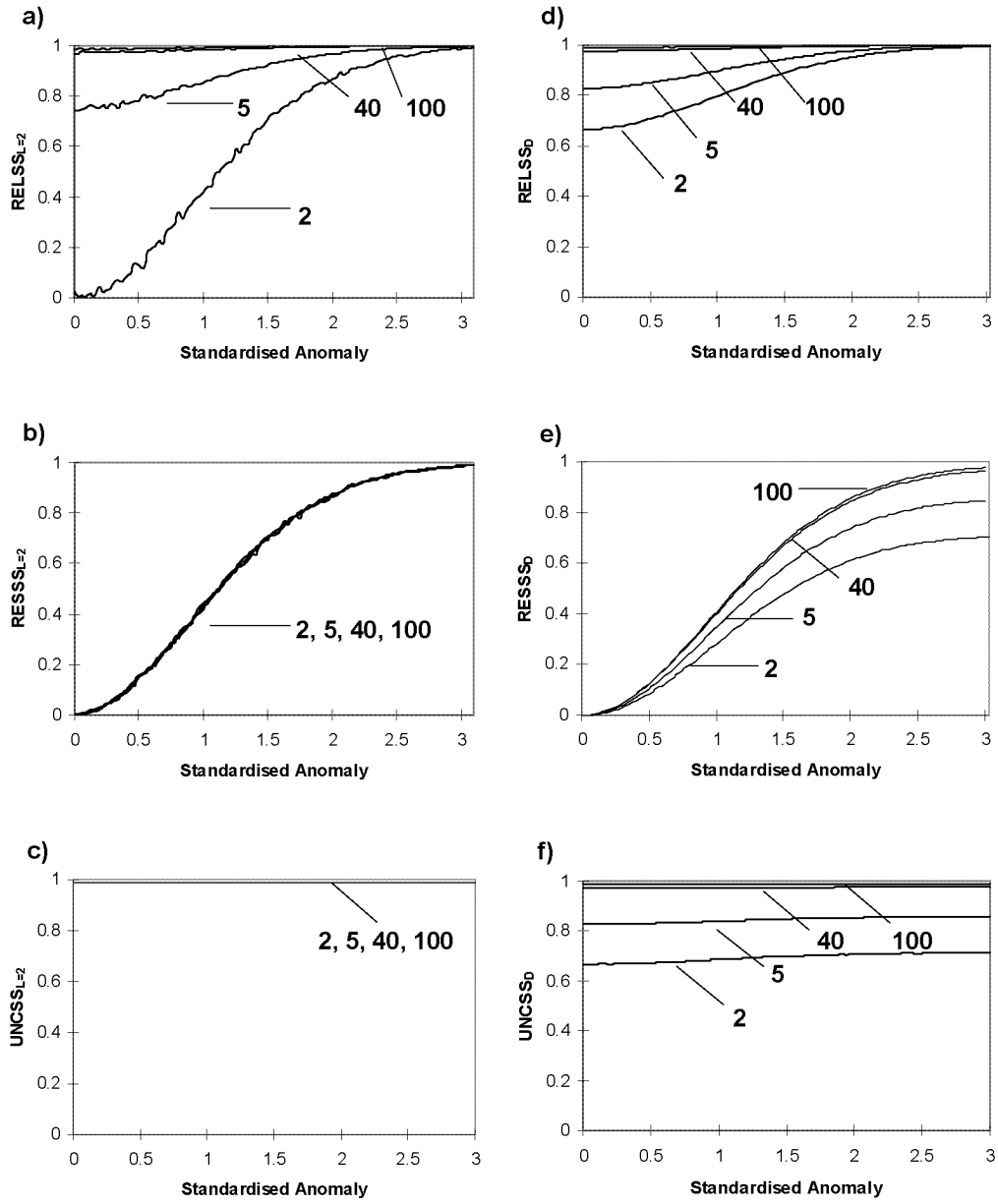


Figure 3.8: The (a) $RELSS_{L=2}$, (b) $RESSS_{L=2}$, (c) $UNCSS_{L=2}$, (d) $RELSS_D$, (e) $RESSS_D$ and (f) $UNCSS_D$ as a function of a standardised anomaly. Numbers denote ensemble size 2, 5, 40 and 100.

For a weak signal-to-noise ratio, the scores range from zero reliability to perfect reliability for a 2-member and 100-member ensemble system, respectively. The $RELSS_D$ is still dependent on ensemble size for weak anomalies, but reveals substantially higher values.

In Fig 3.8b) and e) the results for the $RESSS_{L=2}$ and $RESSS_D$ are shown. Whereas the $RESSS_{L=2}$ proves to be independent of the ensemble sizes and the signal-to-noise ratios, the $RESSS_D$ is reduced for small ensemble sizes and strong anomalies. In Fig 3.8c) and f) the $UNCSS_{L=2}$ and $UNCSS_D$ are shown for varying ensemble sizes. Obviously for A=0, which is the case for the $RPSS_{L=2}$, the $UNCSS_{L=2}$ is one for any ensemble size. The $UNCSS_D$, however, shows dependence on the ensemble size. For a 5-member ensemble system the $UNCSS_D$ is reduced to a value of about 0.85. For the $RPSS_D$ the bias in the RELSS is compensated by introducing ensemble size dependence in the resolution and uncertainty term.

3.6. Conclusions

In this study the mechanics of the $RPSS_L$ are studied in the context of forecast systems with small ensemble sizes and different norms L. In agreement with earlier studies, it is shown that the standard calculation of the $RPSS_{L=2}$ leads to a negative bias that can be even larger than the expected skill of the forecast system itself. This negative bias is a consequence of the squared measure used to quantify the forecast error in the cumulative probability space. It is particularly large for small ensemble sizes. For higher orders of L the bias is further increased.

Two strategies are introduced addressing the bias problem of the $RPSS_{L=2}$. The bias can be eliminated by either one of the two alternative scores. The first one is a proposed modified version of the $RPS_{L=2}$, which is based on the absolute difference ($RPS_{L=1}$) instead of the squared difference of the cumulative probabilities. This score is comparable to the mean absolute error (MAE). The expected skill score is independent of the ensemble size whereas the confidence intervals are related to the ensemble size and the number of forecasts. However, the $RPSS_{L=1}$ is not strictly proper, which means that the probabilities can be changed without impact on the score value. The second score considered is the $RPSS_D$, which represents a de-biased version of the standard $RPSS_{L=2}$. The proposed modification involves a re-sampling of the climatology as a reference forecast ($RPSS_D$). It is shown that this method renders reasonable results even for systems with small ensemble sizes and is strictly proper. A pure random climate forecast is used to show that the $RPSS_{L=1}$ and $RPSS_D$ provide an unbiased estimate of the skill score even for small ensemble sizes. For large ensemble sizes both skill measures are comparable. Random noise forecasts form also the base to determine confidence intervals.

To test the newly proposed scores, two examples are considered. First the operational ECMWF Seasonal Forecast System 2 is used to quantify the skill of near surface winter mean temperature forecasts with the $RPSS_{L=1}$, $RPSS_{L=2}$ and the $RPSS_D$. It is shown that the new skill scores yield increased values for the forecast system, in particular for small ensemble sizes. Furthermore, the grid point based skill score structure is much more

homogeneous and the occurrence of scattered negative values, as is the case with $RPSS_{L=2}$, is largely suppressed.

Second, the $RPSS_D$ and the $RPSS_{L=1}$ are used to find the minimum ensemble size required to predict a given climate signal. Here a white noise climate system is used in which the signal-to-noise ratio is described by the mean shift of a hypothetical distribution of climate anomalies to the climatological distribution in standardised units. By using a hypothetical set up comparable to the ECMWF hindcast system, statistically significant skill scores can be anticipated for climate signal-to-noise ratios larger than ~0.3 (40 members) and ~0.6 (5 members), respectively. These are similar results as found by Kumar et al. (2001), but the statistical significance is now associated with confidence intervals instead of the ensemble size.

In the context of the signal-to-noise detection problem, a decomposition of the quadratic norms identifies the bias of the $RPSS_{L=2}$ as a reliability problem, whereas the resolution is unaffected by the ensemble size. A decomposition of the re-sampling strategy of the $RPSS_D$ shows an improvement in the reliability for small ensemble sizes. However, this is at the expense of the resolution skill score which is reduced. This seems more logical, as in this framework the resolution should also be affected by the ensemble size.

It is argued that for ensemble systems with small ensemble sizes, a de-biased version of the $RPSS_{L=2}$ should be used to quantify the probabilistic skill of the system, either the $RPSS_{L=1}$ or the $RPSS_D$. Since the $RPSS_{L=1}$ proves not to be strictly proper we suggest a preferable use of the $RPSS_D$. But also other formulations could be considered which address the forecasts to be a random guessing, as recently shown in a parallel study of Mason (2004).

Finally, the use of these new skill scores is not restricted to Seasonal Forecast Systems, but is likely to be beneficial for other applications such as climate prediction or short-range limited area ensemble prediction systems (Marsagli et al. 2001). De-biased skill scores are also desirable for comparison of multi-model ensemble systems with different ensemble sizes (Müller et al. 2004).

Chapter 4:

Probabilistic Seasonal Forecast of the European Climate

The aim of the present chapter is to provide an overview of the forecast skill on seasonal time-scales for the European-Atlantic region. First, forecasts of large-scale domain 2m temperatures are compared with emphasis on the tropics and the mid-latitude regions. Since the tropics provide the source of a large fraction of seasonal to interannual climate variability, a more detailed description of the forecast skill in the tropics is given. Second, the seasonal forecasts are examined on grid-point scale. The ECMWF Seasonal Forecast System 2 and the DEMETER system are used to investigate the forecast skill. Within both systems a perfect model approach (PMA) is employed to quantify the potential predictability of the seasonal climate.

4.1. Data and Methods

In this chapter the forecast skills are calculated with the $RPSS_D$, which is described in detail in chapter 3. The distribution of the reference forecast is defined by a Gaussian fit to the observations. Three equiprobable classes are used. They are obtained from the distribution of the reference forecasts. Yet, it must be noted that the forecasts and reference forecasts are not cross-validated, which means that the actual forecasts and observations are included when the underlying scores are evaluated. If only a small number of forecasts and observations are used, the underlying scores might be strongly influenced by the actual values. This in turn may produce scores which are more precise than those for cross-validated forecasts and reference forecasts.

Three approaches are examined to define the forecast skill. First, the reference forecast is based on all ERA40 observations. This attempt is denoted as the forecast approach (FA). Second, the reference forecast is based on the ERA40 observations, but the forecasts are defined in terms of a simple persistence model. The persistence model is defined as:

$$PDF\left(f_{FC}(t_0+I)\right)=\alpha f_{OBS}(t_0)+PDF\left(f_{OBS}(t+I)-f_{OBS}(t)\right) \tag{3.1}$$

where $f_{OBS}(t_0)$ is the observation at the start time t_0 of the forecast, I the lead time, α the autocorrelation coefficient and t denotes the index of all observations $f_{OBS}(t)$. This gives a forecast distribution $PDF\left(f_{FC}(t_0+I)\right)$ for every start time and lead time that depends on

the differences between the observations $f_{OBS}(t+I)$ and $f_{OBS}(t)$ for all t. In this general form α can be adapted for any forecast time. For α=0 and large lead times the distribution becomes the climatology. For α=1 (3.1) becomes a pure persistence forecast. Here α=1 is used. This attempt is denoted as the probabilistic persistence approach (PPA).
Third, the reference forecast is based on all single predicted ensemble members of the model climate, and each single ensemble member is once treated as an observation. This attempt is denoted as the perfect model approach (PMA). This approach removes the bias, since the climatology is fully described by the model. The sample size is the length of the hindcast climatology multiplied by the ensemble size.

The multi-model to be used in section 4.4 is constructed of all individual models of the DEMETER system. The individual models are considered with regard to their own climate. The multi-model distribution for each grid point is then constructed by merging all model anomalies with equal weighting. Within the DEMETER system this gives 63 ensemble members at each grid point. Other methods, such as a Bayesian approach or a linear combination, for which the model anomalies have different weightings exist but are not applied in this thesis (Mason and Mimmack 2002, Rajagopalan et al. 2002, Coelho et al. 2003, Metzger et al. 2004).

In this chapter the forecast systems have different ensemble sizes. Furthermore, different starting months and temporal averages are considered. Table 4-1 gives a detailed overview of the forecast systems, starting months, ensemble and sample sizes and the 95 % confidence level of the $RPSS_D$ which are used in the corresponding sections.

	Forecast System	Start Months	Ensemble Size	Sample Size	95 % Confidence Intervals
Section 4.2	ECMWF System 2	Jan – Dec	5	FA: 180	±0.07
				FA & PPA: 15	±0.19
Section 4.3	ECMWF System 2	May & Nov	40	FA: 15	±0.08
				PMA: 600	±0.02
Section 4.4	DEMETER System (7 models)	Nov	Single Model 9	FA: 15 PMA: 135	±0.15 ±0.03
			Multi Model 63	FA: 15 PMA: 945	±0.05 ±0.01

Table 4-1: Overview of the Seasonal Forecast Systems, start months, ensemble size, sample size and 95 % confidence levels of the $RPSS_D$ which are used in the different sections.

4.2. Large-Scale Domain Averages

Time Series

First the time-series for the large-scale domain averages are examined. In Figure 4-1 the monthly mean 2m temperatures are shown for the global (a) and tropical average (b). The global mean shows a close correspondence between the ensemble mean and the observations. This holds especially for lead time 1 month (red). With increasing lead time the forecasts are further displaced from the observations. However, the variability is captured at all lead times. The tropical mean shows that much of the variability has a tropical origin. It can be seen that the time-series are dominated by the interannual variability of ENSO, with two extra-ordinary boreal winters in 1987/88 and in 1997/98. The ECMWF forecast system 2 is capable of predicting the strong warm events and the La Niña year 1988. However, with increasing lead times the forecasted amplitudes are smaller than the observations. Also the onsets are forecasted with some delay. For lead time 5 months, for example, the maximum amplitude of the 1987/88 event is predicted at least two months too late and by about one third too weak. The coherence between forecasts during weak ENSO episodes is marginal.

In Figure 4-2 the time series for the average northern hemisphere (a) and European (b) 2m temperature are shown. There is less coherence between the ensemble mean and the observations in the northern hemisphere than in the tropics. Although forecasts with lead time 1 month have amplitudes comparable to the observations, longer lead times are characterised by a strongly damped signal and a clear tendency towards anomalies with

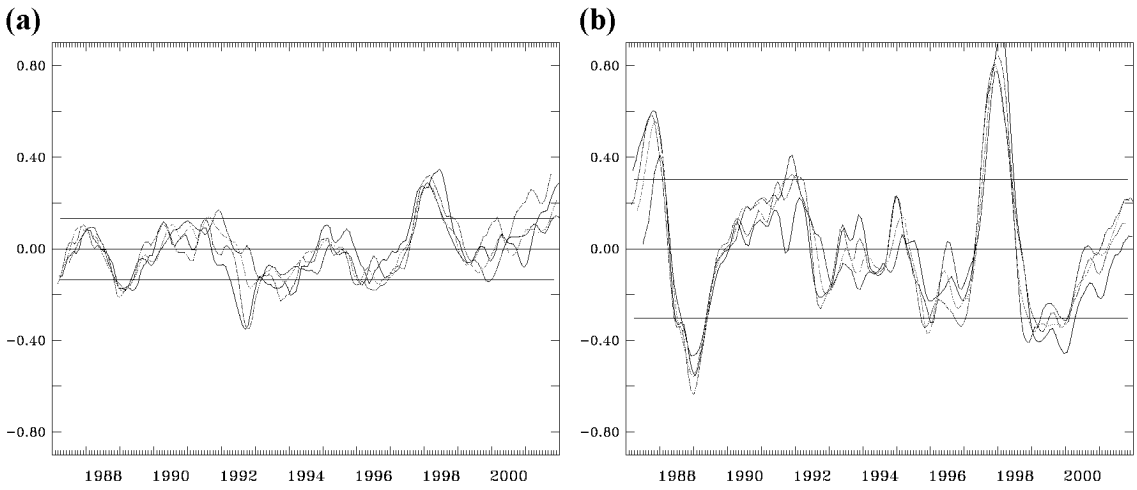


Figure 4-1: Global (a) and tropical average (b) of monthly 2m temperatures of the observations (black) and predicted ensemble mean (colours). The tropical average is defined as the area between 180°W to 180°E and 5°S to 5°. Shown are the ensemble mean for lead time 1 (red), 3 (green) and 5 (blue) and the ±1 standard deviation of the observations (horizontal lines).

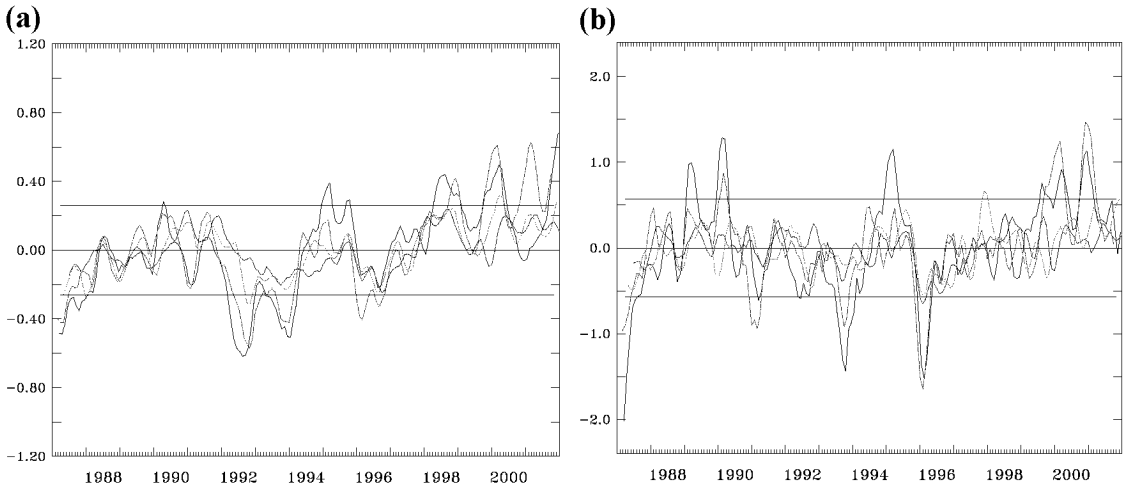


Figure 4-2: Same as Figure 4-1 but for the northern hemispheric (a) and European average (b). The northern hemisphere (European) region covers the area extending from 180°E-180°W and 20°N-90°N (10°W-60°E and 35°N-75°N).

zero amplitude. The same is found for the European mean for lead times greater than 1 month. However, it must be pointed out that for lead time 1 month most of the intense events (>1sd) are captured by the forecasts.

Correlation Coefficients and $RPSS_D$

The correlation between the ensemble mean and the observations of the four regions are shown in Figure 4-3. The correlation for the tropics shows a maximum of about 0.95 for lead time 0 months and decreases systematically to a value of about 0.71 for lead time 5 months. For this sample size a correlation of about 0.15 is significant at the 95 % confidence level (assuming a Gaussian distribution). The correlation coefficients of the global mean are generally smaller than in the tropics. A correlation of about 0.77 is found for lead time 0 months which decreases to about 0.52 for lead time 5 months.

As expected the correlation coefficients of the northern hemisphere and European means are much lower than those of the tropics. For the northern hemispheric mean we find a decrease of the correlation coefficients from 0.76 for lead time 0 months, to 0.23 for lead time 5 months. As for the European domain, the results are further reduced ranging from 0.66 for lead time 0 months, to 0.22 for lead time 5 months.

The forecasts are next examined with the $RPSS_D$. The reference forecast consists of all observations for the whole period which gives a sample size of 180 (see also table 4.1). The $RPSS_D$ of the spatial averaged monthly 2m temperatures are displayed in Figure 4-4. The highest skills are found in the tropical region with about 70 % for lead time 0 months and 55 % for lead time 1 month. Although a strong reduction in skill is found with

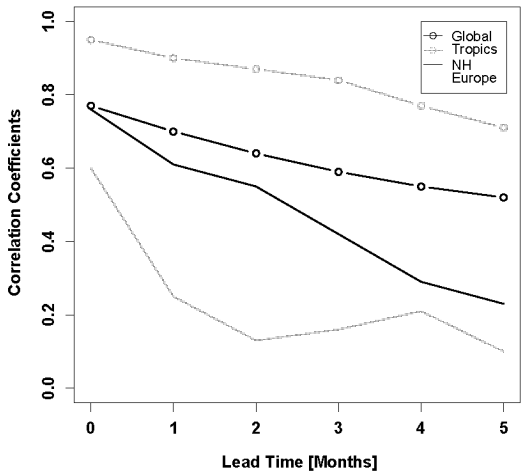


Figure 4-3: The correlation between predicted ensemble mean and observations of monthly 2m temperatures for the period 1987-2001. The different lines denote tropical average (grey circle), global average (black circle), NH average (black solid) and European average (grey solid).

increasing lead time, there still is a positive skill score up to lead time 5 months. The global mean 2m temperature skill scores show a smaller reduction ranging from 37 % for lead time 0 months, to 20 % for lead time 5 months. The probabilistic skill score for the northern hemisphere is also found to be positive for all the lead times with values of about 42 % and 29 % for lead time 0 months and 1 month, respectively. The results for

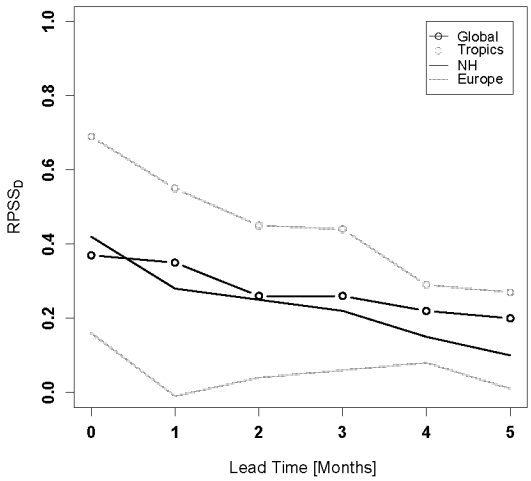


Figure 4-4: Same as Figure 4-3 but for the $RPSS_D$.

lead time 5 months show only a small benefit of about 10 %. If the European domain is considered, the skill scores are further reduced ranging from 16 % for lead time 0 months to 1 % for lead time 5 months. These results are in compliance with other studies (Palmer and Anderson 1994, Pavan and Doblas-Reyes 2000). In a group of multi-model seasonal hindcast experiments Pavan and Doblas-Reyes (2000), demonstrate similar results for meteorological fields such as the 500hPa geopotential and the total precipitation. Interestingly, the skill scores of the European domain show a minimum for lead time 1 month and have a small increase up to lead time 4 months. One possible explanation for this increase could be a delay of the predictive signal from the ocean to the continent. This would be consistent with the skills of other domains where ocean grid points are included. On the other hand the minimum at lead time 1 month could be due to a small spread of the forecast distribution. In such a case there are less ensemble members which are closer to the observation, as is the case for lead times greater than 1 month, where the forecast spread is wider.

Forecast Approach (FA) and Probabilistic Persistence Approach (PPA)

A frequently asked question concerns the benefit of expensive complex models when cheaper statistical methods can be considered. Recently, Landsea and Knaff (2000) examined the prediction performance of El Niño 1997/98 for twelve statistical and dynamical models, ranging from simple analogue methods to more complex coupled atmosphere-ocean model forecasts. They compare these models to an ENSO persistence method (ENSO-CLIPER) which is utilised as a baseline for the determination of the skill score. Based on this single but strong ENSO event, they find no skill for any of the models 1 to 3 seasons ahead, which suggests that the best estimate is performed by the persistence method.

More generally van Oldenborgh et al. (2003) compare the ECMWF Seasonal Forecast System 1 and 2 against a statistical and a persistence models of the form of Landsea and Knaff (2000) for the period 1987-2001. For the monthly Nino3 index they show that the ECMWF systems provide anomaly correlation coefficients (ACC) of about $r>0.7$ throughout the year. In winter the ACCs of the two systems amount to about $r\approx0.9$. The statistical methods perform similar to the dynamical models in the winter months. However, in the summer months the skill scores are strongly reduced. Among all models, the smallest skill score is achieved by the persistence model in the summer months, where the ACCs amount to about $r\approx0.1$ and $r\approx0.2$ for June and July, respectively.

A comparison of the skill scores resulting from the probabilistic persistence approach (PPA) with the skill scores of the forecast approach (FA) provides an alternative method to distinguish whether or not a forecast based on persistence alone is sufficient. Figure 4-5 summarises the results for the Nino3.4 region for lead time 1, 3 and 5 months.

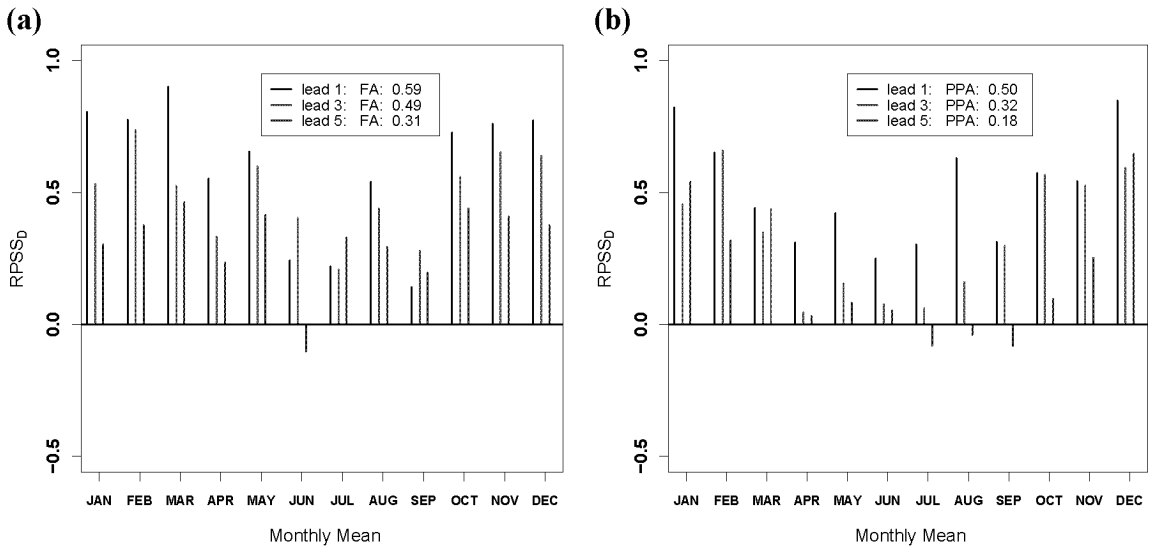


Figure 4-5: The $RPSS_D$ of the tropical 2m temperature average for the ECMWF system 2 for different initial times (calendar months). The colours indicate different lead times (see inlet). Shown are the skills for (a) the FA and (b) the PPA. The inlet shows the corresponding average over all calendar months. The reference forecast is based on the ERA40 re-analysis for the period 1987-2001. Skill scores above 0.19 are statistical significant at the 95 % confidence level.

Obviously, there is a clear seasonal cycle in forecast skill with higher skill scores in winter and lower skill scores in summer. Statistically significant positive skill scores are found for all calendar months for the PPA for lead time 1 month and for the FA for lead time 1 and 3 months. For lead time 3 months the PPA has non-significant skill scores in June and July. For lead time 5 the PPA has a longer period with non-significant skill scores ranging from April to October. For this lead time the FA has only a non-significant skill score in June. The averages of the skill scores through all calendar months are statistically significant for all lead times. The FA performs 10 % to 17 % higher than the PPA.

Finally, Figure 4-6 shows the FA and PPA for the European domain. The skill scores have a high variation from month to month and are mostly not significant. Only in summer there is a tendency towards the positive skill scores for the FA and PPA. However, the interpretation of the strong positive skill scores requires careful considerations, because the forecasts and reference forecasts are not cross-validated. The averages over all calendar months are not significant for the FA and PPA, respectively.

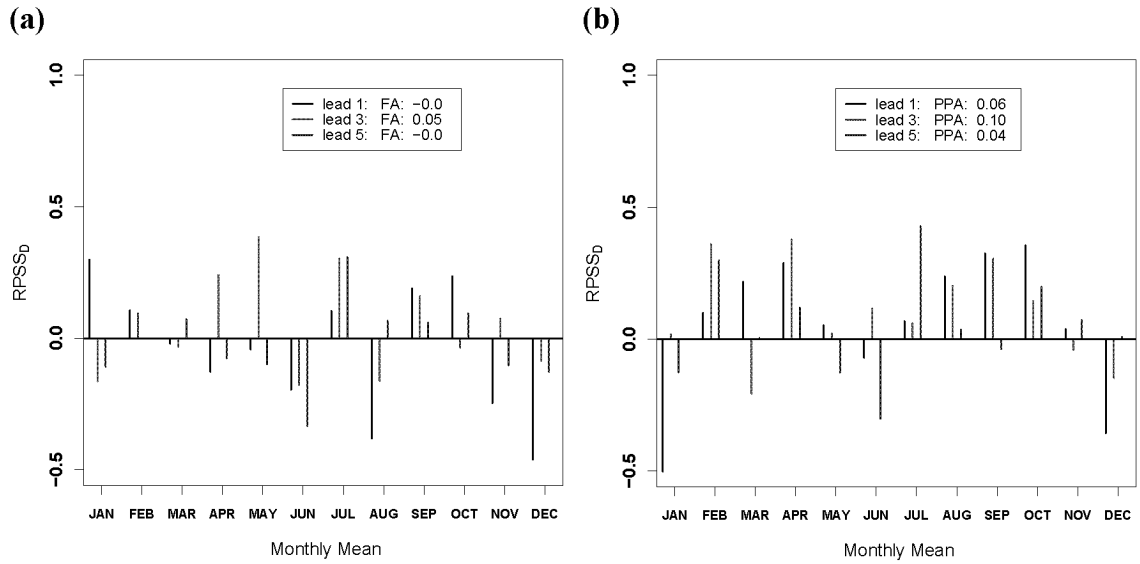


Figure 4-6: Same as Figure 4-5 but for the European average.

In summary the ECMWF Seasonal Forecast System 2 has a high forecast skill in the tropics, even for lead time 5 months. The skill, however, is seasonally dependent and lowest in summertime. A prediction based on observations alone using a probabilistic persistence method (PPA) provides an alternative for producing forecasts in the Nino3.4 region. For the tropics, the forecast benefit is slightly smaller than with system 2. For the extra-tropical mean the skill is generally lower, but highest in summer.

4.3. ECMWF Seasonal Forecast System 2 Grid-Point Skill

The large-scale mean quantities described in the last section provide an identification of the potential predictability in pre-defined areas. Since atmospheric processes are organised in locally and temporally coherent structures, a too large definition of the region of interest can lead to a cancellation of the predicted signal. Furthermore, user specific needs are to a large extent driven by the requirement of local station-based information. In this section the seasonal prediction skill in the Euro-Atlantic sector is investigated on model grid-point scale.

Forecast Approach (FA)

All results presented in this and section 4.3 are based on the same probabilistic measure as in the last subsection, but applied on 3 monthly mean anomalies (for details see also

Table 4-1). In Figure 4-7 the $RPSS_D$ is shown for the 2m temperature of the ECMWF system 2 for winter (a) and summer (b). The skill scores for the DJF mean shows distinct areas of enhanced skill score with emphasis over the ocean. Several local maxima are found off the coast of Newfoundland and the sub-tropical Atlantic. Some significant positive skill scores are located in the northern part of Canada and the European Polar Sea. For the European continent, however, no significantly positive skill is found. Instead the figure is covered by localised regions of strong negative skill. However, they are scattered all over and their occurrence may be artificial due to the limited hindcast period of 15 years (see below).

In JJA (Figure 4-7b) the maximum skill score in the central North Atlantic is aligned towards the Labrador Sea and is increased in magnitude, whereas the sub-tropical positive skill at the African coast is reduced. Central and southern Europe (and the Mediterranean Sea) is now covered by a positive skill of about 25 %. Strong negative skill scores are found over northern Europe as well as scattered over the subtropical Atlantic and the Middle East. The average skill score over the entire domain is about 5 % higher than in the winter season.

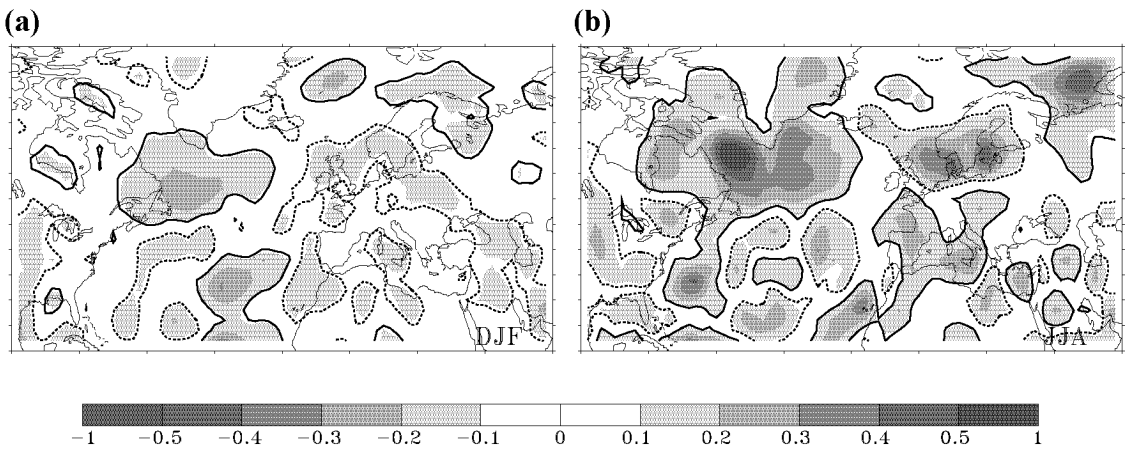


Figure 4-7: The $RPSS_D$ of the 2m temperatures (shadings, 5 % intervals) for the FA. Shown are the seasons DJF (a) and JJA (b) for lead time 1 month. The contour lines show the upper (solid) and lower (dotted) 95 % confidence levels. The $RPSS_D$ at each grid point is smoothed with the adjoining neighbouring skill scores.

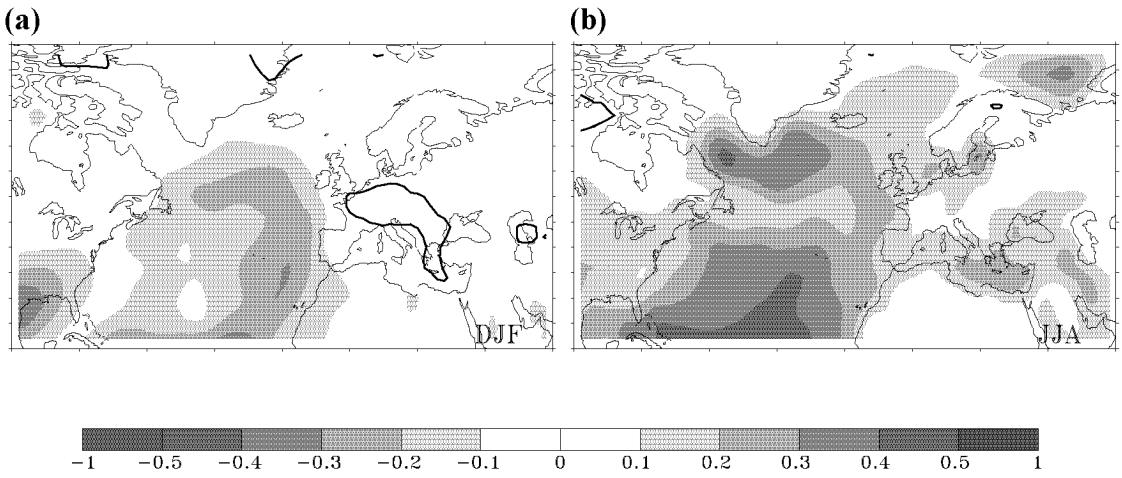


Figure 4-8: Same as Figure 4-7 but for the PMA.

Perfect Model Approach (PMA)

In Figure 4-8 the skill scores for the 2m temperature for the Euro-Atlantic sector are shown for the PMA. Figure 4-8a shows the skill scores for DJF. The picture is covered by positive skill scores over the entire domain. The distinct maxima which are achieved in the FA, are now connected to a large-scale, horseshoe-like pattern with magnitudes up to 35 %. The extra-tropical land masses are mostly covered by small but significantly positive skill scores of 5 %-15 %. Only the central European continent does not show significantly positive skill scores. Finally, the southern part of the US is covered by a maximum of about 40 % with strong linkage to the tropical Pacific (not shown). Please note that the 95 % significance levels are now closer to zero as in the FA because the sample size is increased (see Table 4-1).
In JJA (Figure 4-8b) the horseshoe-like pattern of the skill score is split into a local maximum further aligned to the Labrador Sea (max. 45 %) and a local maximum which is shifted to the sub-tropical Atlantic. The Mediterranean Sea is also covered by relative high skill scores with values up to 25 %.

4.4. DEMETER Multi-Model Grid-Point Skill

Forecast Approach (FA)

The $RPSS_D$ for the winter mean 2m temperatures are shown for the individual models of the DEMETER system in Figure 4-9. Despite the large local variability of the skill scores in the individual models, there are recurrent patterns of positive and negative signs. Regions with recurrent positive skill scores are found off the west coast of Africa and

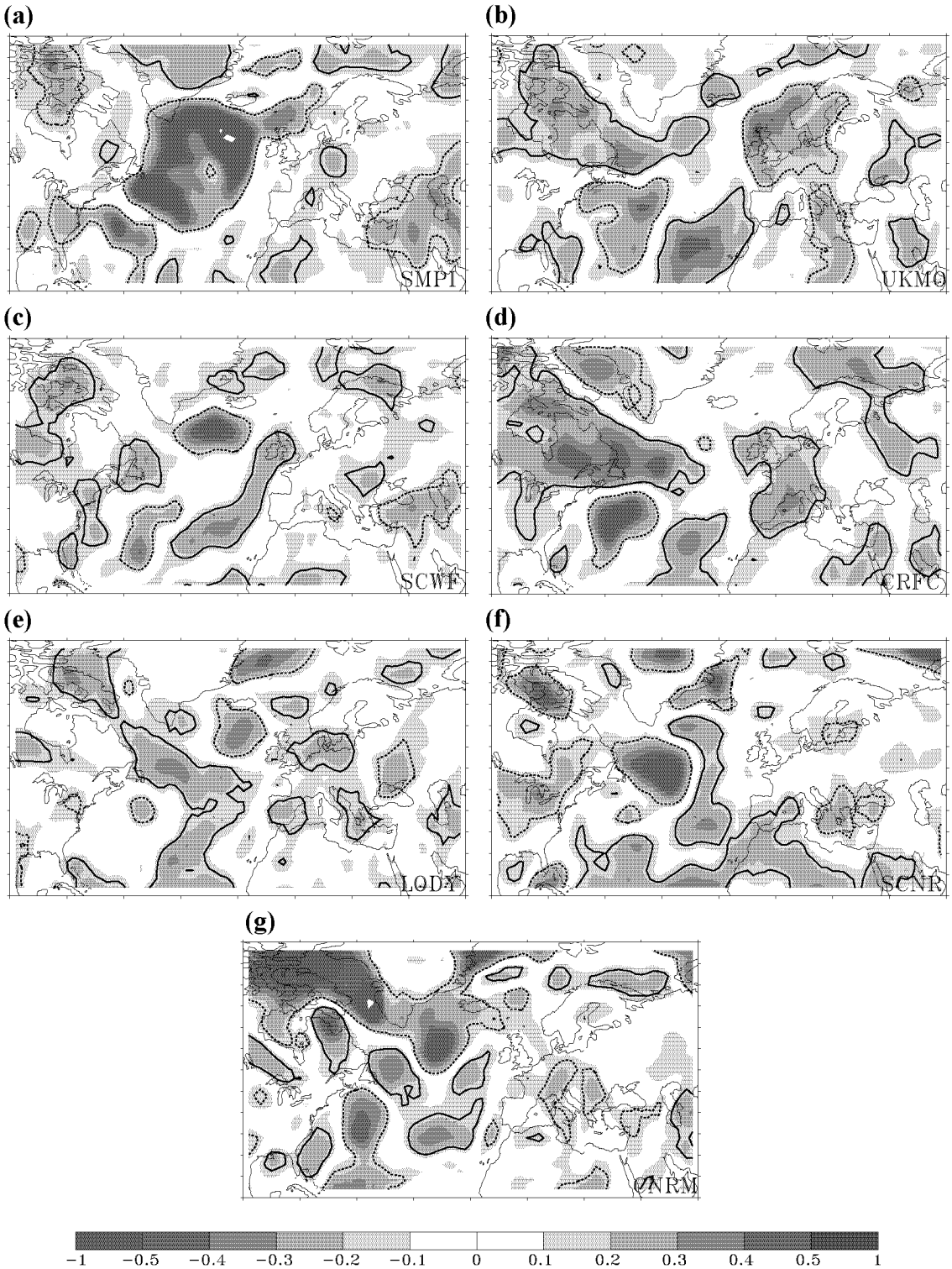


Figure 4-9 The $RPSS_D$ of the DJF mean 2m temperatures of the DEMETER models for the FA and lead time 1 month (shadings, 5 % intervals). The models from top-left to bottom right are: (a) SMPI, (b) UKMO, (c) SCWF, (d) CRFC, (e) LODY, (f) SCNR, (g) CNRM.

Spain, respectively, and off the coast of Newfoundland. In some models such as the SCWF the former is elongated up to the British Isle. Recurrent areas of negative skill scores are found off the US coast at about 30°-50°N latitude. The average of the models' skill score at each grid point is illustrated in (Figure 4-10b). Significantly positive skill scores are located over the eastern sub-tropical Atlantic (up to 25 %) and small regions of Canada and the European Polar Sea. Negative results are found off the US coast at about 40°N (up to -30 %) and over the North Atlantic at about 60°N (up to -25 %). Over the European continent no significant skill scores were found.

The results of the multi-model show regions with positive skills as described above (Figure 4-10a). Strong positive skill scores are found over north-eastern Canada (up to 35%), large areas of western Africa and the sub-tropical Atlantic (up to 20 %), the Middle East (up to 15 %) and the polar sea (up to 25 %). Negative skill scores are located over the western sub-tropical Atlantic, Greenland and northern Europe (up to -20 %). The average of the multi-model skill over the entire domain is about 3 % higher than the average over the mean of all individual models. However, the difference of the skill scores between the multi-model and the average of the models (not shown) varies locally. It provides a major improvement of the multi-model in the north-western Atlantic and Canada (+25 %), but a degradation in the sub-tropics (-25 %) and northern Europe (-15 %).

Perfect Model Approach (PMA)

Finally, the PMA is applied to the DEMETER system. Figure 4-11 shows the $RPSS_D$ for the PMA for all individual models. As expected, there is a high variation in the magnitude of the skill scores. However, as in the FA, there are some recurrent patterns.

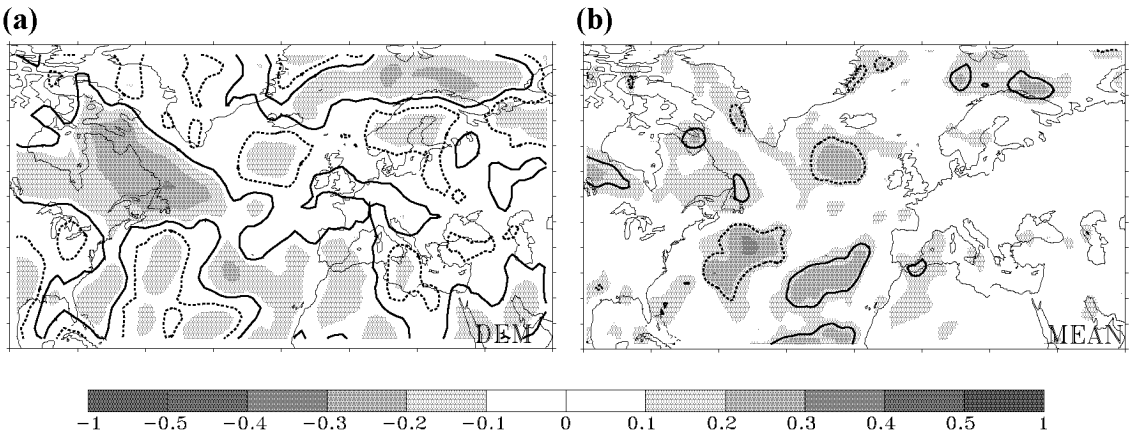


Figure 4-10: The $RPSS_D$ of the DJF mean 2m temperature of the multi-model (a) and the single-models' average (b). Shown are the skill scores for the FA and lead time 1 month. Shadings denote 5 % intervals.

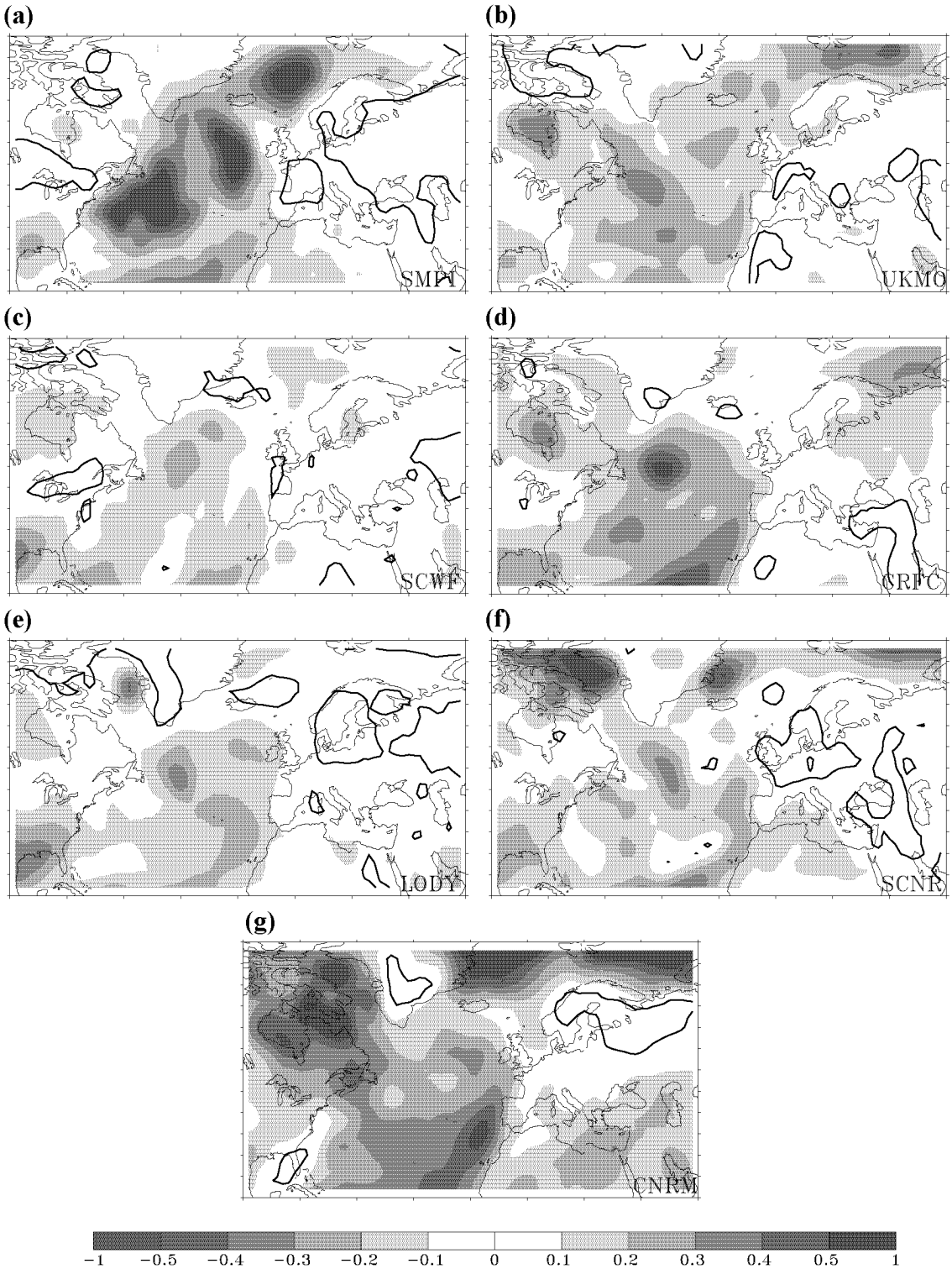


Figure 4-11: Same as Figure 4-9 but for the PMA. The models from top-left to bottom right are: (a) SMPI, (b) UKMO, (c) SCWF, (d) CRFC, (e) LODY, (f) SCNR, (g) CNRM.

Similar to the FA there is a tongue of positive skill score from the tropics along the west coast of Africa and Spain, respectively. Most models also provide a high degree of forecast skill in the mid-latitude Atlantic. However, their structures are less clear and mostly dominated by small localised maxima. Most obvious is the localised maximum in the region where the Gulf Stream leaves the American coast. Over the European continent skill scores with at least 10 % in magnitude are found in all models. Over the American continent the models show some degree of potential predictability. In particular over Canada almost all models provide skill scores of about 15 % and higher. But positive forecast skill scores up to 40 % are found over the south-eastern US. Some models such as the CNRM have exhausting skill over the Polar Sea and the northern American continent.

The average of all models' skill scores at each grid point is shown in Figure 4-12. The average identifies the above described regions of enhanced forecast skill. The high skill scores are found in local regions of the Atlantic basin and partly over Canada. The European and African land masses are covered by skill scores of less than 10 %. It must be noted that for this figure the mean at each grid point is estimated from a sample of only 7 skill scores and hence local maxima are affected by single models. Furthermore, the interpretation of such local maxima requires careful considerations, because the underlying scores are not calculated with cross-validated forecasts and reference forecasts.

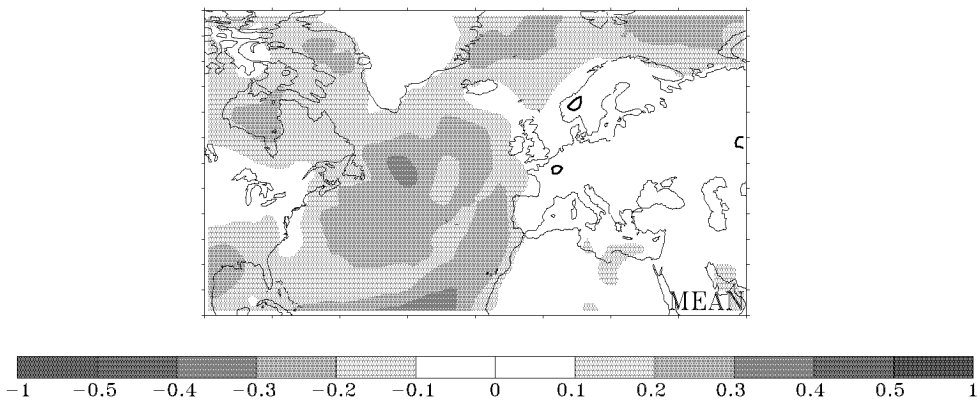


Figure 4-12: Same as Figure 4-10 but for the single models' average.

Chapter 5:

Probabilistic Seasonal Prediction of the Winter North Atlantic Oscillation and its Impact on Near Surface Temperature [9]

Müller, W. A., C. Appenzeller and C. Schär[10]

Abstract:

The North Atlantic Oscillation (NAO) is the major winter climate mode describing about one third of the interannual variability of the upper-level flow in the Atlantic European mid latitudes. It provides a statistically well-defined pattern to study the predictability of the European winter climate. In this paper the predictability of the NAO and the associated surface temperature variations are considered using a dynamical prediction approach. Two state-of-the-art coupled atmosphere-ocean ensemble forecast systems are used, namely the Seasonal Forecast System 2 from the European Centre for Medium Range Weather Forecast (ECMWF) and the multi-model system developed within the joint European project DEMETER (Development of a European Multi-Model Ensemble Prediction System for Seasonal to Interannual Prediction). The predictability is defined in probabilistic space using the de-biased ranked probability skill score with adapted discretization ($RPSS_D$). The potential predictability of the NAO and its impact are also investigated in a perfect model approach, where each ensemble member is used once as "observation". This approach assumes that the climate system is fully represented by the model physics.

Using the perfect model approach for the period 1959-2001, it is shown, that the mean winter NAO index is potentially predictable with a lead time of one month (i.e. from November 1st). The prediction benefit is rather small (6 % skill relative to a reference climatology) but statistically significant. A similar conclusion holds for near surface temperature variability related to the NAO. Again the potential benefit is small (5 %) but statistically significant. Using the forecast approach the NAO skill is not statistically significant for the period 1959-2001, while for the period 1987-2001 the skill is surprisingly large (15 % relative to a climate prediction). Furthermore a weak relation between the strength of the NAO amplitude and the skill of the NAO is found. This contrasts with El Niño/ Southern Oscillation (ENSO) variability where the forecast skill is strongly amplitude dependent. In general, robust results are only achieved if the sensitivity with respect to the sample size (both the ensemble size and length of the period) is correctly taken into account.

---

[9] Accepted by Climate Dynamics, revised version 12 July 2004

[10] Institute for Atmospheric and Climate Science, ETH, Zürich, Switzerland

5.1. Introduction

On seasonal to interannual timescales the North Atlantic Oscillation (NAO) provides the major mode of natural winter climate variability for the Atlantic European region. The NAO is historically described by the NAO index, a normalised sea level pressure difference between the Azores high and the Iceland low. Changes of the phases of the NAO index are accompanied by a modulation of location and strength of the storm tracks and synoptic eddy disturbances (Rogers 1997). Changes in storm track activity and the associated influence on the transport of atmospheric heat and moisture are closely linked to changes in regional winter temperature and precipitation (van Loon and Rogers 1978, Lamb and Peppler 1987, Cayan 1992, Appenzeller et al. 1998). A positive/negative phase of the NAO index is associated with a stronger/weaker than normal mean westerly flow and a pronounced warm/cold air advection towards northern Europe (for details see Hurrell 2003). Such climate anomalies have a major impact on social, economic and ecological sectors (see Marshall et al. 2001, Drinkwater et al. 2003, Mysterud et al. 2003, Straile et al. 2003). For example the energy consumption in Northern Europe is highly correlated to temperature variations related to NAO variability (Hurrell et al. 2003). Hence, a successful prediction of the climate variations in these regions would be beneficial for many applications.

The prediction of the NAO, however, is limited by the intrinsic chaotic nature of the atmosphere-ocean system. Small uncertainties in the initial conditions and model formulation have a strong impact on the seasonal predictability. Since the initial conditions are always faced with some observational inaccuracies, multiple integrations with different starting values are used to import these uncertainties into the forecasts. The predicted values are probabilistic in nature and allow a proper estimate of the level of uncertainty. Such ensemble prediction systems have proven to be successful in increasing the prediction skill for various applications (Palmer and Anderson 1994, Stern and Miyakoda 1995).

The reduction of the full set of the governing equations on a truncated model grid inevitably leads to impreciseness. This led to the idea of combining several models into a multi-model ensemble system (Tracton and Kalnay 1993, Krishnamurti et al. 1999, Kharin and Zwiers 2002) where the uncertainties due to initial conditions and model formulations are accounted for. The improved forecast performance by multi-model formulations has been shown for medium-range forecast (Evans et al. 2000, Graham et al. 2000) as well as for seasonal forecast (Palmer and Shukla 2000, Shukla et al. 2000, Palmer et al. 2004). Among others the aim of such a multi-model system, is to enhance the reliability and finally the predictive skill of the forecast system. Multi-model systems were developed under the PROVOST (Prediction of Climate Variations on Seasonal to Interannual Timescales) project (Palmer and Shukla 2000) and currently under the DEMETER (Development of a European Multi-Model Ensemble Prediction System for

Seasonal to Interannual Prediction) project (Palmer et al. 2004). Within the PROVOST, for example, the probabilistic skill of the multi-model winter mean precipitation and mid-troposphere geopotential height prediction over the northern hemisphere was found to be generally higher than that of any single model (Doblas-Reyes et al. 2000).

The NAO provides a statistically well-defined pattern to study winter climate predictions. It is the aim of this paper to quantify the predictability of the winter NAO and its impact onto near surface temperature taking into account various aspects of uncertainty. The results will allow for the elucidation of two questions: (I) How well can current state-of-the-art seasonal forecast-systems predict this major European climate mode? (II) Is the NAO itself predictable at all; and if so to what extent?

A first attempt to simulate the NAO on seasonal timescales with dynamical models has been performed within the PROVOST project (Doblas-Reyes and Stephenson 2003). In contrast to PROVOST, we use coupled atmosphere-ocean modelling systems to ensure the linkage between the two media that differ in their inherent time scales. One is the ECMWF operational Seasonal Forecast System 2 (Anderson et al. 2003), the second is the multi-model ensemble system DEMETER (Palmer et al. 2004). Special attention is paid to the skill in the framework of the limited available ensemble size. Within the DEMETER project several models with 43 years of seasonal hindcast climatology are available. Together with the long database and a perfect model scenario we provide a robust estimate of the winter mean predictability in the mid latitudes. The data and methods are introduced in section 5.2. An assessment of prediction skill for the NAO with single and multi-models is given in section 5.3. In section 5.4 we provide a skill assessment for predictions of NAO impact on temperature. A concluding discussion is given in section 5.5.

5.2. Data and Methods

Data

Retrospective forecast data of the ECMWF Seasonal Forecast System 2 and the multi-model DEMETER system were used. The observations stem from the ERA40 Re-Analysis Project[11].

All forecasting systems consist of an ensemble of coupled Atmosphere-Ocean integrations. The atmospheric component of System 2 is the ECMWF IFS model version23r4 with 40 vertical levels and reduced horizontal resolution of a T95L40 grid.

[11] http://www.ecmwf.int/research/era

System Name	Hindcast Period	# of Ensembles and Start Months	Sample Size in Perfect Model Approach (PMA)
Operational ECMWF System 2	1987-2001	40 (Nov) 5 (Sep, Oct, Dec)	600 75
DEMETER System	1987-2001	Single-Model 9 Multi-Model 7x9 (all Nov)	135 945
ECMWF System from DEMETER	1959-2001	Single-Model 9 Multi-Model 3x9 (all Nov)	387 1134

Table 5.1: Overview of the applied models, their hindcast period and sample size.

The ocean component is the HOPE (Hamburg Ocean Primitive Equation Model) version 2. A set of hindcasts from system 2 are available for the 15-year period. The hindcasts start at the first of each month and are available in 5-member ensembles except for May and November. Here 40-member ensembles are available. The ensembles are constructed by combining 5 ocean analysis with sea surface temperature perturbations and stochastic physics. We use the term 'lead time' to distinguish between different integration times (i.e. a seasonal forecast with lead time 1 month means e.g. that we consider a forecast for the winter period December-January-February (DJF) that starts on the first of November). For further details on the models see Anderson et al. (2003).

The DEMETER system consists of a set of 7 atmosphere-oceans models. The contributing partners are CERFACS (CRFC), ECMWF (SCWF), INGV (SCNR), LODYC (LODY), MPI (SMPI), MeteoFrance (CNRM) and MetOffice (UKMO). The horizontal resolution of the models varies from T42L19 to T95L40 in the atmospheric component. For further details see Palmer et al. (2004)[12]. We use the abbreviations in brackets for the models of each organisation. Each model is initialised with 9 ensemble members for each season. We refer to all models for the period of 1987 to 2001 and to the individual models (SCWF, UKMO, CRFC) for the period 1959-2001 (see Table 5.1 for details). The maximum forecast period is 6 months for all integrations.

The predictions were validated with ERA40 Re-Analysis data as observations. Two variables have been considered for verifying the forecasts. Monthly mean geopotential height at 500hPa (Φ500) was used to analyze the major mode of variability. The monthly

[12] http://www.ecmwf.int/research/demeter/general/docmodel/index.html

mean 2m temperature was considered as near surface climate impact parameter. We focus on the DJF mean climate for the North Atlantic region (100W-60E, 20N-80N).

Methods

The NAO was defined using an empirical orthogonal functions (EOF) analysis of the observed and modelled winter mean Φ500. The field variables were weighted by the cosine of the latitude. The EOF analysis for the individual models were performed using all ensemble members merged into one single time series, providing a long sample climatology (number of ensemble members times the hindcast period, see Table 5.1). The observed climate variability was analysed using the ERA40 winter mean period 1959/60–2000/01. We have chosen the longer period for the observations to ensure statistically significance following the criterion of North et al. (1982). The field variables of the observations and all ensemble members were then regressed to the corresponding EOF-patterns. The resulting correlation coefficients were finally normalised. This produced the loading of the principal components of each ensemble member and the observations and will be denoted modelled NAO index and observed NAO index, respectively.

The impact on the European winter mean temperature was defined by a singular value decomposition (SVD) analysis (Bretherton et al. 1992). We used the cross-covariance matrix of the two scalar fields of 2m temperature and Φ500. The resulting eigenvectors were denoted left (2m temperature) and right (geopotential) singular vectors. The magnitudes of eigenvalues were quantified with the squared covariance fraction (SCF). The strength of the coupling of the two fields was measured with the reconstructed variance fraction (RSCF) where the SCF was weighted with the homogeneous correlation of the left and right expansion coefficients (Widmann et al. 2003). The field variables were regressed with their corresponding singular vectors. For application perspective the singular vectors are prior multiplied by the standard deviations of the scalar fields (Appenzeller 2000). The resulting time series of the coefficients were denoted as modelled and observed impact index. The correlation between the observations and the ensemble median NAO and NAO-impact index were quantified with the Spearmen correlation coefficients. Statistical significance is shown with the corresponding p-value.

To obtain a multi-model NAO index, each model of the DEMETER project was used to produce an ensemble of NAO indices. Since the component models will each have a different representation of the NAO variability, the multi-model NAO indices were obtained collecting all models' NAO indices into one distribution. The same was done for the multi-model NAO-temperature-impact indices where the multi-model distribution was governed by the models' expansion coefficients (for the 2m temperature). This method inhibits, that the multi-model NAO indices and NAO-temperature-impact indices, respectively are dominated by one single model.

All ensemble prediction systems used here have an implicit measure of uncertainty introduced by the initial conditions (Anderson et al. 2003, Palmer et al. 2004). In these systems the predicted values are probabilistic in nature and hence the verification procedure requires a probabilistic skill measure. There is no general agreement on the best single skill. Widely used scores are Brier scores, relative operating characteristics (ROC), linear error in probability space (LEPS), or ranked probability skill score (RPSS) (Swets 1973; Mason 1982; Ward and Folland 1991; Stephenson and Jolliffe 2003; Wilks 1995). The choice depends on the particular application considered or the forecast system used. In this paper we focussed on the probabilistic forecast which includes the entire shape of the probability distribution. We used the ranked probability score (RPS) but with an adapted discretization of the climatological probabilities (see below). The standard RPS is defined as $RPS = 1/N \sum_{k=1}^{N} RPS_k$ where N is the number of forecasts, k is the forecast index and

$$RPS_k = \sum_{j=1}^{J} \left(Y_j - O_j \right)^2 \tag{5.1}$$

represents the mean square error of the cumulated probabilities of the forecast relative to the observations. Here J denotes the total number of classes. If not explicitly stated we use three equiprobable classes. The cumulative probabilities of the forecasts Y_j and the observations O_j are defined as $Y_j = \sum_{i=1}^{j} y_i$ and $O_j = \sum_{i=1}^{j} o_i$, where y_i and o_i are the probability of the forecast and observation, respectively, of class i. The cumulative probability of the observation becomes one for the observations lying in the corresponding and following classes. The RPS_k is zero in case of a perfect forecast (i.e. all ensemble members correctly predict the event) and positive otherwise. The skill calculation is based on the comparison of the forecast score (RPS_{FC}) to a reference forecast score (RPS_{REF}) relative to a perfect forecast. In this study the reference forecast is given by the observed climatology ($RPS_{REF} = RPS_{CL}$).

However, since a forecast system generates only discrete probability values the probabilities of the reference forecast also needs to be calculated with discrete probability values. This was done by choosing many random samples out of the climatological distribution, with their size equal to the number of ensemble members. For details see Müller et al. (2004). Here a re-sampling method is used by returning the selected climatological values into the original distribution. After summation the RPSS with adapted climatology ($RPSS_D$) becomes:

$$RPSS_D = 1 - \frac{q \sum_{k=1}^{N} RPS_{FC,k}}{\sum_{k=1}^{N} \sum_{l=1}^{q} RPS_{CL,k,l}} \tag{5.2}$$

where q is the number of randomly chosen climatological sub-samples. The so-defined de-biased $RPSS_D$ is free of the systematic bias of the RPSS (Müller et al. 2004). For a

perfect forecast the $RPSS_D$ is one and zero values indicate no benefit against the reference forecast.

The $RPSS_D$ values at the 95 % confidence level were calculated using a Monte Carlo approach and differ in ensemble size and forecast length. The boundaries of the classes of the forecast and reference forecast distribution were obtained from the tertiles of the full distribution of the climate forecast (i.e. by rank ordering of the NAO index and NAO impact index, respectively, and defining boundaries that separate the tertiles). If not explicitly stated otherwise we use this method. A more realistic situation was achieved by excluding the forecasted value from the climatology for each single forecast.

When using the operational ECMWF Seasonal Forecast System 2 in a forecast approach the database consists of only 15 winter seasons (FA15). To obtain a robust estimate of the winter climate predictability a larger sample size is desirable. In this paper two ways were investigated to enlarge the sample. In a first attempt we used the extended operational system 2 model data base from the DEMETER project, giving a sample size of 43 years. A forecast approach for this period is denoted as FA43. The second attempt was performed in terms of a perfect model approach (PMA). In this approach each ensemble member was considered once as the "observation" and not as member of the forecast distribution. This reduces the number of ensemble members by one, but including the "observation" would lead to an artificially high skill that is dependent on the ensemble size. In a synthetic white noise climate model with an expected skill of zero (see also Müller et al. 2004) the difference can be quantified. For a 5-member ensemble this amounts to about 25 %. The PMA substantially increases the sample size with respect to the number of ensemble members and hindcasts (see table 5.1 for details).

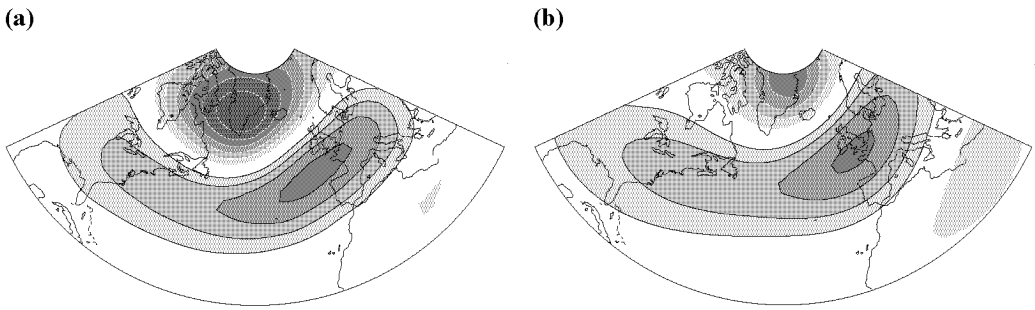


Figure 5.1: The first EOF of the winter (DJF) mean geopotential of (a) the observed atmosphere and (b) the ECMWF Seasonal Forecast System 2 with lead time 1 month. This mode explains 37 % and 30 %, respectively, of the total variance. The observed climatology is based on ERA40 data for the period 1959-2001 and the model climatology on the period 1987-2001 times the number of ensemble members (40x15). The isolines depict anomalies of Φ500, with an interval of 10hPa. Dashed lines refer to negative values.

5.3. Probabilistic Prediction of the NAO

Single Model Approach

The primary mode of the northern hemisphere seasonal winter climate variability in the observed data is displayed in Fig. 5.1a. The figure shows the characteristic dipole of the NAO with an anomalous low over Iceland and a high over the central Atlantic. This significant mode explains 37 % of the total variance. The first mode in the operational ECMWF System 2 with lead time 1 month is shown in Fig. 5.1b. This mode explains 30% of the total variance. The overall structure is comparable to the observations but the amplitude is damped. For all lead times the modelled NAO is the primary mode of the forecast data. The explained variances are displayed in the first line of Table 5.2. Figure 5.2 shows the observed and modelled NAO index for different lead times. For lead time 0

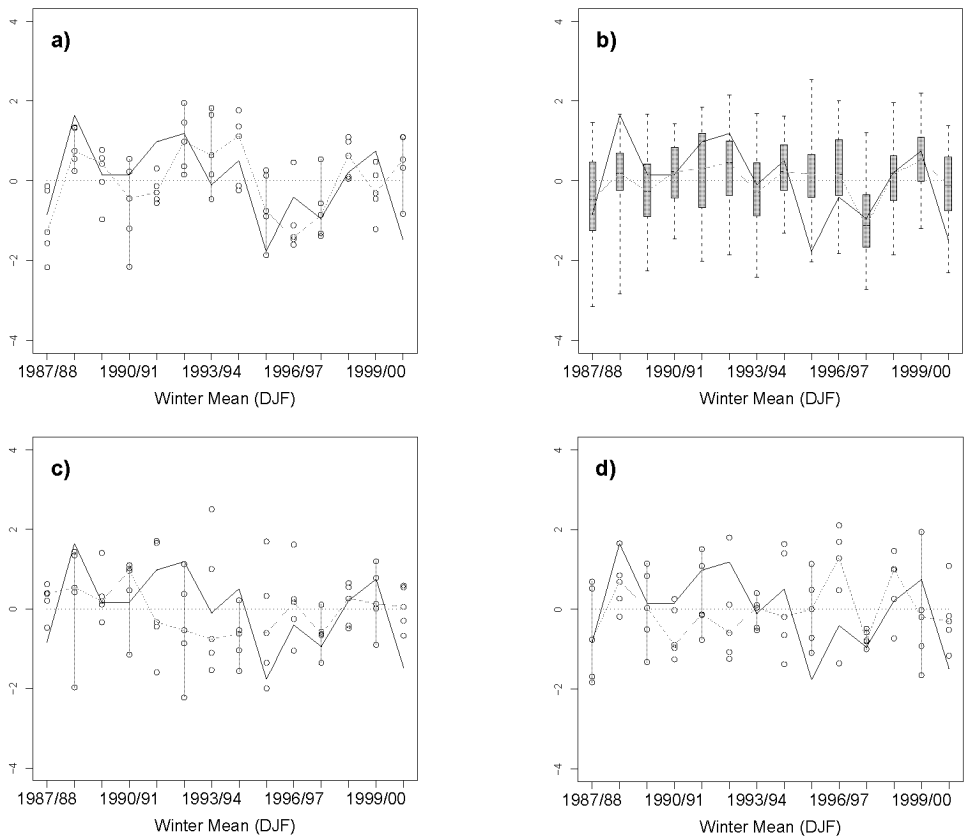


Figure 5.2: The observed NAO index (black lines) and the distribution of the NAO index modelled by the ECMWF Seasonal Forecast System 2 for (a) lead time 0 months, (b) 1 month, (c) 2 months and (d) 3 months. The grey lines show the ensemble median. The boxes in (b) are divided into quartiles. Dashed vertical lines denote the upper and lower quartiles. Note that there are 5 ensemble members in (a), (c) and (d), while there are 40 ensemble members in (b).

months (Fig 5.2a) the model NAO variability is similar to the observed NAO index. The correlation coefficients between the ensemble median NAO-index and the observations amount to 0.56 (p-value 0.03). However, these results reflect a strong contribution from relatively high predictability at a short-range to medium-range forecast. In Fig.5.2b the results for lead time 1 month are shown. The spread is much larger than for lead time 0 months, which is also due to the increased number of ensemble members (40). Obviously there is a lower variability in the ensemble mean NAO index than for lead time 0 months. However, the ensemble mean NAO index mostly follows the observations and a correlation coefficient with a statistically significant value of 0.76 (p-value 0.02) is found. For the longer integration periods (Figs 5.2c and 5.2d) the modelled NAO index does not follow the observations, although years with relatively marked NAO index are best predicted. In particular predictions for the years 1988/89 and 1997/98 show a good consistency with the observed NAO index for all lead times. However, the corresponding correlation coefficients indicate no statistically significant resemblance for lead times greater than one month (Table 5.3).

We used the distributions of the modelled and observed NAO indices to quantify the probabilistic forecast skill. Figure 5.3 shows the $RPSS_D$ as a function of integration time for the operational ECMWF Seasonal Forecast System 2. Statistically significant skills are found for lead time 0 and lead time 1 month. Longer integration periods

	Lead 0	Lead 1	Lead 2	Lead 3
EOF	33 %	29 %	31 %	37 %
SCF	58 %	60 %	58 %	71 %
RCOR	0.55	0.57	0.54	0.67

Table 5.2: The statistical quantities of the EOF and SVD analysis. Shown are the explained variances in the EOF analysis (EOF) and the reconstructed (RCOR) and the fraction of covariance (SCF) in the SVD analysis of the ECMWF Seasonal Forecast System 2 for different lead times.

System 2	Lead 0	Lead 1	Lead 2	Lead 3
NAO	**0.56** (0.03)	**0.76** (0.02)	0.24 (0.39)	0.08 (0.78)
NAO-impact	**0.49** (0.05)	0.39 (0.16)	0.08 (0.77)	0.05 (0.85)

Table 5.3: The correlation coefficients and the corresponding p-values of the NAO and NAO-temperature-impact of the operational ECMWF Seasonal Forecast System 2 for different lead times. Statistically significant results with at least 95 % confidence are printed in bold.

(lead time 2 and 3 months) did not show statistically significant skills. A closer inspection of these skills revealed that single winter seasons have a strong impact on the skill. For instance, when the winter 1997/98 was excluded from the analysis the skill for lead time 1 month decreased by about 6 %. This effect has also been considered in the work of Doblas-Reyes and Stephenson 2003) where the prediction skill of the NAO was examined for the PROVOST data.

In order to increase the robustness of the analysis the sample size was first increased by using the perfect model approach (PMA). The results as a function of integration time are shown in Fig. 5.3 (white bars). It can be seen that for all lead times the skills of PMA are smaller than those obtained from the forecast approach, but the values are still statistically significant for lead times 0 and 1 month. Furthermore, the skills in the PMA are not as sensitive to the climatology as in the case of the FA. However, since the PMA indicates a robust skill measure, the results suggest an overestimation of the skills for the forecast approach (grey bars).

We further determined the skill based on a longer period. For this purpose we used the period 1959-2001 of the DEMETER System. In Fig. 5.4 the $RPSS_D$ for lead time 1 month is shown for the operational ECMWF Seasonal Forecast System 2 (left bars) and the SCWF model (middle bars) for the period 1987-2001. The skills of the SCWF model for the shorter period are comparable with system 2. The results of the SCWF model for the extended period are shown in the right bars. The skills of the FA are reduced to about 3%. The skills of the PMA amount to 5 %-6 % for all periods.

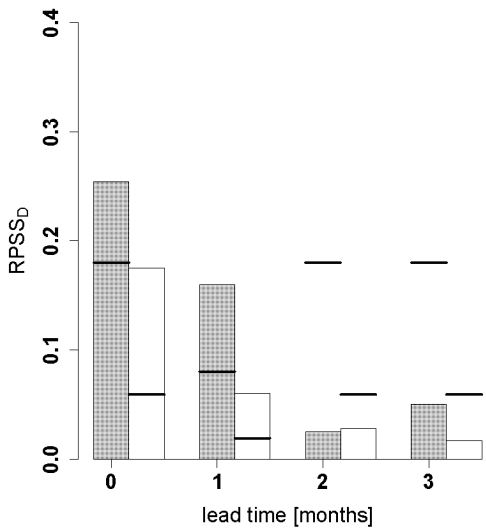


Figure 5.3: The debiased ranked probability skill score ($RPSS_D$) of the NAO index modelled by the ECMWF Seasonal Forecast System 2 as a function of lead time. Three equiprobable classes were used. The reference forecast is based on the observed data referring to the period 1987-2001. Forecasts for lead time 0 month, 2 months, 3 months are based on 5-member ensembles, whereas forecast for lead time 1 month result from 40-member ensembles. Grey bars show the skill of the forecast with respect to the observed NAO index (FA15). White bars correspond to the perfect model approach (PMA). The horizontal lines denote the 95 % significance levels, which are a function of sample size.

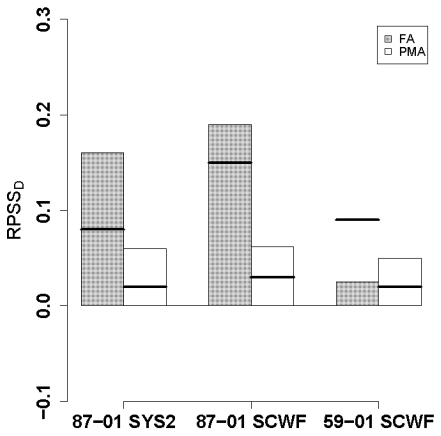


Figure 5.4: The $RPSS_D$ of the NAO index for lead time 1 month for the operational ECMWF Seasonal Forecast System 2 (SYS2) and the ECMWF model from DEMETER system (SCWF) for different periods. Grey bars denote the $RPSS_D$ with the forecast approach; white bars show results for the perfect model approach. Three classes were used. The horizontal lines denote the 95 % significance levels. (See Table 5.1 for the ensemble and sample size).

In summary, these results suggest, that the prediction with one month lead time of the winter mean NAO index based on coupled models give an improvement of about 6% relative to a simple prediction based on climatology. Our results underline the large sensitivity of any quantitative prediction skill on the sample size and the climatology chosen and that the most robust estimation of the potential predictability of the NAO index is achieved by the PMA method.

Multi Model Approach

In this section we use the recently developed DEMETER multi-model data to quantify the prediction of the NAO. The NAO is the primary mode of variability for all models with about one third of the total variance. The distribution of the multi-model NAO indices (not shown) is characterised by a large spread of the predicted ensemble members, but the ensemble mean index generally follows the observed NAO variability. From the last sub-section it can be seen that the skill is strongly dependent on the sample size. We therefore start the discussion of the forecast verification with the period 1959-2001. For both the multi-model and all individual models (see Table 5.4) no significant correlation coefficients between the predicted medians and the observations are found. The $RPSS_D$ of the multi-model ensemble and the individual models are shown in Fig.5.5a. In the FA43 the skill of the multi-model ensemble is about 4 % higher than the average of the individual models. However, the skills are not significant. Within the PMA the skills of the individual models are in the range from 5 %-8 %, and the average is about 6 %.

For the period 1987-2001 the results are different. The correlation coefficient between the multi-model ensemble median NAO index and the observations amounts to 0.67 (p-value of 0.01). The correlation coefficients and significances of all individual models are

displayed in Table 5.4. Three single models (SCWF, LODY and CRFC) show significant correlation at the 95 % confidence level. The probabilistic skills for individual models, the model average and multi-model are shown in Fig. 5.5b. The values range from -3 % (CNRM) to 21 % (LODY). Only the models of SMPI, SCWF, LODY and SCNR show significant values. The average of all models amounts to about 10 %. The skill of the multi-model distribution, however, is higher than the mean of the models and amounts to about 17 %. The potential predictability for the single-model forecasts are shown in Fig. 5.5b (white bars). It can be seen that the skills are generally decreased but remain statistically significant for all models except UKMO and SCNR. The average skill for the models´ PMA amounts to a value of 7 %.

	DEM	CNRM	SMPI	UKMO	SCWF	SCNR	LODY	CRFC
87-01								
NAO	**0.67** (0.01)	0.10 (0.71)	0.10 (0.74)	0.01 (0.9)	0.56 (0.3)	0.13 (0.6)	**0.61** (0.02)	**0.64** (0.02)
NAO-Impact	0.37 (0.19)	0.01 (0.98)	0.05 (0.86)	0.14 (0.69)	0.49 (0.07)	0.06 (0.84)	0.40 (0.15)	0.24 (0.4)
59-01								
NAO	0.01 (0.9)	0.19 (0.24)		-0.29 (0.11)	0.05 (0.9)			
NAO-Impact	0.17 (0.25)	0.31 (0.06)		-0.08 (0.61)	0.14 (0.35)			

Table 5.4: Same as Table 5.3 but for all DEMETER models and the multi-model ensemble (DEM) for lead time 1 month. Shown are the results for the period 1987-2001 and 1959-2001.

(a)

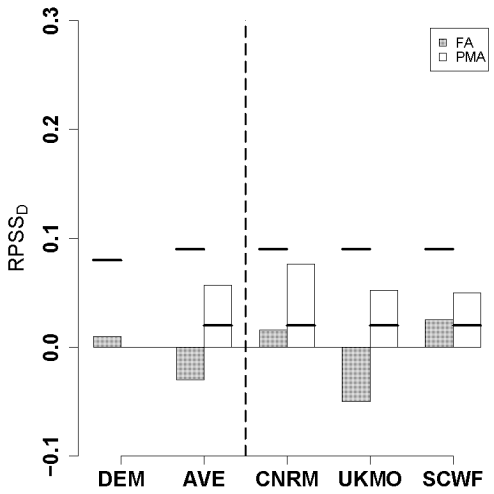


Figure 5.5: The $RPSS_D$ of the NAO index of the DEMETER models for lead-time one month for the periods (a) 1959-2001 and (b) 1987-2001. The skills for the multi-model ensemble (DEM), the model average (AVE) and the individual models are shown. Grey bars denote the $RPSS_D$ for the forecast approach while the white bars denote the perfect model approach. The horizontal lines denote the 95 % significance levels. Three equiprobable classes were used.

Skill vs. Amplitude

As suggested by several authors, the skills of seasonal to interannual predictions of climate patterns are a function of the signal-to-noise ratio (Shukla 2000, Kumar et al. 2001). For instance, in the dynamical seasonal prediction project (DSP) it has been shown that the skills of the extra-tropical Pacific North America pattern can be associated with the intensity of El Niño events. Strong events are usually forecasted with high skill.

(b)

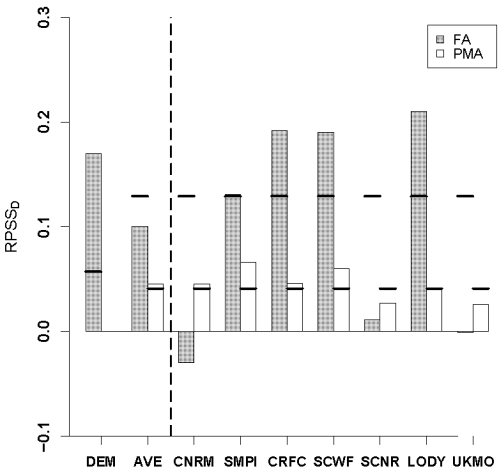


Figure 5.5: continued.

Hence it is attractive to consider the skills of single NAO events with respect to their amplitude. Figure 5.6 shows the $RPSS_D$ for the multi-model DEMETER ensemble, the multi-model and the Seasonal Forecast System 2 as a function of the observed NAO amplitude for the FA15. There is a large variability in the results for the single winter seasons and the models. For example, the SMPI model performs best with a maximum skill for the 1997/98 winter, while the CNRM model shows a minimum in 1989/90. Some single winter seasons such as the strong 1988/89 NAO show positive skill throughout all the models. Others such as 1987/88 or 1997/98 have at least a strong positive tendency. Furthermore, there are a number of winters with low intensity and low skill such as the winters 1990/91 and 1998/99. In summary, there is a tendency of pronounced skills for intense winter seasons. But overall, no clear separation between the high and low intensities can be found. Winter seasons with low intensity neither show intense skills, nor a uniform direction of the skills. Similarly no clear separation was found for the period 1959-2001 (not shown).

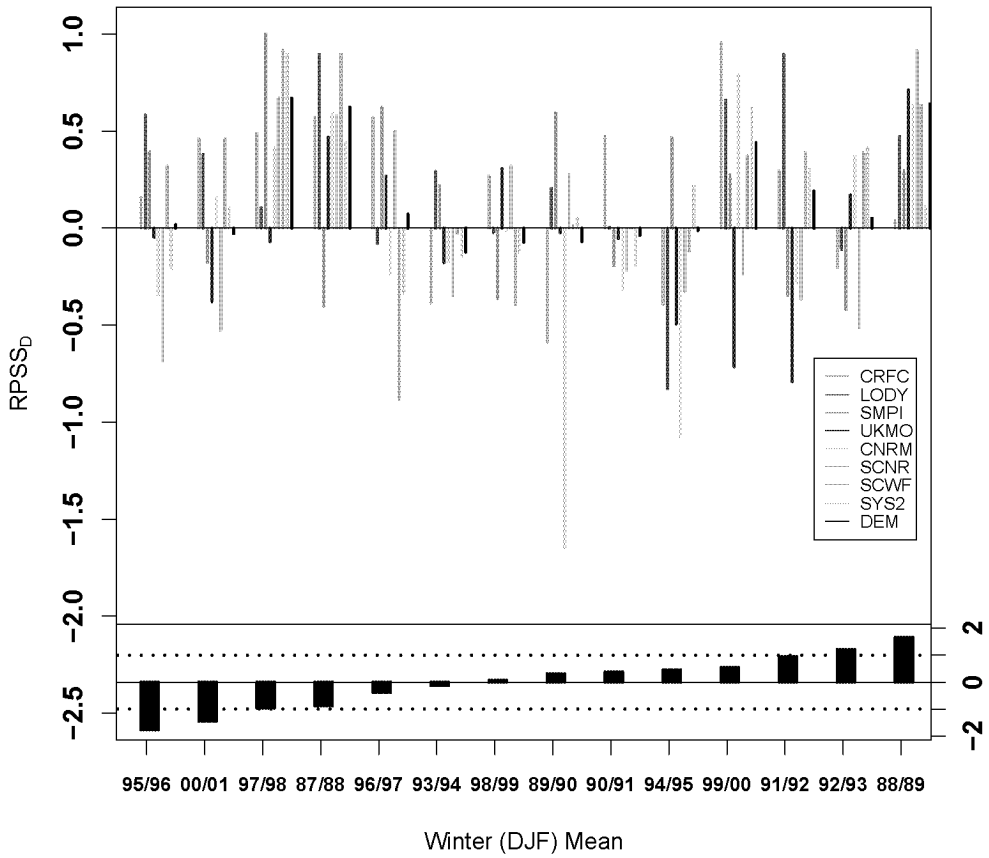


Figure 5.6: The $RPSS_D$ for various winter NAO predictions (upper) as a function of the observed NAO index (lower). Shown are the operational ECMWF Seasonal Forecast System 2 (SYS2), the individual DEMETER models (see inlet) and the DEMETER multi-model (DEM) for the period 1987-2001. In the lower part the amplitude of the NAO and its standard deviation is shown

In order to obtain a predictive relation, Fig. 5.7a shows the $RPSS_D$ (from Fig 5.6) as a function of the predicted ensemble mean NAO index. For the period 1987-2001 a hyperbolic relationship is found between the forecast skill and the forecasted intensity. Although single outliers exist for single winter seasons and individual models, a strongly positive and negative forecasted ensemble median NAO is associated with an enhanced predictive skill, indicating higher credibility. For seasons with low to intermediate forecasted intensity the results suggest an interpretation with proviso. For example three single seasons such as the intense winter NAO of 1988/89, the intermediate season of 1990/91 and the El Niño season 1997/98 are shown in colour. It can clearly be seen that the strong winter season 1988/89 is located in the right tertiles of the figure suggesting a high predictive value. Moreover, in 1990/91, a year with near-zero observed NAO index, there were no forecast false alarms, with systems predicting a range of NAO values in the centre of distribution, but no extreme values. The winter 1997/98 again shows high predictive skill, this time associated with a distinctly negative amplitude. A similar analysis of the period 1959-2001, when the skill of the NAO index was not significant, does not support such a predictive relation (Fig. 5.7b).

A separate examination for the period 1987-2001 (not shown) reveals that the scores of the forecasts decrease for strong NAO amplitudes at the ends of the diagram. The scores of the reference forecasts are independent of the predicted NAO amplitude index. However, they increase for stronger observed winter seasons. This supports the results that the increase of the skills in Fig 5.7a can be partly explained by the forecasts.

5.4. Probabilistic Prediction of the NAO Related Temperature Variability

Single Model Approach

In general, practical applications are not dependent on the state of the NAO index directly, but rather on the associated changes in near surface climate parameters. To define a NAO-related near surface temperature impact we used an SVD analysis (see section 5.2). Figure 5.8 shows the leading pair of eigenvectors of the observations and system 2 for lead time 1 month. As can be seen the right eigenvector (geopotential) of the observations and the Seasonal Forecast System display the NAO-like pattern. These modes explain 56 % and 60 % of the SCF of the observation and system 2, respectively. The SCF and their RSCF for all other lead times up to three months are displayed in Table 5.2. The left eigenvector (temperature) shows the characteristic features of responding warm

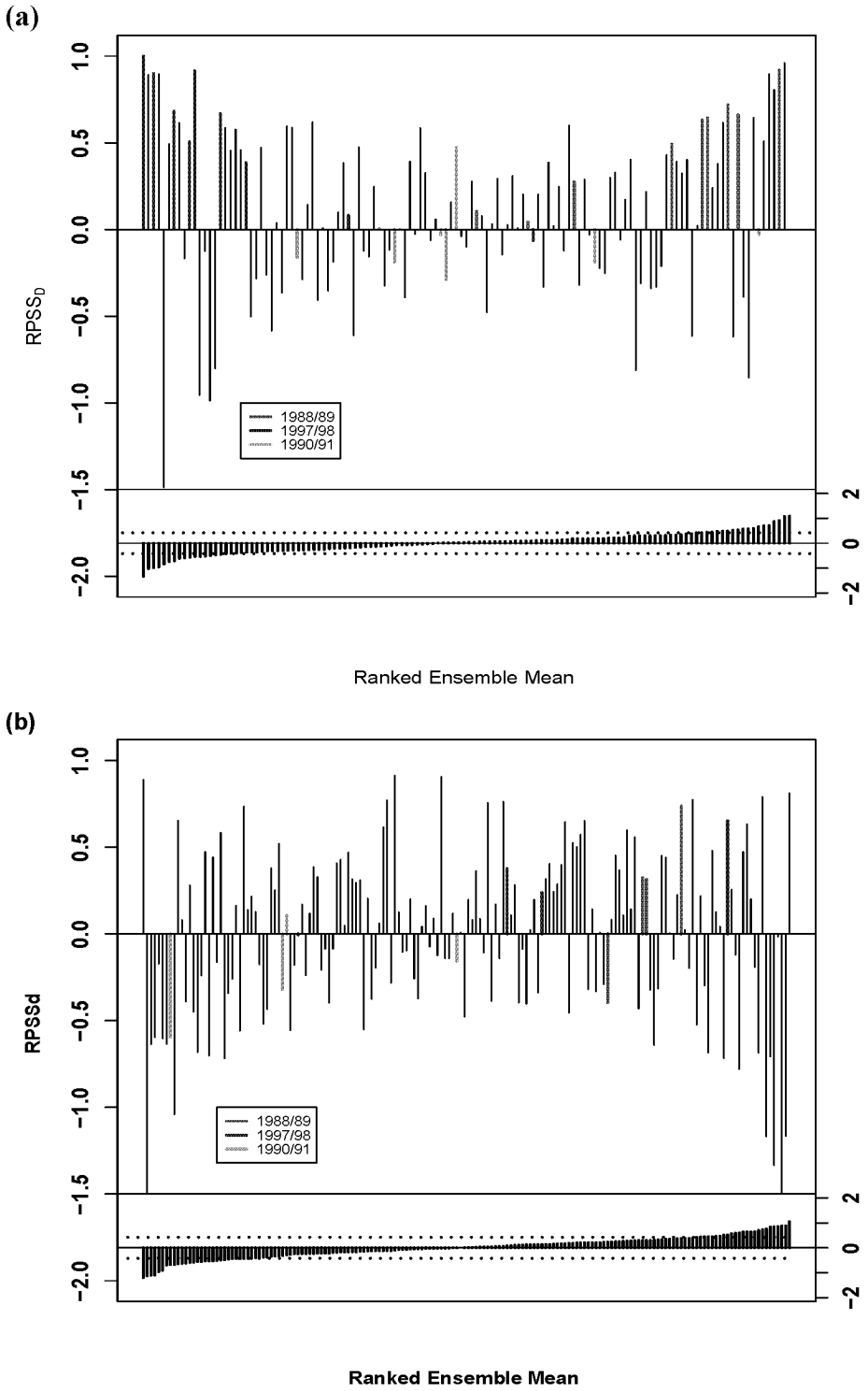


Figure 5.7: The RPSS$_D$ as a function of the strength of the predicted ensemble mean NAO index (upper) for lead time 1 month. The results show the ECMWF model and the DEMETER models for the periods (a) 1987-2001 and (b) 1959-2001. In the lower panel the amplitudes of the predicted ensemble mean NAO are shown. Dotted lines show the standard deviations. Single winter seasons (see inlet) are displayed in colours..

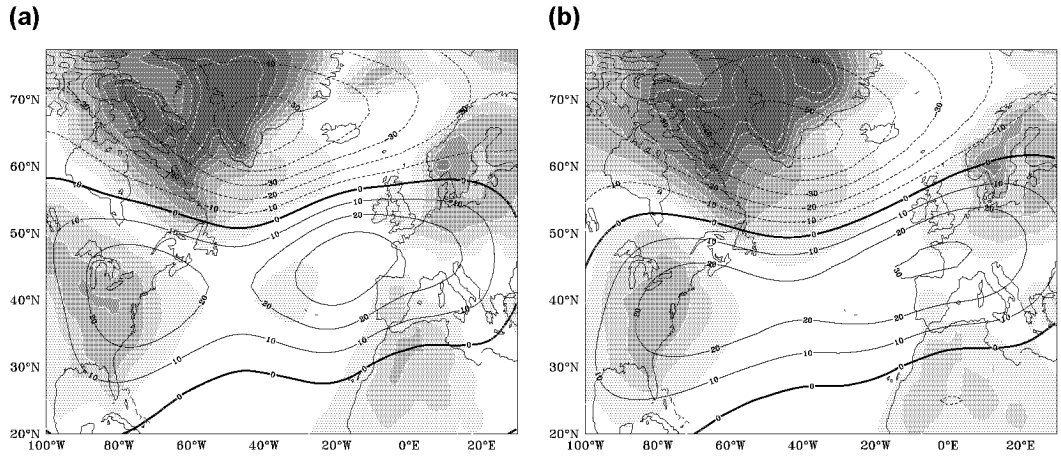


Figure 5.8: Same as Fig. 5.1 but for the first pair of eigenvectors of the SVD analysis of the winter mean (DJF) Φ500 (contours, 10hPa interval) and 2m temperature fields (shadings, contour interval 0.1°). This mode explains 56 % and 60 %, respectively of the fraction of squared covariance.

anomalies in northern Europe and south-eastern America and cold anomalies over northern Africa and north-eastern America. The correlation coefficients of the resulting pair of expansion coefficients amount to 0.90 for the observations and 0.94 for the forecasts for lead time 1 month, respectively.

As described in the previous section the regression of the temperature fields onto the corresponding eigenvectors gives the distribution of the temperature-impact indices. Figure 5.9 shows the observed and modelled mean NAO-impact index for different lead times. The modelled mean NAO-impact index shows its highest variability for lead time 0 months. The correlation coefficient between the observed and the forecasted ensemble median amounts to a statistically significant value of 0.49 (Table 5.3). For lead time 1 month the results are dominated by the spread of each forecast. The variability of the predicted ensemble median NAO-index roughly follows the observations but is smaller than for lead time 0 months. The correlation coefficients between the ensemble median and the observation, however, are not statistically significant.

We used these distributions to quantify the skill with the $RPSS_D$. In Fig. 5.10 the skills are shown for the FA15 (grey bars) and the PMA (white bars). The $RPSS_D$ has a major decrease with increasing lead time. Statistically significant values of about 25% and 16% were only achieved for lead time 0 and 1 month, respectively. Longer integration periods show no significant positive skill. As pointed out in the previous section these results are strongly affected by the small sample size. A more robust estimation of the skill was achieved by the two methods introduced in the previous section. For the PMA a similar reduction was found with skills of 12% and 6% for lead time 0 and 1 month, respectively. Longer integration time yields no significant results. The robustness with respect to the extended database is illustrated in Fig 5.11. The left and middle bars show the results of

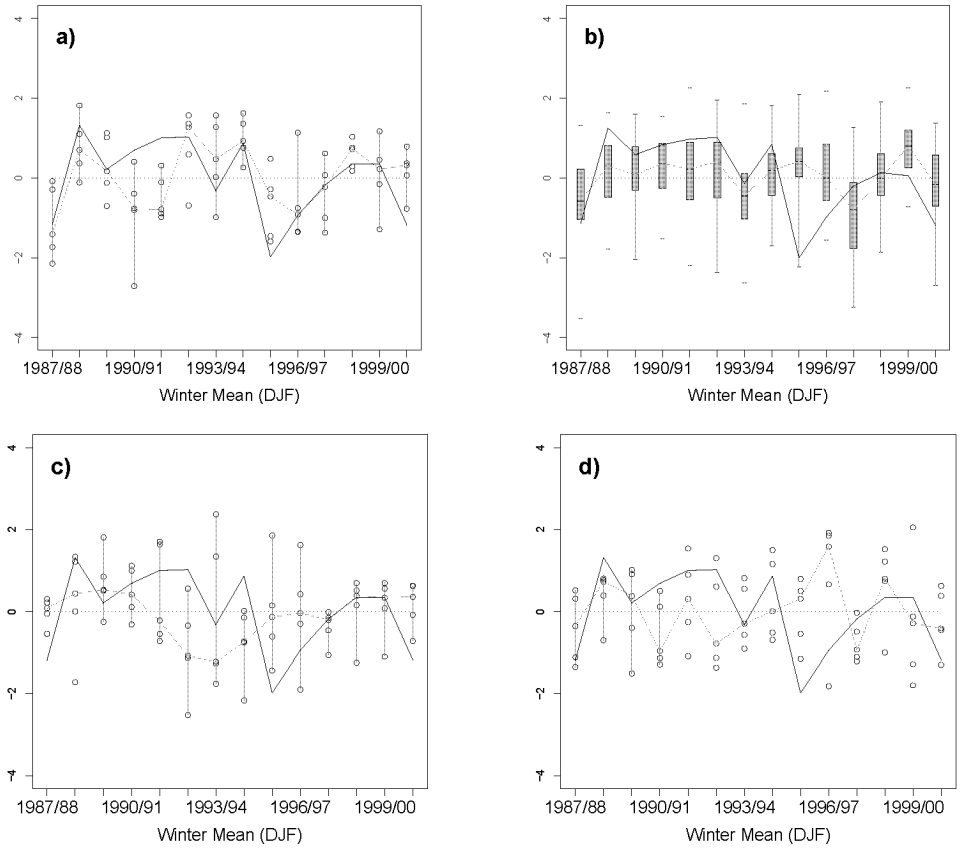


Figure 5.9: Same as Fig. 5.2 but for the NAO-temperature-impact index (black lines). The coefficients refer to the corresponding leading left eigenvector (T2m) of the SVD analysis. Note that there are 5 ensemble members in (a), (c) and (d), while there are 40 ensemble members in (b).

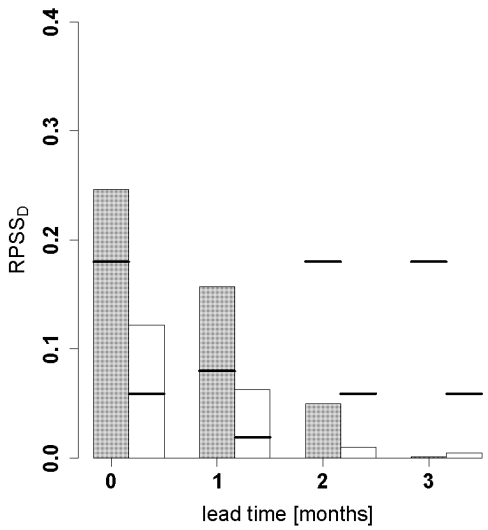


Figure 5.10: Same as Fig 5.3 but for the modelled NAO-temperature-impact index.

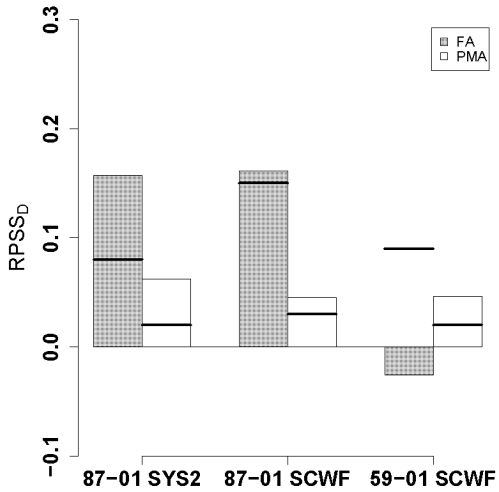


Figure 5.11: Same as Fig.5.4 but for the modelled NAO-temperature-impact index.

the operational ECMWF Seasonal Forecast System 2 and the SCWF model for the period 1987-2001 and lead time 1 month. As can be seen the results are comparable. The right bars show the results for the extended period. The skill for the FA43 has dramatically decreased to a value of about -3 % which is not statistically significant. The PMA, however, remains consistently significant with a value of about 5 %-6 %. Hence the results indicate that the NAO-related temperature impact has a potential benefit greater than the climatology for lead time 1 month.

Multi Model Approach

Next we investigate the predictability of the NAO-temperature-impact index for the multi-model distribution. As in the previous section we start with the period 1959-2001. The correlation of the ensemble median with the observed NAO-temperature-impact index (Table 5.4) shows no significance, neither for the multi-model ensemble, nor for individual models. The corresponding skill is shown in Fig. 5.12a. The NAO-impact index shows no significant positive skill. The skills range from -8 % (UKMO) to 3 % (SCWF) and 4 % (CNRM). The skill of the multi-model ensemble is somewhat higher (1%) than the average of the individual models. The skills of the PMA (white bars) suggest a potential predictability of 5 % to 8 %.

For the period 1987-2001 the correlation coefficients of the ensemble median with the observed NAO-temperature-impact index are also listed in Table 5.4. The highest values are achieved with the multi-model, SCWF and LODY. However, the models do not

(a)

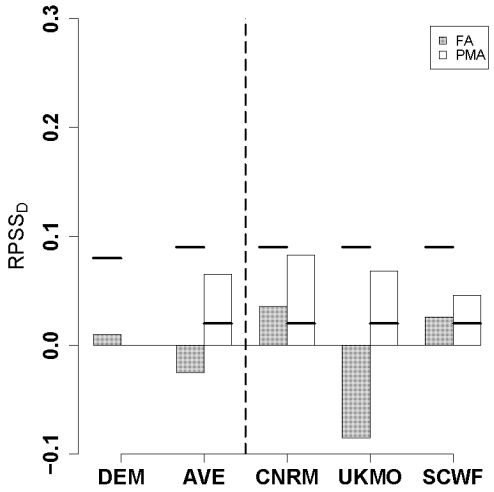
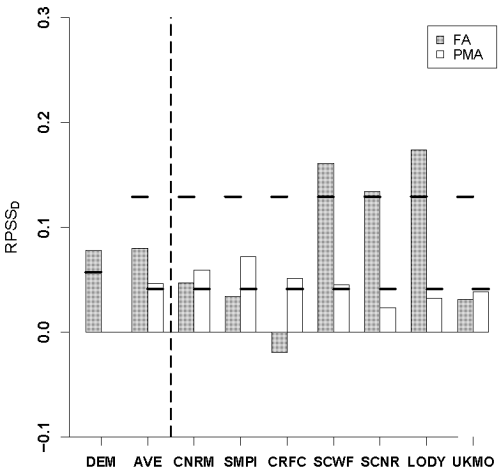


Figure 5.12: Same as Fig. 5.5 but for the modelled NAO-temperature-impact index.

provide statistical significance. The probabilistic skills are summarised in Fig. 5.12b. The skills range from -2 % (CRFC) to 17 % (LODY) for the FA15 (grey bars). The models of the SCWF, LODY and SCNR show a significant segregation to purely random chosen values. The average of all models amounts to a non-significant value of about 8 %. The skill of the multi-model distribution also amounts to about 8 %. However, due to the larger sample size the skill is statistically significant. The skill for the PMA is also shown in Fig. 5.12b (white bars). The potential predictability is massively reduced ranging from 3 % (UKMO) to 7 % (SMPI). Statistical significance is found throughout all models apart from SCNR and LODY. The skills for the average of all models amount to 5 %.

(b)

Figure 5.12: continued.



Skill vs. Amplitude

Finally, the skills of the single NAO-impact seasons were considered with respect to the observed and forecasted amplitudes. As described in the previous section first the NAO-impact skills are ordered with respect to the intensity of the observed NAO-related temperature impact index. Figure 5.13 shows the results for the DEMETER models, the multi-model and the operational ECMWF Seasonal Forecast System 2. In general there is no clear separation of pronounced skill with intense NAO-temperature-impact, and the figure shows a larger variability of skills among the models. An almost perfect forecast is obtained by the multi-model and SMPI for the winter season 1997/98. The minimum skill is achieved by the CNRM model in 1994/95. There are several seasons where the models show a positive tendency of the skills such as the El Niño winters of 1987/88 or 1997/98. The NAO-impact skills ordered with respect to the predicted ensemble mean impact index (Fig. 5.14a) show no clear separation either.

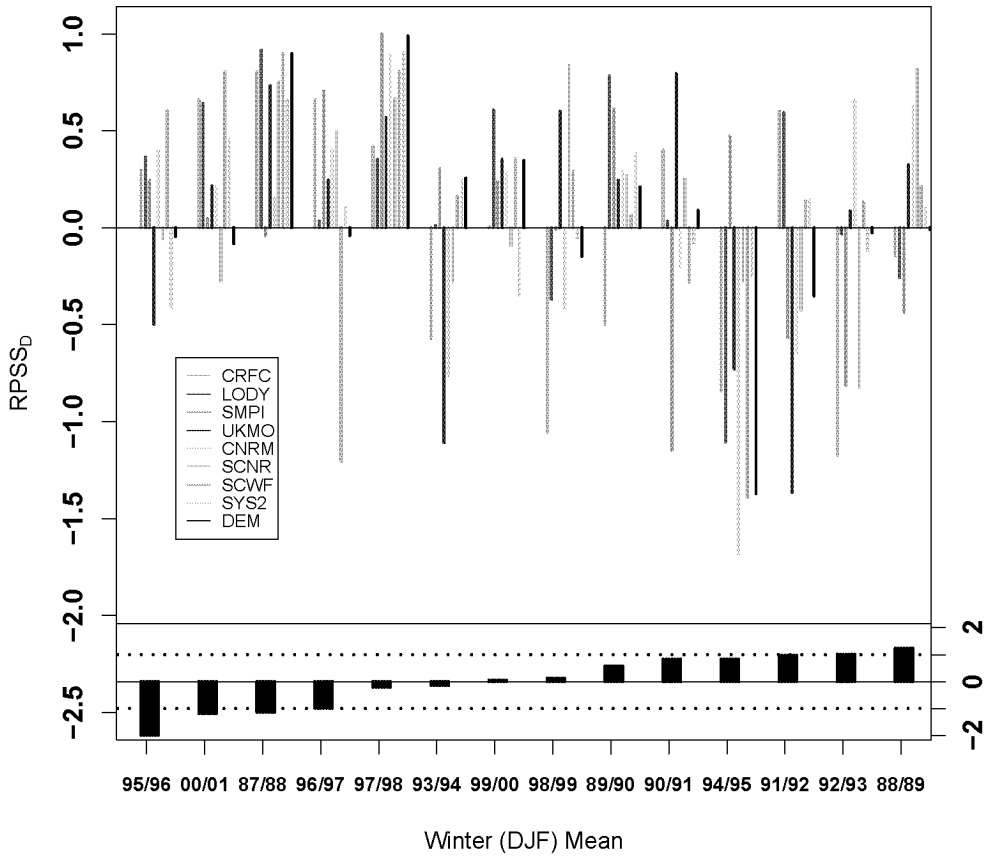


Figure 5.13 Same as Fig. 5.6 but for the NAO-impact onto 2m temperature.

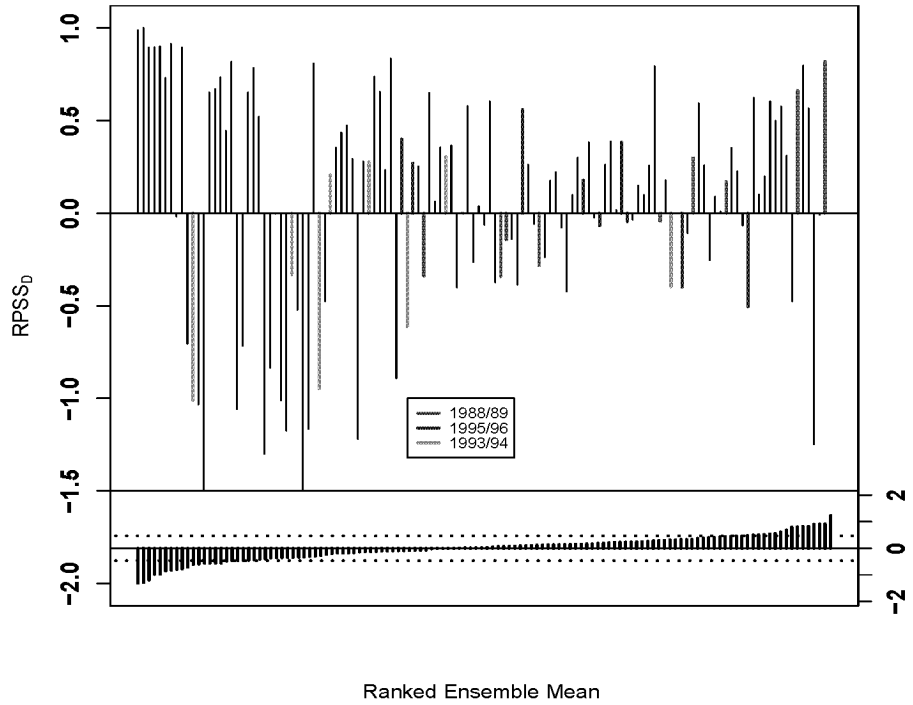


Figure 5.14: Same as Fig. 5.7 but for the NAO-impact on 2m temperature.

5.5. Conclusion

In this study the probabilistic predictability of the NAO and its impact on near surface temperature has been investigated. The models used were fully coupled atmosphere-ocean systems, namely the ECMWF Seasonal Forecast System 2 and a set of 7 models of the European DEMETER project.

The NAO is known to be the major mode of winter variability in the observations over the European-Atlantic region. The same was found to be true for the ECMWF system2 and the component models of the DEMETER multi-model. The probability distributions of the modelled and observed NAO indices have been obtained by regression of the eigenvectors to the corresponding field variables. We quantified the predictability in terms of a probabilistic verification measure developed in Müller et al. (2004), namely the debiased ranked probability skill score with adapted discretization ($RPSS_D$). In general the NAO skills are small but positive with respect to a climatology forecast and have a major decrease with integration time and sample size. The best estimate of the potential predictability was achieved by using the perfect model approach (PMA). The skill is found to be about 6 % for lead time 1 month and is consistent for different periods. A similar value is found for the forecast approach of the ECMWF model from

the DEMETER data set (SCWF) and the extended period of 1959-2001. However, if only the short period of 1987-2001 is taken into account a probabilistic skill of about 15 % and 18 % is obtained for the operational ECMWF System 2 and the SCWF, respectively. The multi-model suggests a benefit of about 17 % for the same period and performs better than the average of the DEMETER models (10 %). Hence the multi-model yields an enhancement of NAO skill related to the improved representation of uncertainty.

In summary, the results suggest a small but statistically significant skill for lead time 1 month. These results also show that the high skill for the short period 1987-2001 can be attributed to the small sample size. Single forecasts have a high impact on the skill for the short period (Doblas-Reyes and Stephenson 2003). For the extended period 1959-2001 the impact is considerably low.

Additionally, we examined the influence of the amplitudes of the observed and forecasted NAO indices on the forecast skills. Such an amplitude-dependent forecast skill was found in the equatorial Pacific region, where ENSO is the primary interannual variability, but also in the extra-tropics where it has a modulating impact on the PNA region (Straus and Shukla 2000, Shukla et al. 2000). In general the skills of the modelled NAO indices show a large variability in time and models. However, if all models are considered for the period 1987-2001 a tendency towards higher skill is detectable for stronger predicted ensemble mean NAO amplitudes. Predicted winter NAO with lower intensity show no preferred sign of the skill. For the period 1959-2001 no tendency is found. Although no general quantification can be given, the results encourage NAO prediction as long as its predicted intensity is strong enough.

The predictability of the NAO-related near-surface temperature impact was also quantified. In a cross-covariance analysis it was shown that the Seasonal Forecast Systems are capable to forecast the pattern of near-surface temperature linked with the NAO. It was found that models have skill in predicting climate variability impact at least one season (one month lead) in advance. However, similar to the NAO index the benefit is small and the best estimate of the potential predictability is achieved from the PMA. The PMA provides a skill of the NAO-temperature-impact index of about 5 %. A similar skill is achieved, if the extended period is applied for the forecast approach 1959-2001. A general enhancement was found for the short period 1987-2001. Here a beneficial value of about 16 % was assigned to the operational ECMWF system 2 for lead time 1 month. The multi-model NAO-temperature-impact index was found to have a statistically significant value of about 8 %. However, we found no further increase with respect to the average of all DEMETER models.

Finally, the skills of the single NAO-temperature-impact events were considered with respect to the observed and forecasted amplitudes. The skills show no clear tendency for higher skill with stronger predicted NAO-impact amplitude.

Chapter 6:

Summary and Concluding Remarks

Seasonal climate predictions are operational at several meteorological centres, as the European Centre for Medium-Range Weather Forecast (ECMWF), and are used for many applications (Anderson et al. 2003, Palmer et al. 2004, Goddard et al. 2003). In this thesis, the potential seasonal predictability of the atmosphere and its associated impact on surface related quantities is explored for the European-Atlantic winter climate. Two state-of-the-art forecast systems are applied, namely the ECMWF Seasonal Forecast System 2 and the multi-model system developed within the joint European project DEMETER. Both systems consist of coupled atmosphere-ocean general circulation models and are designed to produce forecasts with 5 or 40 ensemble members for system 2, and 9 ensemble members per model for the DEMETER system, respectively.

Four major issues are discussed. First, the mechanics of the RPSS are studied in the context of forecast systems with a small ensemble size. Second, the seasonal forecasts of the European climate are examined for large-scale averages and on the grid-point scale. Third, the probabilistic seasonal prediction of the winter NAO and its impact on near surface temperature is considered. And finally, the seasonal forecasts are evaluated from an end-user perspective. The key findings are summarised below.

- In agreement with earlier studies (e.g. Kumar et al., 2001), it is shown that the standard calculation of the RPSS leads to a negative bias that can be even larger than the expected skill of the forecast system itself. This negative bias is a consequence of the squared measure used to quantify the forecast error in the cumulative probability space. It is particularly large for small ensemble sizes. Two strategies are introduced that circumvent the bias problem of the RPSS. First, it is shown that the $RPSS_{L=1}$, which is based on the absolute rather than the squared difference between forecasted and observed cumulative probability distribution, is unbiased. However, the $RPSS_{L=1}$ is not strictly proper, which means that the probabilities can be changed without impact on the skill score. A second approach, which is based on the quadratic norm, reduces the reference forecast to sub-samples of the same size as the forecast ensemble size. The resulting unbiased and proper skill score is denoted as the debiased ranked probability skill score ($RPSS_D$).

  The new skill scores are employed to find the minimum ensemble size required to predict a given climate signal. By using a hypothetical set up comparable to the ECMWF hindcast system, statistically significant skill scores can be anticipated for signal-to-noise ratios larger than ~0.3 (40 members) and ~0.6 (5 members) respectively.

- First the forecast skill of the 2m mean temperatures of the European-Atlantic domain is examined from a grid-point perspective. For the period 1987-2001 the ECMWF Seasonal Forecast System 2 shows the highest skill scores over the subtropical Atlantic Ocean amounting to 20 %-30 %. Over the European continent, no significant positive skill score is found. The perfect model approach supports these findings and shows that the potential predictability is confined to the North Atlantic Ocean. It reveals a horse-shoe like structure over the North-Atlantic. Additionally, it is demonstrated that the skill scores of the multi-model approach provide higher skill scores than those of the individual models.
- The winter predictability is next examined via the dominant mode of the European winter climate variability, namely the NAO. This major winter climate mode describes about one third of the interannual variability of the upper level flow in the Atlantic European mid-latitudes. Using a perfect model approach for the period 1959-2001, it is shown, that the mean winter NAO index is potentially predictable for lead time 1 month. The prediction benefit is rather small (7 % relative to the model climatology) but statistically significant. A similar conclusion holds for near surface temperature variability related to the NAO. Again the potential benefit is small (8 %) but statistically significant. However, for verification against observations the results are quite different: For the period 1959-2001, the NAO skill score is not statistically significant, while for the period 1987-2001 the skill score is surprisingly large (16 % to 27 % relative to the observed climatology). Furthermore, a weak relation between the strength of the NAO amplitude and the skill score of the NAO is found. This contrasts with ENSO forecasts which have a strong amplitude dependent forecast skill.
- Finally, the seasonal forecasts are examined from an end-user perspective. A systematic analysis reveals that the forecast skill can be improved in a relative sense, looking to spatial and temporal averaged quantities. In the northern Extra-tropics for instance, 6-monthly mean forecasts provide approximately 5 % and 10% higher skill than seasonal and monthly means, respectively. Spatial averaging of the near surface temperature shows an increase in forecast skill only in the tropics, with increased skill scores for an average over approximately 30-140 grid points. This analysis provides the basis for a newly designed climate forecast visualisation, the so-called "Klimagram".

Conclusively, the seasonal predictability of the European-Atlantic domain is small and not comparable to the tropics. The potential benefit for the end user requires careful considerations about the location and season. The forecast skills discussed in this thesis, however, depend on the dynamical model used, and an increase in forecast skill is possible for improved forecast systems. It has been shown that a multi-model provides an alternative approach to increase the forecast skill (see Palmer et al. 2004, Barnston et al. 2003). But also post-processing techniques, such as downscaling methods can be considered to improve applications based on the forecast systems (see Palmer et al. 2004).

Appendix A: End User Application of Seasonal Forecasts

Many user applications require seasonal climate forecasts for particular cities or regions. Ideally such forecasts should contain as many details on the predicted quantity as possible over the entire six month forecast period. The so called "Klimagram", introduced below, has been developed in this thesis to meet such requirements. Seasonal forecasts often reveal limited skill on local station or grid-point scale, in particular over the mid-latitude land areas. A systematic analysis reveals that the forecast skills can be improved in a relative sense, looking to spatial and temporal averaged quantities.

The Klimagram

The "Klimagram" is a compact way to display a probabilistic seasonal forecast (Fig. A1). It shows the evolution of monthly mean forecasts with regard to the model climatology and in a specific geographic area. The abscissa corresponds to the forecast's integration time and the ordinate shows the anomaly of the forecast relative to the mean of the model climate. The probabilistic forecast ensemble itself is denoted by blue notched box-plots divided into quartiles. The black line in the boxes represents the median and the extreme values are indicated by the end of the dashed lines. The model climatology is displayed in the background as grey area. The quartiles are shaded grey with the median as black line. Further, the inner 90 % and the envelope are marked as black solid and black dotted lines, respectively. The notched box-plots also include a significance test between the forecasted ensemble distribution and the model climate distribution evaluated for each forecast month separately. Relevant are the notches of the forecast ensemble and the two dots of the model climate. The notches display a confidence interval of the median. If both notches of the box-plot and both dots of the model climate do not overlap, the median of the two samples differ in a statistical sense at the 95 % confidence level.

In Figure A1 two examples of Klimagrams are shown for monthly mean near surface temperature forecasts. The model climatology is based on the hindcast period 1987-2001 of the corresponding starting month. Panel (a) displays the forecasted temperature values spatially averaged over Western Europe (grid points over land only). The spread of the forecast is close to the one of the model climatology indicating a limited skill of this particular forecast. In panel (b) a Klimagram is shown for a temperature forecast for the equatorial Pacific Nino3.4 region. In this case the ensemble spread is much smaller than the one of the model climatology and the forecasted median is significantly above the model climatology for the entire forecast period, forecasting an El Niño like situation.

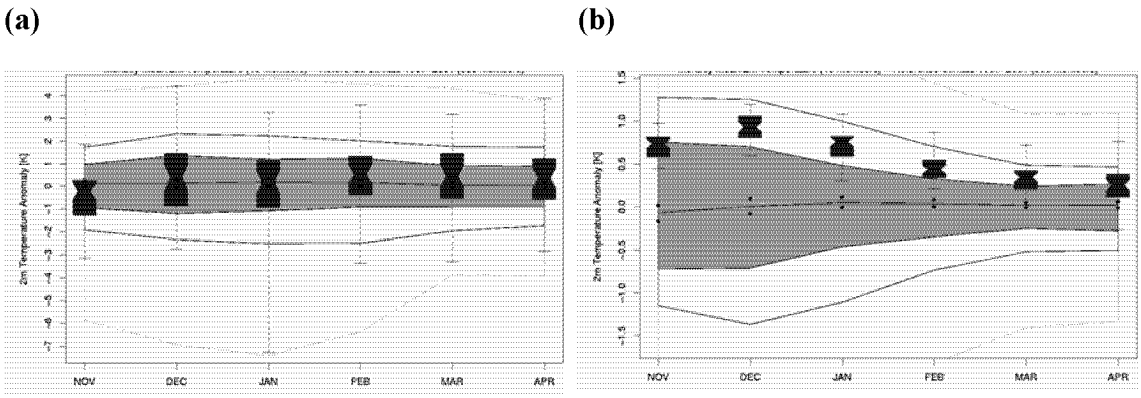


Figure A-1: Monthly mean near surface temperature forecasts displayed with Klimagrams for (a) western Europe (4°W-16°E, 48N-52N) and (b) the Nino3.4 region. Forecasts are from the ECMWF Seasonal Forecast System 2 started on November 1st 2003.

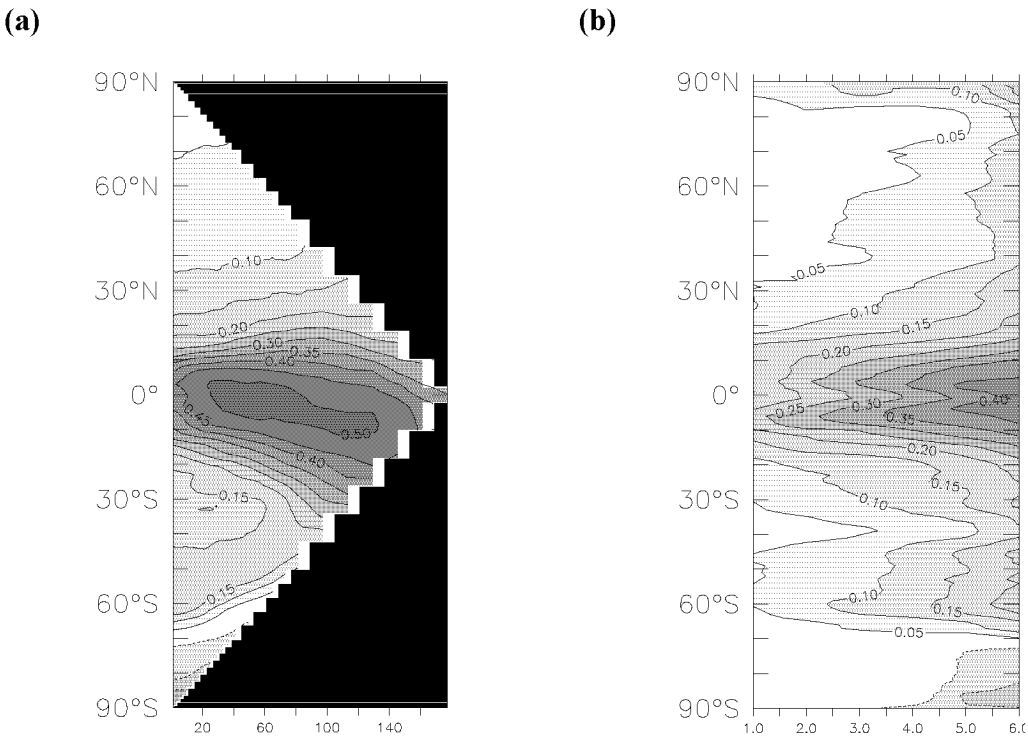


Figure A-2: The zonal averaged $RPSS_D$ as a function of (a) latitude and spatial average applied (units given in grid points) and (b) latitude and temporal average applied (units given in months). The skill scores are evaluated on the basis of the ECMWF Seasonal Forecast System 2 forecasts started at 1st of all months and for the period 1987-2001. For panel (a) 3-month mean forecasts with lead time 1 month are used.

Spatial and Temporal Averaging

The sensitivity of the seasonal forecast skill to spatial and temporal averaging is illustrated in Figure A2. Panel (a) shows the zonal mean of the skill scores ($RPSS_D$) as a function of latitude and the spatial average applied. The skill scores are calculated at each grid point by an average of the forecasted 2m temperatures over a surrounding region of variable size. For example the point (0° latitude, 160 grid points) represents the globally averaged skill score. As verification analysis the ERA40 data set was used. Using larger region averages increase the skill in the tropical region, whereas in the mid latitudes spatial averaging has little effect. The sensitivity of the skill scores to time-averaging is shown in panel (b). The skill scores increase with longer time-averages. This holds for the tropics, where the skill scores are increased from about 15 % to 40 %, but also for the extra-tropics, where the zonal averaged skill scores in the northern and southern hemisphere amount to 10 % and 20 %, respectively. A more detailed analysis is in progress as a joint project and is planed to be submitted at full length.

References

Alves, J. O. S., M. Balmaseda, D. Anderson, and T. Stockdale, (2003): Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *ECMWF Technical Memorandum*, **369:** 1-26.

Ambaum, M. H. B., B. J. Hoskins, and D. B. Stephenson, (2001): Arctic Oscillation or North Atlantic Oscillation. *Journal of Climate*, **14:** 3495-3507.

Ambaum, M. H. B. and B. J. Hoskins, (2002): The NAO troposphere-stratosphere connection. *Journal of Climate*, **15:** 1969-1978.

Anderson, D. L., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitard, F. J. Doblas-Reyes, R. Hagedorn, T. Jung, A. Vidard, A. Troccoli, and T. N. Palmer, (2003): Comparison of the ECMWF Seasonal Forecast Systems 1 and 2, including the relative performance for the 1997/8 El Niño. *ECMWF Technical Memorandum*, **404:** pp93.

Appenzeller, C., J. Schwander, S. Sommer, and T. F. Stocker, (1998): The North Atlantic Oscillation and its imprint on precipitation and ice accumulation in Greenland. *Geophysical Research Letters*, **25:** 1939 - 1942.

Appenzeller, C., T. F. Stocker, and M. Anklin, (1998): North Atlantic oscillation dynamics recorded in Greenland ice cores. *Science*, **282:** 446-449.

Appenzeller, C., T. F. Stocker, and A. Schmittner, (2000): Natural climate variability and climate change in the North-Atlantic European region: chance for surprise. *Integrated Assessment*, **1:** 301-306.

Argyis, J., G. Faust, and M. Haase, (1995): *Die Erforschung des Chaos*. Vieweg, 790 pp.

Arpe, K. and E. Klinker, (1986): Systematic errors of the ECMWF operational forecasting model in mid-latitudes. *Quarterly Journal of the Royal Meteorological Society*, **112:** 181-202.

Bacon, S. and D. J. T. Carter, (1993): A connection between mean wave height and atmospheric pressure gradient in the North Atlantic. *International Journal of Climatology*, **13:** 423-436.

Bader, J. and M. Latif, (2003): The impact of decadal-scale Indian Ocean sea surface temperature anomalies on Sahelian rainfall and the North Atlantic Oscillation. *Geophysical Research Letters*, **30:** 1-4.

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

Barnston, A. G. and R. E. Livezey, (1987): Classification, seasonality, and persistence of low frequency atmospheric circulation patterns. *Monthly Weather Review*, **115:** 1083-1126.

Barnston, A. G., S. J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak, (2003): Multimodel ensembling in seasonal climate forecasting at IRI. *Bulletin of the American Meteorological Society*, **84:** 1783-1796.

Barsugli, J. J. and D. Battisti, (1998): The basic effects of atmosphere-ocean coupling on mid-latitude variability. *J. Atmos. Sci.*, **55:** 477-493.

Battisti, D. S. and A. C. Hirst, (1989): Interannual variability in the tropical atmoaphere/ocean system: Influence of the basic state, ocean geometry and nonlinearity. *J. Atmos. Sci.*, **46:** 1687-1712.

Battisti, D. S., U. S. Bhatt, and M. A. Alexander, (1995): A modelling study of the interannual variability of the North Atlantic Ocean. *Journal of Climate*, **8:** 3067-3083.

Bjerknes, J., (1964): Atlantic air-sea interaction. *Advances in Geophysics*, **10:** 1-82.

——, (1969): Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, **97:** 163-172.

Boulanger, J. P., (2001): Role of non-linear oceanic processes in the response to westerly wind events: New implications for the 1997 El Niño onset. *Geophysical Research Letters*, **28:** 1603-1606.

Bretherton, C., C. Smith, and J. M. Wallace, (1992): An intercomparison of methods for finding coupled patterns in climatic data. *Journal of Climate*, **5:** 541-560.

Bretherton, C. and D. Battisti, (2000): An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophysical Research Letters*, **27:** 767-770.

Buizza, R. and T. N. Palmer, (1998): Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, **126:** 2503-2518.

Buizza, R., T. Pertoliagis, T. N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, (1998): Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quarterly Journal of Royal Meteorological Society*, **124:** 1935-1960.

Cayan, D. R., (1992): Latent and sensible heat flux anomalies over the northern oceans: The connection to monthly atmospheric circulation. *Journal of Climate*, **5:** 354-369.

Coelho, C. A. S., S. Pezzullin, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, (2003): Forecast calibration and combination: a simple Bayesian approach for ENSO. *Journal of Climate*, **(submitted)**.

——, (2003): Skill of coupled seasonal forecasts: A Bayesian assessment of ECMWF ENSO forecasts. *ECMWF Technical Memorandum*, **426:** 1-16.

Cook, E. R., 2003: Multi-Proxy Reconstructions of the North Atlantic Oscillation (NAO): A Critical Review and a New Well-Verified NAO Index Reconstruction Bach to AD 1400. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 63-82.

Czaja, A. and C. Frankignoul, (1999): Influence of the North Atlantic SST on the atmospheric circulation. *Geophysical Research Letters*, **26:** 2969-2972.

Czaja, A. and J. Marshall, (2001): Observations of atmosphere-ocean coupling in the North Atlantic. *Quarterly Journal of the Royal Meteorological Society*, **127:** 1893-1916.

Czaja, A., P. van der Vaart, and J. Marshall, (2002): A diagnostic study of the role of remote forcing in Tropical Atlantic Variability. *Journal of Climate*, **15:** 3280-3290.

Czaja, A., A. W. Robertson, and T. Huck, 2003: The role of Atlantic ocean-atmosphere coupling in affecting North Atlantic Oscillation variability. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 147-172.

Davis, R., (1976): Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific ocean. *Journal of Physical Oceanography*, **8:** 249-266.

——, (1978): Predictability of sea-level pressure anomalies over the North Pacific oceans. *Journal of Physical Oceanography*, **8:** 233-246.

Delecluse, P. and G. Madec, 1999: Ocean modelling and the role of the ocean in the climate system. *Modelling the Earths Climate and its Variability*, W. R. Holland, Ed., Elsevier Science, 237-313.

Delworth, T., S. Manabe, and R. J. Stouffer, (1993): Interdecadal variations in the thermohaline circulation in a coupled ocean-atmosphere model. *Journal of Climate*, **6:** 1993-2010.

Delworth, T. and R. T. Greatbach, (2000): Multidecadal thermohaline circulation variability driven by atmospheric surface flux forcing. *Journal of Climate*, **13:** 1481-1495.

Deque, M., (2001): Seasonal predictability of tropical rainfall: probabilistic formulation and validation. *Tellus*, **53A:** 500-512.

Dequé, M., (1997): Ensemble size for numerical seasonal forecasts. *Tellus A*, **49:** 74-86.

Deser, C. and M. L. Blackmon, (1993): Surface climate variations over the North Atlantic during winter: 1900-1989. *Journal of Climate*, **10:** 393-408.

Deser, C., (2000): On the teleconnectivity of the Arctic Oscillation. *Geophysical Research Letters*, **27:** 779-782.

Diaz, H. F. and V. Markgraf, (2000): *El Niño and the Southern Oscillation: Multiscale variability and global and regional impacts.* Cambridge University Press.

Dickson, R., T. J. Osborn, J. Hurrell, J. Meincke, B. Blindheim, Vigne, T., and W. Maslowski, (2000): The Arctic Ocean response to the North Atlantic Oscillation. *Journal of Climate*, **13:** 2671-2696.

Doblas-Reyes, F. J., M. Deque, and J. P. Piedelievre, (2000): Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society*, **126:** 2069-2088.

Doblas-Reyes, F. J. and D. B. Stephenson, (2003) Multi-model seasonal hindcasts of the North Atlantic Oscillation. *Clim Dyn*, **21**, 501-514.

F. J. Doblas-Reyes, R. Hagedorn and T. Palmer (2004): The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and Combination. *Tellus* **(submitted).**

Dommenget, D. and M. Latif, (2002): A cautionary note on the interpretation of EOFs. *Journal of Climate*, **11:** 216-225.

Drévillion, M., L. Terray, P. Rogel, and C. Cassou, (2001): Mid-latitude Atlantic SST influence on European winter climate variability in the NCEP reanalysis. *Climate Dynamics*, **18:** 331-344.

Drinkwater, K. F., A. Belgrano, A. Borja, A. Conversi, M. Edwards, C. H. Greene, G. Ottersen, A. J. Pershing, and H. Walker, 2003: The response of maritime ecosystems to climate variability associated with the North Atlantic Oscillation. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 211-234.

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

Evans, R. E., M. S. J. Harrison, R. J. Graham and K. R. Mylne, (2000) Joint medium-range ensembles from The Met Office and ECMWF systems. *Mon Weather Rev*, **128**, 3104-3127.

Fedorov, A. V., (2002): The response of the coupled tropic ocean-atmosphere to westerly wind burst. *Quarterly Journal of the Royal Meteorological Society*, **128:** 1-23.

Fedorov, A. V., S. L. Harper, S. G. Philander, B. Winter, and A. Wittenberg, (2003): How predictable is El Niño? *Bulletin of the American Meteorological Society*, **84:** 911-919.

Feldstein, S. B., (2000): The timescale, power spectrum, and climate noise properties of teleconnection patterns. *Journal of Climate*, **13:** 4430-4440.

Folland, C. K. and a. co-authors, 2001: Observed climate variability and change. *Climate Change 2001, The scientific basis*, J. Houghton, T., Y. Ding, D. J. Griggs, M. Nogouer, P. J. van der Linden, and D. Xiaosu, Eds., Cambridge University Press, 99-181.

Fraedrich, K. and K. Müller, (1992): Climate anomalies in Europe associated with ENSO extremes. *International Journal of Climatology*, **12:** 25-31.

Fraedrich, K., (1994): ENSO impact on Europe. *Tellus*, **46A:** 541-552.

Frankignoul, C. and K. Hasselmann, (1977): Stochastic climate models. Part II. Application to sea surface temperature variability and thermocline variability. *Tellus*, **29:** 284-305.

Frankignoul, C., (1985): Sea surface temperature anomalies, planetary waves and air-sea feedbacks in the middle latitudes. *Reviews of Geophysics*, **23:** 357-390.

Frankignoul, C. and A. Czaja, (2002): Observed impact of Atlantic SST on the North Atlantic Oscillation. *Journal of Climate*, **15:** 606-623.

Gillett, N. P., H.-F. Graf, and T. J. Osborn, 2003: Climate Change and the North Atlantic Oscillation. *The North Atlantic Oscillation: Climate Significance and Environmental Impact*, J. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck, Eds., AGU, 193-210.

Goddard, L., A. G. Barnston, and S. J. Mason, (2003): Evaluation of the IRI's Net Assessment seasonal climate forecasts. *Bulletin of the American Meteorological Society*, **84:** 17611781.

Gordon, C., C. Cooper, A. Senior, H. Banks, J. M. Gregory, T. C. Johns, F. B. Mitchell, and R. A. Wood, (2000): The simulation of SST, sea ice extends and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16:** 147-168.

Graham, R. J., A. D. L. Evans, K. R. Myline, M. S. J. Harrison, and K. B. Robertson (2000) An assessment of seasonal predictability using atmospheric general circulation models. *Q J R Meteorol Soc* , **126,** 211-2240.

Grazzini, F., L. Ferranti, F. Lalaurette, and F. Vitard, (2003): The exceptional warm anomalies of summer 2003. *ECMWF Newsletter*, **99:** 2-8.

Grazzini, F. and P. Viterbo, (2003): Record-breaking warm sea surface temperature of the Mediterranean Sea. *ECMWF Newsletter*, **98:** 30-31.

Greatbach, R. T., (2000): The North Atlantic Oscillation. *Stochastic and Environmental Risk Assessment*, **14:** 213-242.

Gregory, D., J. J. Morcette, C. Jacob, A. C. M. Beljaars, and T. Stockdale, (2000): Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecast System. *Quarterly Journal of the Royal Meteorological Society*, **126:** 1685-1710.

Griffies, S. M. and E. Tziperman, (1995): A linear thermohaline oscillator driven by staochastic atmospheric forcing. *Journal of Climate*, **8:** 2440-2453.

Grötzner, A., M. Latif, and T. P. Barnett, (1998): A decadal climate cycle in the North Atlantic ocean as simulated by ECHO coupled GCM. *Journal of Climate*, **11:** 831-847.

Grötzner, A., M. Latif, A. Timmermann, and R. Voss, (1999): Interannual to decadal predictability in coupled ocean-atmosphere general circulation model. *Journal of Climate*, **12:** 2607-2624.

Guckenheimer, J. and P. Holmes, (1983): *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. 459 ed. Vol. 42, *Applied Mathematical Sciences*, Springer.

Hagedorn, R. , F. J. Doblas-Reyes and T. Palmer (2004): The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concepts. *Tellus* **(submitted).**

Halliwell, G. R. and D. A. Mayer, (1996): Frequency response properties of forced climate SST anomaly variability in the North Atlantic. *Journal of Climate*, **9:** 3575-3587.

Hartmann, D. L., J. M. Wallace, V. Limpasuvan, D. W. J. Thompson, and J. R. Holton, (2000): Can ozone depletion and global warming interact to produce rapid climate change? *Proc. Nat. Acad. Sci.*, **97:** 1412-1417.

Hasselmann, K., (1976): Stochastic climate models, Part I: Theory. *Tellus*, **28:** 473-485.

Hersbach, H., (2000): Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15:** 559-570.

Hoerling, M. P., J. Hurrell, and T. Y. Xu, (2001): Tropical origins for recent North Atlantic climate change. *Science*, **292:** 90-92.

Hurrell, J., (1995): Decadal Trends in the North Atlantic Oscillation Regional Temperatures and Precipitation. *Science*, **269:** 676-679.

Hurrell, J., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 1-36.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 2003: Pressure-Based Measures of the North Atlantic Oscillation (NAO): a comparison and a Assessment of Changes in the Strength of the NAO and its Influence on surface Climate Parameters. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 51-62.

Jones, R. H., (1975): Estimating the variance of time averages. *Journal of Applied Meteorology*, **14:** 159-163.

Kovats, S., T. Wolf and B. Menne, (2004): Heat waves of August 2003 in Europe: provisional estimates of the impact on mortality. *Eurosurveillance Weekly*, **8 (11)**.

Kharin, V. V. and F. Zwiers, (2002): Improved seasonal probability forecasts. *Journal of Climate*, **16:** 1684-1701.

Krishnamurti, T. N., C. M. Kishtawai, L. T., D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, (1999): Improved weather and seasonal climate forecast from multimodel superensemble. *Science*, **285:** 1548-1550.

Kumar, A. and M. P. Hoerling, (1995): Prospects and limitations of atmospheric GCM climate predictions. *Bulletin of the American Meteorological Society*, **76:** 335-345.

Kumar, A., A. G. Barnston, and M. P. Hoerling, (2001): Seasonal predictions, probabilistic verifications and ensemble size. *Journal of Climate*, **14:** 1671-1676.

Kushnir, Y., V. J. Cardone, J. G. Greenwood, and M. Cane, (1997): On the recent increase in North Atlantic wave height. *Journal of Climate*, **10:** 2107-2113.

Lamb, P. J. and R. A. Peppler, (1987): The North Atlantic Oscillation: Concept and an application. *Bulletin of the American Meteorological Society*, **68:** 1218-1225.

Landsea, W. C. and J. A. Knaff, (2000): How much skill was there in Forecasting the very strong 1997-98 El Niño? *Bulletin of the American Meteorological Society*, **81:** 2107-2119.

Latif, M. and T. P. Barnett, (1994): Causes of decadal climate variability over the North Pacific and North America. *Science*, **266:** 634-637.

Latif, M., T. Stockdale, J. Wolff, G. Burgers, E. Maier-Remer, M. Junge, K. Arpe, and L. Bengtsson, (1994): Climatology and variability in the ECHO coupled GCM. *Tellus*, **46A:** 351-366.

Latif, M. and T. P. Barnett, (1996): Decadal climate variability over the North Pacific and North America: Dynamics and Predictability. *Journal of Climate*, **9:** 2407-2423.

Latif, M., (1998): Dynamics of inter-decadal variability in coupled ocean-atmosphere models. *Journal of Climate*, **11:** 602-624.

Latif, M., D. Anderson, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider, (1998): A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, **103:** 14375-14393.

Leith, C. E., (1973): The standard error of time - average estimates of climatic means. *Journal of Applied Meteorology*, **12:** 1066-1069.

——, (1974): Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, **102:** 409-418.

Lorenz, E. N., (1963): Deterministic nonperiodic flow. *Journal of Atmospheric Science*, **20:** 130-141.

Madden, R. A., (1976): Estimates of the natural variability of time-averaged sea-level pressure. *Monthly Weather Review*, **104:** 942-952.

Madden, R. A. and D. J. Shea, (1978): Estimates of the natural variability of time-averaged temperature over the United States. *Monthly Weather Review*, **106:** 1695-1703.

Madden, R. A., (1981): A quantitative approach to long-range prediction. *Journal of Geophysical Research*, **86:** 9817-9825.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, 1997: OPA Version 8.0 Ocean General Circulation Model Reference Manual11, 200 pp.

——, 1998: OPA Version 8.1 Ocean General Circulation Model Reference Manual11, 91 pp.

Marsagli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza (2001): A strategy for high resolution ensemble prediction II: Limited-area experiments in four Alpine flood events. *Quarterly Journal of Royal Meteorological Society*, **127:** 2095-2115.

Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravan, and M. Visbeck, (2001): North Atlantic Climate Variability: Phenomena, Impacts and Mechanisms. *International Journal of Climatology*, **21:** 1863-1898.

Marsland, S. H., H. Haak, J. H. Jungclaus, M. Latif, and F. Röske, 2002: The Max-Plack-Institute global sea/ice model with orthogonal curvilinear coordinates. *Ocean Modelling*, In Press.

Mason, I., (1982): A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30:** 291-303.

Mason, S. J. and G. M. Mimmack, (2002): Comparison of some statistical methods of probabilistic forecasting of ENSO. *Journal of Climate*, **15:** 8-29.

Mason, S. J., (2004): On using climatology as a reference strategy in the Brier and ranked probability skill scores. *Monthly Weather Review*: (submitted).

McPhadden, M. J. and X. Yu, (1999): Equatorial waves and the 1997798 El Niño. *Geophysical Research Letters*, **26:** 2961-2964.

Merkel, U. and M. Latif, (2002): A high resolution AGCM study of the El Niño impact on the North Atlantic/ European sector. *Geophysical Research Letters*, **29:** 1-4.

Metzger, S., K. Fraedrich, and M. Latif, (2004): Combining ENSO forecasts: A feasible study. *Monthly Weather Review*, **132:** 456-472.

Molteni, F. and R. Buizza, (1999): Validation of the ECMWF ensemble prediction system using empirical orthogonal functions. *Monthly Weather Review*, **127:** 2346-2358.

Morse, A. P., M. Hoshen, F. J. Doblas-Reyes, and M. C. Thompson, (2003): Towards forecasting epidemics in Africa - the use of seasonal forecasting. *CLIVAR Exchanges*, **8:** 50-52.

Müller, W. A., C. Appenzeller, F. J. Doblas-Reyes and M. A. Liniger, (2004): A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal of Climate*, **(submitted)**.

Müller, W. A., C. Appenzeller, and C. Schär, (2004): Probabilistic seasonal prediction of the winter North Atlantic Oscillation and its impact on near surface temperature. *Climate Dynamics*: (**accepted**).

Murphy, A. H., (1969): On the ranked probability skill score. *Journal of Applied Meteorology*, **8:** 988-989.

——, (1971): A note on the ranked probability skills score. *Journal of Applied Meteorology*, **10:** 155-156.

——, (1973): A new partition of the probability score. *Journal of Applied Meteorology*, **12:** 595-600.

Murphy, A.H. and R. W Daan, (1985): Forecast Verification. Probability, Statistics and Decision Making in the Atmospheric Science, Murphy, A.H. and Katz, R. W., Westview Press, 379-437.

Mysterud, A., N. C. Stenseth, N. G. Yoccoz, R. Langvatn, and G. Ottersen, 2003: The response of terrestrial ecosystems to climate variability associated with the North Atlantic Oscillation. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 235-262.

Neelin, J. D., M. Latif, and F. F. Jin, (1994): Dynamics of ocean-atmosphere models: The troipical problem. *Ann. Rev. Fluid. Mech.*, **26:** 617-659.

Neelin, J. D., D. S. Battisti, A. C. Hirst, F. F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, (1998): ENSO theory. *Journal of Geophysical Research*, **103:** 14261-14290.

Nichollis, N., (2001): The insignificance of significance testing. *Bulletin of the American Meteorological Society*, **81:** 981-986.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, (1982): Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, **109:** 2080-2092.

Palmer, T. N. and Z. Sun, (1985): A modelling and observational study of the relationship between sea surface temperature anomalies in the North-West Atlantic and the atmospheric general circulation. *Quarterly Journal of the Royal Meteorological Society*, **111:** 947-975.

Palmer, T. N., (1993): A nonlinear dynamical perspective on climate change. *Journal of Climate*, **48:** 314-327.

Palmer, T. N. and D. L. T. Anderson, (1994): The prospect of seasonal forecasting - a review paper. *Quarterly Journal of the Royal Meteorological Society*, **120:** 755-793.

Palmer, T. N., C. Brankovic, and D. S. Richardson, (2000): A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quarterly Journal of Royal Meteorological Society*, **126:** 2013-2033.

Palmer, T. N. and J. Shukla, (2000): Editorial to DSP/PROVOST special issue. *Quarterly Journal of the Royal Meteorological Society*, **126:** 1989-1990.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantalaube, M. Davey, P. Delecluse, M. Deque, E. Diez, F. J. Doblas-Reyes, H. Fedderson, R. Graham, S. Gualdi, J. F. Gueremy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J. M. Terres, and M. C. Thomson, (2004): Development of a European Multi-Model ensemble system for seasonal to interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, **85,** 853-872.

Pavan, V. and F. J. Doblas-Reyes, (2000): Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climate Dynamics*, **16:** 611-625.

Peng, S., W. A. Robinson, and S. Li, (2003): Mechanism for the NAO response to the North Atlantic SST tripole. *Journal of Climate*, **16:** 1987-2004.

Perlwitz, J. and H.-F. Graf, (1995): The statistical connection between tropospheric and stratospheric circulation of the Northern Hemisphere winter. *Journal of Climate*, **8:** 2281-2295.

Pope, V. D., M. L. Gallani, P. R. Rowtree, and R. A. Stratton, (2000): The impact of new physical parameterization in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, **16:** 123-146.

Rajagopalan, B., U. Lall, and S. E. Zebiak, (2002): Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Monthly Weather Review*, **130:** 1792-1811.

Röckner, E., 1996: The Atmospheric General Circulation Model ECHAM-4: Model description and simulation of the present-day circulation. Technical Report 218, 90 pp.

Rodwell, M. J., D. P. Rowell, and C. K. Folland, (1999): Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, **398:** 320-323.

Rodwell, M. J. and C. K. Folland, (2002): Atlantic air-sea interaction and seasonal predictability. *Quarterly Journal of the Royal Meteorological Society*, **128:** 1413-1443.

Rogers, J. C., (1984): The association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere. *Monthly Weather Review*, **112:** 1999-2015.

——, (1990): Patterns of low-frequency monthly sea level pressure variability (1899-1986) and associated wave cyclone frequencies. *Journal of Climate*, **3:** 1364-1379.

——, (1997): North Atlantic storm track variability and its association to the North Atlantic Oscillation and climate variability of northern Europe. *Journal of Climate*, **10:** 16-35-1647.

Schär, C., T. D. Davies, C. Frei, H. Wanner, M. Widmann, M. Wild, and H. C. Davies, 1998: Current Alpine Climate. *A View from the Alps: Regional Perspectives on Climate Change*, P. Cebon, U. Dahinden, H. C. Davies, D. Imboden, and C. Jaeger, Eds., MIT Press, 21-72.

Schär, C., D. Lüthi, U. Beyerle, and E. Heise, (1999): The soil-precipitation feedback: A process study with a regional climate model. *Geophysical Research Letters*, **23:** 669-672.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, (2004): The role of increasing temperature variability in European summer heatwaves. *Nature*, **427:** 332-336.

Scherrer, S., C. Appenzeller, P. Eckert, and D. Cattani, (2004): Analysis of the spread-skill relation using the ECMWF Ensemble Prediction System over Europe. Weather and Forecasting, **(submitted)**.

Schmidli, J., C. Schmutz, C. Frei, H. Wanner, and C. Schär, (2002): Mesoscale precipitation variability in the region of the European Alps during the 20th century. *International Journal of Climatology*, **22:** 1049-1074.

Schopf, P. S. and M. J. Suarez, (1988): Vacillations in a coupled ocean-atmosphere model. *Journal of Atmospheric Science*, **45:** 549-566.

Serreze, M. C., F. Carse, R. G. Barry, and J. C. Rogers, (1997): Icelandic low activity: climatological features, linkages with the NAO, and relationship with recent changes in the Northern Hemisphere circulation. *Journal of Climate*, **10:** 453-464.

Shukla, J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, L. Marx, T. N. Palmer, D. Paolino, J. Ploshay, S. Schubert, D. M. Straus, M. Suarez, and J. Tribba, (2000): Dynamical seasonal prediction. *Bulletin of the American Meteorological Society*, **81:** 2593-2609.

Smith, N., J. Blomley, and G. Meyers, (1991): A univariate statistical interpolation scheme for subsurface thermal analysis in the tropical oceans. *Progress of Oceanography*, **28:** 219-256.

Stephenson, D. B. and I. T. Jolliffe, (2003): *Forecast Verification: A Practitioner`s Guide in Atmospheric Science*. Wiley, 240 pp.

Stern, W. and k. Miyakoda, (1995): Feasibility of seasonal forecast inferred from multiple GCM simulations. *Journal of Climate*, **8:** 1071-1085.

Straile, D., D. M. Livingston, G. A. Weyhenmeyer, and D. G. George, 2003: The response of freshwater ecosystems to climate variability associated with the North Atlantic Oscillation. *The North Atlantic Oscillation*, Y. K. J. W. Hurrell, G. Ottersen and M. Visbeck, Ed., AGU, 263-279.

Straus, D. M. and J. Shukla, (2000): Distinguishing between the SST-forced variability and internal variability in mid-latitudes: Analysis of observations and GCM simulations. *Quarterly Journal of the Royal Meteorological Society*, **126:** 2323-2350.

Suarez, M. J. and P. S. Schopf, (1988): A delayed action oscillator for ENSO. *Journal of Atmospheric Science*, **45:** 3283-3287.

Swets, J. A., (1973): The relative operating characteristic in psychology. *Science*, **182:** 990-1000.

Thompson, D. W. J. and J. M. Wallace, (1998): The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, **25:** 1297-1300.

——, (2000): Annular modes in the extra-tropical circulation, Part I: Month-to-month variability. *Journal of Climate*, **13:** 1018-1036.

Tibaldi, S. and F. Molteni, (1990): On the operational predictability of blocking. *Tellus*, **42A:** 343-365.

Tibaldi, S., T. N. Palmer, C. Brankovic, and U. Cubasch, (1990): Extended-range predictions with ECMWF models: influence of horizontal resolution on systematic error and forecast skill. *Quarterly Journal of the Royal Meteorological Society*, **116:** 835-866.

Tibaldi, S., E. Tosi, A. Navarra, and L. Pedulli, (1994): Northern and southern hemisphere seasonal variability of blocking frequency and predictability. *Monthly Weather Review*, **122:** 197-200.

Timmermann, A., M. Latif, R. Voss, and A. Grötzner, (1998): Northern hemispheric interdecadal variability: a coupled air-sea mode. *Journal of Climate*, **11**.

Toth, Z. and E. Kalnay, (1993): Ensemble forecasting at NMC. The generation of perturbations. *Bulletin of American Meteorological Society*, **74:** 2317-2330.

Tracton, M. S. and E. Kalnay, (1993): Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Weather and Forecasting*, **8:** 379-398.

Tracton, M. S. and Z. Toth, (1993): Ensemble Forecast at NMC. Operational implementation. *Weather and Forecasting*, **8:** 379-398.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropolewski, (1998): Progress during TOGA in understanding and modelling global teleconnections associated with tropical sea surface temperature. *Journal of Geophysical Research*, **103:** 14291-14324.

Unger, D. A., 1985: A method to estimate the continuous ranked probability score. *Ninth Conf. on Probability and Statistics in Atmospheric Science*, American Meteorological Society, 206-213.

van Loon, H. and J. Rogers, (1978) The seesaw in winter temperatures between Greenland and Norhtern Europe, Part I: General description. *Mon Weather Rev*, **106,** 296-310.

van Oldenborgh, G. J., M. Balmaseda, L. Ferranti, T. Stockdale, and D. L. T. Anderson, (2003): Did the ECMWF seasonal forecast model outperform a statistical model over 15 years. *ECMWF Technical Memorandum*, **418:** 1-32.

van Oldenborgh, G. J. and G. Burgers, (2004): The effects of El Niño on precipitation and temperature, an update. *Bulletin of the American Meteorological Society*, **submitted**.

Vialard, J., C. Menkes, J. P. Boulanger, P. Delecluse, E. Guilyardi, M. J. McPhadden, and G. Madec, (2001): A model study of the oceanic mechanism affecting equatorial Pacific sea surface temperature during the 1997/98 El Niño. *Journal of Climate*, **13:** 1814-1830.

Vidale, P. L., D. Lüthi, C. Frei, S. I. Seneviratne, and C. Schär, (2003): Predictability and uncertainty in a regional climate model. *Journal of Geophysical Research*, **108**.

Vitart, F., D. L. T. Anderson, L. Ferranti, and M. Balmaseda, (2003): Westerly wind events and the 1997/98 El Nino event in the ECMWF Seasonal Forecast System: A case study. *Journal of Climate*, **16:** 3153-3170.

Walker, G. T., (1924): Correlation in seasonal variations of weather IX. A further study of world weather. *Mem. India Meteorol. Dep.*, **24:** 275-332.

Walker, G. T. and E. W. Bliss, (1930): World Weather IV. *Mem. Roy. Meteo. Soc.*, **3:** 81-95.

——, (1932): World Weather V. *Mem. Roy. Meteo. Soc.*, **4:** 53-84.

Wallace, J. M., (2000): The North Atlantic Oscillation/Annular Mode: Two paradigms - one phenomenon. *Quarterly Journal of the Royal Meteorological Society*, **126:** 791-805.

Wallace, J. M. and D. W. J. Thompson, (2002): The Pacific center of action of the Northern Hemispheric Annular Mode: Real or artefact? *Journal of Climate*, **15:** 1987-1991.

Ward, M. N. and C. K. Folland, (1991): Prediction of seasonal rainfall in the North Northeste of Brazil using eigenvectors of sea-surface temperature. *International journal of Climatology*, **11:** 711-743.

Widmann, M., C. S. Bretherton, and E. P. Salathe, (2003): Statistical precipitation downscaling over the North-western United States using numerically simulated precipitation as a predictor. *Journal of Climate*, **16:** 799-816.

Wilks, D. S., (1995): *Statistical Methods in the Atmospheric Sciences*. Vol. 59, *International Geophysics Series*, Academic Press, 467 pp.

Wolff, J., E. Maier-Reimer, and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model. Technical Report No.13.

Zwiers, F. and V. V. Kharin, (1998): Changes in the extremes of the climate simulated by the CCC GCM2 under CO2 doubling. *Journal of Climate*, **11:** 7295-7315.

Acknowledgements

I would like to thank the following persons for their engagement and endeavours and that with their help this thesis was made possible at all:

Andre Walser, Christof Appenzeller, Christoph Schär, Francesco J. Doblas-Reyes, Heike Kunz, Mark A. Liniger, Mathias Rotach, Mojib Latif, Sabine Kümmerle and Simon Scherrer.

Veröffentlichung

MeteoSchweiz

Veröffentlichungen der MeteoSchweiz

Kürzlich erschienen:

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

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, 40 Fr.

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., 100 Fr.

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

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

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

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

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

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, 36 Fr.

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

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

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

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

55 Wolfensberger, H: 1994, Chronik der Totalisatoren, Handbuch zu den Niederschlags-Totalisatoren, 390pp, 78 Fr.

54 Fankhauser, G A: 1993, Einfluss der Witterung auf den Ertrag und die Qualität von Zuckerrübenkulturen, 116pp, 36 Fr.

53 de Montmollin A. : 1993, Comparaisons de différentes méthodes de calcul de la température journalière dans leurs influences sur les longues séries d'observations, 144pp, 41 Fr.

Frühere *Veröffentlichungen* und *Arbeitsberichte* finden sich unter
www.meteoschweiz.ch

Veröffentlichung

MeteoSchweiz

Arbeitsberichte der MeteoSchweiz

Kürzlich erschienen:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

192 Maurer H: 1997, Frostprognose in der Schweiz: neue Methode mit automatischen Stationen, 38pp, 16 Fr.

191 Schönbächler M: 1996, Objektive Kontrolle der Textprognose SMA OPKO, 31pp, 14 Fr.

190 Brändli J: 1996, Statistische Auswertungen von täglichen und monatlichen Verdunstungswerten an 22 Standorten der Schweiz, 52pp, 19 Fr.

189 Schneiter D: 1994, SMI contribution to ETEX project in 1994, 24 Fr.

188 Fröhlich C: 1996, Internationaler Pyrheliometervergleich Comparison IPC VIII 25 September - 13 October 1995 Results and Symposium, 35 Fr.

187 Calame F: 1996, Evolution de la température de l'air et de la phénologie d'espèces végétales entre 1952 et 1992 dans la région genevoise et sur le Plateau Suisse, 19pp, 11 Fr.

186 Spinedi F., et al.: 1995, Le alluvioni del 1993 sul versante subalpino, 42pp, 20 Fr.

185 Held E: 1995, Radarmessung im Niederschlag und der Einfluss der Orographie, 98pp, 33 Fr.

Frühere *Veröffentlichungen* und *Arbeitsberichte* finden sich unter
www.meteoschweiz.ch