The role of boundary and initial conditions for dynamical seasonal predictability

T. J. Reichler, J. O. Roads

The role of boundary and initial conditions for dynamical seasonal predictability

T. J. Reichler and J. O. Roads

Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0224, USA

Abstract. The importance of initial state and boundary forcing for atmospheric predictability is explored on global to regional spatial scales and on daily to seasonal time scales. A general circulation model is used to conduct predictability experiments with different combinations of initial and boundary conditions. The experiments are verified under perfect model assumptions as well as against observational data. From initial conditions alone, there is significant instantaneous forecast skill out to 2 months. Different initial conditions show different predictability using the same kind of boundary forcing. Even on seasonal time scales, using observed atmospheric initial conditions leads to a substantial increase in overall skill, especially during periods with weak tropical forcing. The impact of boundary forcing on predictability is detectable after 10 days and leads to measurable instantaneous forecast skill at very long lead times. Over the Northern Hemisphere, it takes roughly 4 weeks for boundary conditions to reach the same effect on predictability as initial conditions. During events with strong tropical forcing, these time scales are somewhat shorter. Over the Southern Hemisphere, there is a strongly enhanced influence of initial conditions during summer. We conclude that the long term memory of initial conditions is important for seasonal forecasting.

1   Introduction

Atmospheric flow is sensitively dependent on initial conditions. Lorenz (1969) showed how small scale errors grow progressively and ultimately affect the largest scales. Such small errors are unavoidable because of inaccuracies in the specification of initial conditions, boundary conditions, and various model approximations. The practical implication has been that daily weather events become unpredictable after 2–3 weeks.

Correspondence to: T. J. Reichler (reichler@ucsd.edu)

However, seasonal forecasts are skillful (e.g. Roads et al., 2001; Kanamitsu et al., 2002). The physical basis for making forecasts on such long time scales rests primarily on the sensitivity of the atmosphere to anomalous structures at its lower boundaries, such as sea surface temperatures (SSTs), sea ice, soil moisture, or snow. Fluxes across these boundaries can excite large scale forced modes of variability in the atmosphere. The conditions at these boundaries typically evolve on a much slower time scale than daily weather events. Atmospheric predictability is therefore prolonged to the extent that the future evolution of these boundary conditions can be predicted.

It is well established that on interannual timescales, El-Niño Southern Oscillation (ENSO) induced heating anomalies over the tropical Pacific are the dominant mechanism for predictability over both the tropics and the extratropics. Over the tropics, the response to ENSO forcing is strong and the internal variability is low, leading to a detectable shift in the atmosphere's mean state during ENSO (Kumar and Hoerling, 1998). ENSO can also significantly impact the planetary-scale circulation over the extratropics (e.g. Lau and Nath, 1994). In fact, it has been found that ENSO events are the single most important source for extratropical boundary forced predictability (e.g. Kumar and Hoerling, 1995) on seasonal time scales.

Besides the dominant ENSO signal, there may be other, less regularly occurring and weaker sources of boundary forced predictability. For example, it is reasonable to assume that SST anomalies over the tropical Indian and Atlantic oceans have similar effects as Pacific SST anomalies. Although the SST variability between the three ocean basins is to some degree dependent, these basins also exhibit some variability on their own (e.g. Landman and Mason, 1999). There may also be some influences from midlatitude SST anomalies on the atmosphere, although the predominant opinion is that on time scales shorter than a decade the influences are small (e.g. Robinson, 2000). Low-frequency variations in the land surface boundary conditions like soil moisture, snow cover and soil temperature over both the tropical

and extratropical continents can also contribute to variations in predictability (e.g. Fennessy and Shukla, 1999; Hong and Pan, 2000; Dirmeyer, 2000).

If land or ocean models are coupled to GCMs, then their initial conditions also impact predictability. Because of the much slower boundary time scales, these initial states will influence the forecast during much of the integration. Anderson and Ploshay (2000), for example, found a significant sensitivity in predictability by including observed land surface initial conditions in a forecast.

Even though there is a general agreement that boundary forcing is the primary source of predictability at longer time scales, there is reason to believe that atmospheric initial conditions also influence the evolution of the atmosphere on time scales longer than just a few days. First, it is quite likely that predictability is variable, and that it is also a function of initial conditions. For some initial states, predictability may extend well beyond the mean predictability limit. Other states may be virtually unpredictable, for example those, which are close to regime shifts (Palmer, 1993), but how to identify a priori those predictions with potential extended range skill is still unknown. Second, Shukla (1981) hypothesized that fast growing synoptic instabilities and slowly varying planetary waves might possess different predictability characteristics. There is in fact some evidence from modeling studies for an extended memory of the larger scales (e.g. Roads, 1987). There are probably also regional differences in the importance for initial conditions. Shukla (1998), for example, found that in the tropics the link of the atmosphere to the ocean, and thus the influence of boundary conditions, is too strong in order for initial conditions to have much importance. However, it remains to be seen whether the initial condition effect is limited to the extratropics.

Tribbia and Baumhefner (1988) showed that the effect of initial conditions on daily unaveraged skill in a GCM lasts for roughly 2 months, which they interpreted as an indicator for certain long term predictable components of the atmosphere. They also found little skill in a one month mean forecast with climatological boundary forcing. Barsugli et al. (1999) found that effects of strong tropical boundary forcing appear 5–10 days after initialization. Recently, several large experiments, e.g. "dynamical seasonal prediction" (DSP) (Shukla et al., 2000), and "prediction of climate variations on seasonal to interannual time-scales" (PROVOST) (Palmer et al., 2000), investigated seasonal predictability, which included the effect of initial conditions. Chang et al. (2000), for example, found that initial conditions are the major factor leading to skillful extratropical forecasts at lead times of less than a month, while boundary forcing is the dominant factor at a seasonal time-scale.

Here, we again attempt to understand the relative role of initial and boundary conditions with a series of idealized GCM experiments which are internally consistent with respect to model and data. The experiments are forced with different combinations of initial and boundary conditions to diagnose the effects of both on the forecast. The main difference between this study and previous studies is that the ear-

lier studies examined either daily forecast skill on relatively short time scales, or simply the seasonally averaged skill. In the present study, we attempt to bridge the gap between short term weather forecasting and long term climate prediction by examining the daily breakdown of instantaneous predictability from day one out to a season to derive the various time scales of interest. The results are then compared to time averaged skill for different lead times.

We mostly apply the perfect model approach (Buizza, 1997; Anderson et al., 1999) in this paper. That is, one forecast is verified against another forecast with the same model, and errors concerning uncertainties in the model formulation are eliminated. This allows us to better focus on the basic questions of this paper and to determine the theoretical upper predictability limit for this model. To further examine the influence of model errors on actual forecast skill, we also verify our experiments against reanalysis.

Forecast skill on longer time scales is actually quite variable, especially in the extratropics. Boundary forcing exerts only a weak constraint on the flow, and predictable signals from the boundary forcing and the initial state are largely dominated by noise, which are the unpredictable components of internal variability. In order to forecast on longer time scales, predictable components need to be separated from the noise. The solution is a statistical treatment in the form of time averaging, ensemble averaging, and various composites. The low-pass filtering structure of time and ensemble averaging eliminates unpredictable high-frequency components from the flow. At the same time more and more information about individual weather events is lost. Thus, the problem of forecasting beyond the deterministic predictability limit becomes essentially one of stochastic prediction (Barnett, 1995).

The organization of this paper is as follows. In the following chapter we describe briefly our model and data. In chapter three, we explain the experimental design of the predictability experiments. In chapter four, we illustrate how we analyze the experimental results using the perfect model framework. In chapter five, we present our results. In the last chapter, we summarize our findings and draw some conclusions. Basically, we conclude that the effect of atmospheric initial conditions should be considered for seasonal forecasts, especially if the boundary forcing is weak.

2 Model and data

We used the seasonal forecasting model from the National Centers for Environmental Prediction (NCEP), which is also known as the Global Spectral Model (GSM). The model uses the spectral transform method to solve the atmospheric primitive equation system for vorticity, divergence, virtual temperature, specific humidity, and logarithm of surface pressure. A triangular truncation of 42 spherical harmonics (T42), equivalent to a horizontal resolution of about 280 km, and a vertical sigma coordinate system which contains 28

layers and 29 levels from the surface to zero hPa is used for these experiments.

The GSM originates from the NCEP medium range forecast model (MRF) that was operational in 1995. The GSM was used for the NCEP/NCAR reanalysis (Kalnay et al. 1996; Kistler et al. 2001), and the NCEP/DOE reanalysis-2 (Kanamitsu et al. 1999). The model version of this study was released in October 2000, and differs from the reanalysis-2 model mainly by its parallelized code structure and improved physical parameterizations. Deep cumulus convection is now parameterized using the "Relaxed Arakawa-Schubert" (RAS) scheme for deep cumulus convection (Moorthi and Suarez, 1992), which is physically more realistic, and which is likely to show a better midlatitude response to tropical heating than the "Simplified Arakawa-Schubert" (SAS) scheme (Grell, 1993) of the older version. A new scheme for long-wave radiation (Chou and Suarez, 1994; Chou et al., 1999) replaces the old GFDL scheme (Fels and Schwarzkopf, 1975). Radiative effects of convective and stratiform clouds are now parameterized using the Slingo and Slingo (1991) formulation. Much effort has also been devoted to improve the physical processes in the soil and land-surface model through a better description of surface vegetation type, vegetation fraction, and soil type, and compatible physics (Hong, 2001; Pan and Mahrt, 1987). A detailed description of the model's formulation is given in the comprehensive documentation of the 1988 version of the model (NMC-Development-Division, 1988). Subsequent model improvements are summarized in Kanamitsu (1989); Kanamitsu et al. (1990); Kalnay et al. (1990); Kanamitsu et al. (1991); Kalnay et al. (1996); Caplan et al. (1997); Kanamitsu et al. (1999); and Kanamitsu et al. (2002).

Observed, climatological or model generated boundary conditions are used to force the various experiments of this study. Over oceans, the observed history of SST and sea ice is imposed as the evolving lower boundary condition. The SST data for the 1948-1981 period were taken from the UKMO Global Ice and Sea Surface Temperature (GISST) data set. After 1982, a satellite-in situ blended SST analysis based on the method of optimum interpolation (Reynolds and Smith 1994) was available from the reanalysis archives. The sea ice distribution was taken from the NCEP/NCAR reanalysis (Kalnay et al., 1996; Kistler et al., 2001). Some experiments were forced every year with the climatological seasonal cycle of SST and sea ice. The climatologies were calculated by averaging the observed monthly mean fields over the 50 year period 1950–1999. The monthly mean data were then linearly interpolated to daily values.

Over land, three different types of boundary conditions were used: For most experiments, land boundary conditions were determined internally by the land surface scheme of the model. In some cases, a climatological mean seasonal cycle for soil moisture and snow cover was prescribed, calculated from the NCEP/DOE reanalysis-2 by averaging over the period 1979-1998. For some experiments, daily observed fields of soil moisture and snow cover, which were also derived from NCEP/DOE reanalysis-2, were specified.

3 Experimental design

The simulations of this study can be roughly divided into three categories, depending on the initial and boundary conditions. Tables 1a–1c provides the name of each experiment, the kind of boundary conditions over oceans and over land, the type of initial conditions, the ensemble size, and the simulation time and period.

3.1 Experiments

Two base runs constitute the first group of experiments (Table 1a). These continuous multi-year GCM simulations produced the initial conditions for the subsequent ensemble predictability experiments. Both base runs started from observed initial conditions at 1 January 1948, which were derived from the NCEP/NCAR reanalysis. Simulation BASE-O was forced with the observed history of SST and sea ice over the 1948–2000 period. Land surface boundary conditions were generated internally by the GCM. The second base run, BASE-C, was run for 77 years, forced with a seasonal cycle of climatological SST and sea ice every year. Land surface boundary conditions were generated internally by the model.

The ensemble predictability experiments were reinitialized every year at 15 December (15 June) and run continuously through the $3\frac{1}{2}$ month long northern hemispheric winter (summer) season. Individual members of each experiment were forced with identical boundary conditions, but were started from slightly different initial conditions. The initial conditions were derived by breeding (discussed later) from the BASE runs or from NCEP/NCAR reanalysis. The validation of the runs depended on the type of experiment and will be explained later.

To test the effect of initial conditions on predictability, two experiments forced with climatological boundary conditions (Table 1b) were conducted. The initial conditions for experiment CC were produced by breeding from BASE-C. We call these initial conditions "climatological", since BASE-C was forced with climatological ocean boundary conditions. Since neither the initial conditions nor the boundary conditions of CC contain anomalies, CC can be regarded as a pure predictability experiment (e.g. Lorenz, 1969). Experiment IC, on the other hand, was initialized by breeding from BASE-O. The initial conditions for this experiment can be thought of as being "anomalous" in the sense that they contain information about the anomalous boundary forcing acting prior to the initialization time. This information is then carried over into the subsequent integration.

The remaining four experiments were forced with observed SST and sea ice, and model generated or reanalysis land boundary conditions (Table 1c). These experiments differ mainly in the way how they were initialized. ICBC was initialized from BASE-O. Under the perfect model assumption, this is equivalent to using "observed" initial conditions. ICBC-r was initialized from reanalysis, which is also equivalent to "observed" initial conditions when verifying against

Table 1. (a) Base run characteristics. Specified are boundary ("BCs") and initial conditions ("ICs"), as well as ensemble size ("size"), and simulation period ("days" and "years") **(b)** Experiments with climatological boundary conditions. "r-2" indicates NCEP/DOE reanalysis-2 were used. See (a) for other details. **(c)** Experiments with observed ocean boundary conditions. See (a) for other details

(a)

name	ocean BCs	land BCs	atmos. ICs	land ICs	sizes	days	years
BASE-O	observed	model	observed 1/1/48	observed 1/1/48	1	1/1–12/31	1948–2000
BASE-C	climatology	model	observed 1/1/48	observed 1/1/48	1	1/1–12/31	1948–2024

(b)

name	ocean BCs	land BCs	atmos. ICs	land ICs	sizes	days	years
IC	climatology	r-2 climatology	BASE-O	—	10	12/15–03/31 6/15-09/31	1979–2000
CC	climatology	r-2 climatology	BASE-C	—	10	12/15–03/31	1979–2000

(c)

name	ocean BCs	land BCs	atmos. ICs	land ICs	sizes	days	years
ICBC	observed	model	BASE-O	BASE-O	20 (10)	12/15–03/31 6/15–09/31	1979–2000
BC	observed	model	BASE-C rndm.	BASE-C rndm.	10	12/15–03/31 6/15–09/31	1979–2000
ICBC-r	observed	r-2	r-2	—	10	12/15–03/31	1979–2000
iBC	observed	model	ICBC, 1 yr lag	ICBC, 1 yr lag	10	12/15–03/31 6/15–09/31	1980–2001

reanalysis. Through initial and boundary conditions, these two simulations receive the maximum possible amount of information about the state of the system.

Simulation iBC is very similar to ICBC. The only difference are the initial conditions, which were produced by integrating ICBC for one whole year, i.e. from 15 December (June) of the current year to 15 December (June) of the next year. The final state was then used to initialize iBC. For example, to initialize iBC for 15 December 1989, we would run ICBC from 15 December 1988 for one whole year to 15 December 1989. These initial conditions have completely lost their memory from the previous year, but they are adjusted to the boundary forcing at the new initialization time. The initial conditions of iBC and ICBC for the same year evolved under the influence of the same boundary conditions, but through a different history of complex non-linear interactions. It can therefore be expected that the amplitude and phase of the main energy-carrying planetary waves are similar, at least to the extent that they are controlled by boundary conditions. The synoptic and smaller scale components of the flow fields, however, are completely different. Another way of thinking is that two different AMIP type base runs were used to initialize iBC and ICBC. Both experiments, however, were verified against ICBC. There-

fore, ICBC initial conditions were almost perfect, whereas iBC initial conditions are not. The motivation for this experiment was to find out how much predictability is lost by excluding the beneficial effects of synoptic scales in the initial conditions on dynamical predictability, but by retaining the effect of evolving boundary forcing on initial conditions. In this respect, experiment iBC is comparable to an ensemble of continuous AMIP-type integrations. Note that iBC is also comparable to the operational seasonal forecasts at the International Research Institute for Climate Predictions (IRI). In their two-tiered approach, the GCM is initialized from a continuous AMIP-type run, which is forced with observed ocean boundary conditions (see Mason et al., 1999).

The final experiment, BC, contains only information from boundary forcing. It was started from randomly chosen "climatological" initial conditions, which were derived from BASE-C. This experiment allows us to study the effect of boundary forcing alone without the possible influences from initial conditions. It should be noted here that the design of some of our experiments contains certain unavoidable caveats. Atmospheric initial conditions are not unrelated to the boundary conditions for the same day. Initial conditions, which are not in balance with the imposed boundary forcing, may lead to spin-up problems, which tend to make the

model less predictable (e.g. Anderson and Ploshay, 2000). Experiments BC and IC may be affected to some extent by this, but we assume that this influence is small. We note also that due to technicalities with the model, the climatological or observed values for soil moisture and snow for experiments CC, IC, and ICBC-r were prescribed only every 24 h. Between these intervals, the land model is allowed to communicate interactively with the model atmosphere to update land boundary conditions. The 24 h update period, however, is much smaller than the time scales that are typical for land boundary conditions. We believe therefore that this effect is negligible.

3.2 Simulation time and period

The predictability experiments cover the 22 year period from 1979 to 2000. Each year, the model was initialized at 15 December (June) 00:00 Z, and then continuously integrated for 107 days through 1 April (October) 00:00 Z of the following year (see Table 1). The experiments for the Northern Hemisphere winter will be denoted by DJFM, and the experiments for summer by JJAS (it should be noted that this simulation time and period is identical to the DSP project). The model output was saved in 12 h intervals for further analysis. The simulations were carried out for both the northern hemispheric winter and summer to capture the effects of seasonal variations in the strength of the ENSO signal and in the background state of the atmosphere.

3.3 Breeding

Ensemble predictions are used to isolate reproducible atmospheric signals from unpredictable components of internally generated variability or noise. Forecasts start from slightly different initial conditions, which should represent the observational uncertainty in the atmospheric initial conditions. Each ensemble member can then be thought of as tracing one of the many possible paths in the phase space of the atmosphere. The average over all realizations filters out the noise and returns the most likely evolution of the forced atmosphere.

The probabilistic nature of the forecasting problem requires a sufficiently large ensemble. Our standard ensemble size contained 10 members (see Table 1). For reference experiment ICBC, which was used to verify other simulations, an ensemble size of 20 was used during winter. From earlier studies we concluded that for seasonal means in the extratropical atmosphere, 8–10 members produce a reasonably sized ensemble (e.g. Leith, 1974; Palmer et al., 1990; Barnett, 1995; Kumar and Hoerling, 1995; Shukla et al., 2000). Sardeshmukh et al. (2000) stress the need for larger ensembles, but since our results are based on composites over many years, and since we are examining skill over very large regions, we believe that an ensemble size of 10 was more than adequate for all of our investigations.

We were also concerned about producing the best possible initial conditions for the ensembles. Some of the early methods to perturb initial conditions for short term numerical weather forecasts were "Monte Carlo", which just added suitable scaled random perturbations (e.g. Tribbia and Baumhefner, 1988), and "lagged average forecasting", which used the difference between a previous forecast and the present analysis as perturbation (Hoffman and Kalnay, 1983). More modern methods include "breeding of fast-growing modes" (Toth and Kalnay, 1993; Toth and Kalnay, 1997) and the "singular vector method" (Lorenz, 1965; Buizza and Palmer, 1995). On the longer climate time scales, where the influence of the initial conditions is assumed to be minimal, not much attention has been paid to the problem of creating perturbations. Common methods are Monte Carlo (e.g. Shukla, 1981), using analyses for the same date but from different years (e.g. Barnett, 1995), or using analyses centered around the date but several hours apart (e.g. Chang et al., 2000; Anderson and Ploshay, 2000). With the latter method, all initial conditions originate from the same trajectory of the observed atmosphere. Consequently, different forecasts tend to evolve along the same trajectory, which may result in an overestimate of predictability.

We used the breeding method, which is a methodology for getting the fastest growing modes for a particular initial state from the atmospheric model itself. As described by Toth and Kalnay (1993, 1997), the resulting breeding vectors are mathematically related to the Lyapunov exponents, which point into the direction of the phase space that can grow fastest in a sustainable manner. The method is standard for operational forecasting at NCEP. A breeding cycle starts by adding and subtracting ten different small initial perturbations to the initial atmospheric field five days ahead of the actual initialization date (10 December or 10 June). The initial perturbations are those used operationally at NCEP truncated to T42 resolution (the breeding algorithm and the initial perturbations were made available to us by Z. Toth from NCEP). The model is integrated for 24 h from the 20 perturbed and from the original unperturbed initial conditions. The differences between the unperturbed and each of the perturbed forecasts are then used to calculate the next 20 perturbed initial conditions for the following 24 h breeding forecast. These differences are scaled down to the size of the initial perturbation. The scaling is based on the 500 hPa rotational kinetic energy, and uses a time and space dependant mask which takes into account observational uncertainties. The scaled differences are added to the initial atmospheric state for the following day to create the next set of perturbations. The last two steps, i.e. 24 h forecast and scaling of the differences, are repeated five times. The final product includes twenty perturbed initial conditions for the 15th of the month, which contain the fastest growing modes for the particular atmospheric state. The mean rms differences between two individual perturbed initial conditions at the 500 hPa level over the northern hemisphere were ∼3 m/s for the u and v wind components, and ∼20 m for the geopotential height. The maximum differences at individual grid points reached ∼17 m/s for the wind, and 110 m for the geopotential height.

It should be noted that the breeding method of this study

creates fastest growing modes within the context a few days, i.e. for the weather forecasting range. Since we are concerned with a much wider time range from days out to one season, it would be interesting to use an alternative method which creates fastest modes for longer time scales. This issue, however, can only be addressed by future studies.

4 Analysis procedure

Predictability is measured from the spatial anomaly correlation (AC) of a certain pressure surface over four specific areas. Areas of interest were (a) the Northern Hemisphere (NH) (20° N–90° N), the Pacific North American (PNA) region (180° W–60° W, 15° N–70° N), the Southern Hemisphere (SH) (20° S–90° S), and the tropics (20° N–20° S). We calculated the AC from the 500 hPa geopotential height over the extratropical regions (NH, PNA, SH), and from the 200 hPa geopotential height over the tropics.

The AC is designed to detect similarities in spatial patterns (Wilks, 1995) of two fields, which we will refer from now on as "forecast" and "verification". The AC is a common measure for skill of field forecasts. It is used in many studies of atmospheric predictability, but it is highly sensitive to the technicalities of the analysis procedure (Anderson et al., 1999). The main shortcoming of the AC is that it is insensitive to systematic biases and to scale; it does not change when a field is scaled by a positive constant or when a constant is added to the field. The AC also depends strongly on the choice of a reference climate for the computation of the anomalies (e.g. Barker and Horel, 1989). In this study, anomalies are calculated with respect to the daily climatology of each simulation. This removes the seasonal cycle from the data and corrects systematic model errors. The climatologies for a specific simulation are computed by taking daily averages from all members of that simulation over the 1979–2000 period.

The AC is defined as the spatial correlation coefficient over some region between the anomalies of the forecast and the verification. Let $F = F(\mathbf{x}, t)$ be the forecast, and $V = V(\mathbf{x}, t)$ and verification field at location $\mathbf{x}$ of the 12 h model output at time t. Then,

$$AC_i(t) = \frac{\int_A (\tilde{F} - \bar{F})(\tilde{V}_i - \bar{V}_i)\,dA}{\sqrt{\int_A (\tilde{F} - \bar{F})^2\,dA \int_A (\tilde{V}_i - \bar{V}_i)^2\,dA}} \quad (1)$$

defines the anomaly correlation using verification member i. dA is the differential surface element of the geographical area A under consideration, $\bar{F}$ and $\bar{V}_i$ are the respective area averages, e.g. $\bar{F} = \frac{1}{2}\int_A \tilde{F}\,dA$, $\tilde{F}$ is the ensemble and time averaged filtered anomaly of the forecast, and $\tilde{V}_i$ is the time averaged filtered anomaly of only one member of the verification. More specifically,

$$\tilde{F}(\mathbf{x}, t) = \left| \overline{\overline{F_j(\mathbf{x}, t) - \langle F(\mathbf{x}) \rangle}^J}^T \right|_{n1}^{n2} \quad (2)$$

and

$$\tilde{V}_i(\mathbf{x}, t) = \left| \overline{V_i(\mathbf{x}, t) - \langle V(\mathbf{x}) \rangle}^T \right|_{n1}^{n2}, \quad (3)$$

where $F - \langle F \rangle$ and $V - \langle V \rangle$ are the anomalies of the two fields with $\langle \rangle$ being the climatological mean over all ensemble members and years, $\overline{()}^T$ a temporal mean over some time interval T, $\overline{()}^J$ an ensemble average over ten or nine ensemble members $j = 1...9$ or 10 of the forecast, and $||_{n1}^{n2}$ a spatial filter which includes total wave numbers from $n1$ to $n2$. Usually, the spatial filtering coefficients are chosen as $n1 = 0$ and $n2 = 42$, i.e. all spatial scales from the flow field are retained. The time average interval T ranges between 1 (instantaneous 12 h skill), 10 (5 days), 30 (2 weeks), 60 (1 month), 120 (2 months) and 180 (one season) output intervals.

Note that each AC_i is calculated from the ensemble mean $\overline{F_j}^J$ of the forecast, but that ensemble averaging is not applied to the verification field V_i. Leith (1974) showed that the ensemble mean provides in a statistical sense a forecast more reliable than any of the single forecasts, including that started from the control analysis. This is because of the optimal filtering nature of the averaging procedure, which damps phase decorrelated, erroneous, small scale structures in favor of more predictable large scales. Consequently, ensemble averaging will increase the value of the AC. Since real world atmospheric states do not occur more than once, we use only one single realizations V_i to verify the ensemble mean forecast. Since each member of the verification ensemble can represent real observations, a more robust estimate of the skill is obtained by computing the skill with each member of the ensemble being treated as verification in turn and averaging over the individual results. Thus, the final AC for a particular date is simply given by averaging over all I members of the verification experiment, i.e.

$$AC(t) = \frac{1}{I} \sum_{i=1}^{I} AC(t)_i. \quad (4)$$

If a simulation was verified against itself, one ensemble member was taken as verification, and the ensemble mean over the remaining $I - 1$ members was taken as forecast.

AC values are highly non-normally distributed since their range of values is limited between -1 and $+1$. In some cases, however, it is desirable to have a more normal distribution of the skill values. This can be achieved by applying a Fisher z-transformation (e.g. Roads, 1988) to the AC values, i.e.

$$z_i = \frac{1}{2}ln\frac{(1 + AC_i)}{(1 - AC_i)}. \quad (5)$$

The mean spread SP between ensemble members is given by the mean rms difference between all possible pairs of forecast members over some area A, i.e.

$$SP(t) = \frac{1}{N} \sum_{i=0}^{J-2} \sum_{j=i+1}^{J-1} \sqrt{\frac{1}{A}\int_A \left(\tilde{F}_i - \tilde{F}_j\right)^2\,dA}. \quad (6)$$

Note that this measure of spread does not depend on a verifying field. It will be later used as potential indicator for the skill of the ensemble mean.

In some of our analysis we use a wavenumber representation in terms of spherical harmonic base functions. The spherical triangular expansion of a field Z is given by

$$Z(\phi, \lambda, t) = \sum_{m=-N}^{N} \sum_{n=|m|}^{N} \hat{A}_{nm}(t) Y_n^m(\phi, \lambda), \qquad (7)$$

where Y_n^m are the spherical harmonics, $\hat{A}_{nm}$ are the expansion coefficients, N is the truncation, n is the total wavenumber, and m is the zonal wavenumber. This expansion is a standard numerical method. It is, for example, described in more detail in Hoskins (1975). Prior to the expansion, the climatological mean is removed from the data at each grid point. We use a truncation of $N = 42$, so that each field is represented by 946 expansion coefficient for different combinations of n and m. The coefficients are appropriately scaled so that they carry the units of the field.

We calculate a wave number spectrum from the modulus of the complex coefficients by taking the average over various ensemble members, time steps and years, i.e.

$$P_{nm}^2 = \left\langle \hat{A}_{nm} \hat{A}_{nm}^* \right\rangle. \qquad (8)$$

A one-dimensional spectral representation is achieved by summing over all zonal wave numbers m for a given n. Since $P_{n(-m)}^2 = P_{nm}^2$ (see e.g. Boer, 1984) it follows that

$$P_n^2 = \sum_{m=-n}^{n} P_{nm}^2 = \sum_{m=0}^{n} (2 - \delta_0^m) P_{nm}^2. \qquad (9)$$

The spectral error energy at a given lead time t between the coefficients of a forecast and a verification is determined by

$$E_{nm}(t)^2 = \left\langle \left| \hat{F}_{nm}(t) - \hat{O}_{nm}(t) \right|^2 \right\rangle, \qquad (10)$$

where the averaging $\langle \rangle$ is taken over many members and years at the same lead time. It is useful to express E_{nm}^2 relative to the maximum error $E_{nm}^{*}{}^2$

$$S_{nm}^2 = E_{nm}^2 / E_{nm}^{*}{}^2, \qquad (11)$$

where the saturation error is given by the sum of the two climatological variances, i.e.

$$E_{nm}^*(t)^2 = \left\langle \left| \hat{F}_{nm}(t) \right|^2 \right\rangle + \left\langle \left| \hat{O}_{nm}(t) \right|^2 \right\rangle. \qquad (12)$$

Note that $E_{nm}^{*}{}^2$ is also a function of time because of seasonal changes in signal and noise.

5 Results

Below we present a statistical analysis of the various ensemble experiments. The goal is to document the long-term effects of initial conditions and boundary conditions on predictability, and to find out how the results change for different regions, seasons and years. First, the effects of averaging

are discussed, and we illustrate the time evolution of skill with very high temporal resolution. Then, we investigate the long-term effects of initial conditions on predictability without influences from the boundaries. Next, we examine the effects of boundary forcing, and we look at the regional and seasonal aspects of long-term predictability. We compare the perfect model results with real world skill to find out how large model induced error components are. Then, we present a year-to-year breakdown of seasonally averaged predictability and examine the relationship between skill and spread. Finally, we examine how predictability varies for different scales of motion, and show which scales are affected by boundary forcing.

5.1 Averaging

What are the effects of averaging over ensemble members, over many years, and over increasingly longer lead times? We choose simulation CC to investigate the time evolution of skill at different stages of the averaging process. Figure 1a depicts the time evolution of the unaveraged 12 hourly AC of the global 500 hPa height of simulation CC from day 0 (15 December) out to day 106 (31 March). Shown are the results from 10 individual verifying members for one arbitrarily selected model year (1989/1990). Under the perfect model assumption, each forecast member has been used in turn to verify the ensemble average of the 9 remaining members. By taking ensemble averages, unpredictable flow components are filtered out from the forecast, so that the overall predictive skill increases. Figure 1b shows the average over the ten individual verification members for the same year, and the length of the associated 95% confidence interval. The interval was calculated from the Fisher z-transformed skill values of all individual verifying members, using a two sided Student's t-test (e.g. Wilks, 1995), and assuming that the individual values are normally distributed and independent.

Figure 1c depicts the ensemble mean skill averaged over all 22 model winters (1979–2000), and the 95% confidence interval. Each data point of the mean curve represents the average of 220 individual AC calculations, from 10 verification members and 22 winter seasons. The averaging processes reduces the sampling uncertainty to tolerable levels, even for instantaneous forecast skill. Even though simulation CC was forced with climatological boundary conditions, instantaneous forecast skill can be detected out to several weeks. The AC values are actually very small, but as will be shown later, the skill increases considerably by using time averaging, by adding anomalous initial conditions and boundary forcing, and by focusing on certain better predictable regions. The main purpose of displaying unaveraged forecast skill is to determine later the exact time ranges for the influence of initial conditions and boundary forcing. For comparison, Fig. 1c also shows the skill of a persistence forecast (red curve), which is made simply by persisting day 0 for all lead times. The skill of the persistence forecast is much lower than that of the dynamical forecast, and demonstrates the potential value of the dynamical model. Note that the
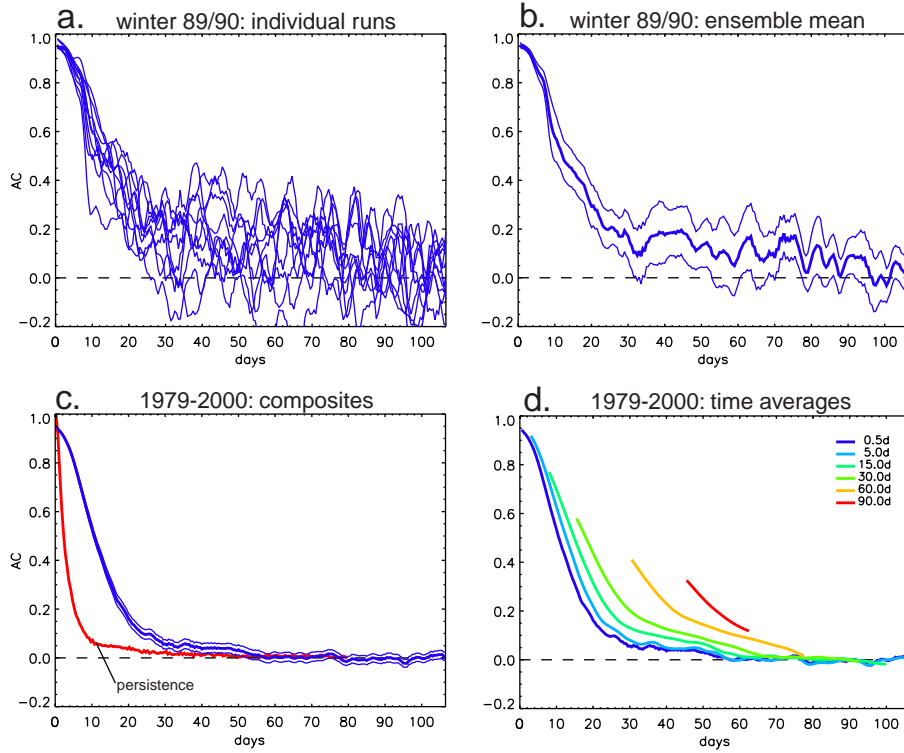
Fig. 1. Time evolution of global skill for simulation CC verified against itself: **(a)** instantaneous forecast skill from verifying against individual members during model winter 1989/1990; **(b)** ensemble averaged skill for the same year, and 95% confidence interval; **(c)** composite forecast skill, averaged over 22 years and 10 verifying ensemble members (blue thick) and associated 95% confidence interval (blue thin); skill of persistence forecast (red); **(d)** composite skill for different time averaging intervals.

average skill over all years (Fig. 1c) is lower than the skill for the specific year 1989/1990 (Fig. 1b). This means that 1989/1990 initial conditions were associated with very high long-term predictability.

Figure 1d demonstrates the effect of taking increasingly longer time averages of the input data before ACs are calculated. The curves are for different lengths of the time averaging interval T in Eqs. (2) and (3), reaching from no averaging (0.5 days) to seasonal averaging (90 days). The curves are centered at the middle of the averaging interval. For example, the first (last) data point of the seasonal skill corresponds to the mean of day 0–89 (17–106) and is centered at day 45 (62). With longer averaging periods, the forecast skill improves at the expense of losing information about individual weather events and explaining less of the original unaveraged variance. The improvement in skill is mainly a consequence of the low-pass filtering structure of time averaging process, which eliminates unpredictable features which propagate with periods less than the averaging time. To a lesser extent, the increase of skill is also consequence of the serially correlated data and the transfer of information from higher skill at the beginning of the forecast to lower skill at later times (e.g. Roads, 1986). To find out how strong this latter effect is, we approximated the instantaneous skill curve by an AR(1) process (not shown). From the AR(1) process it is straightforward to derive a 90-day time averaged skill. It

turns out that the time averaged skill from serial correlation alone is much smaller than the 90-day time averaged forecast skill shown in Fig. 1d. This indicates that the increase in temporal averaged skill is more a consequence of low frequency predictability than simply initial state contribution to time averages.

We further investigate the effects of time and ensemble averaging from a spectral analysis of the global 500 hPa geopotential height fields. The left panel of Fig. 2 shows the one-dimensional power spectrum P_n^2 calculated after Eq. (9) from simulation CC for different time averaging periods with and without ensemble averaging. The spectral maximum at wavenumbers 3–4 corresponds to the main energy carrying planetary waves. The reduction in power relative to the unaveraged spectrum is shown in the right panel of Fig. 2, which demonstrates that time averaging acts like a spatial low-pass filter. The dashed curves in Fig. 2 are for ensemble averaged data. They show that ensemble averaging is more effective in variance reduction than time averaging alone. Because of the relative strong serial correlation of atmospheric fields, different dates of the same member are less phase decorrelated than various members for the same date. Roughly speaking, the effect of averaging over 10 ensemble members of 12 h data is comparable to taking 90 day time averages of individual members.

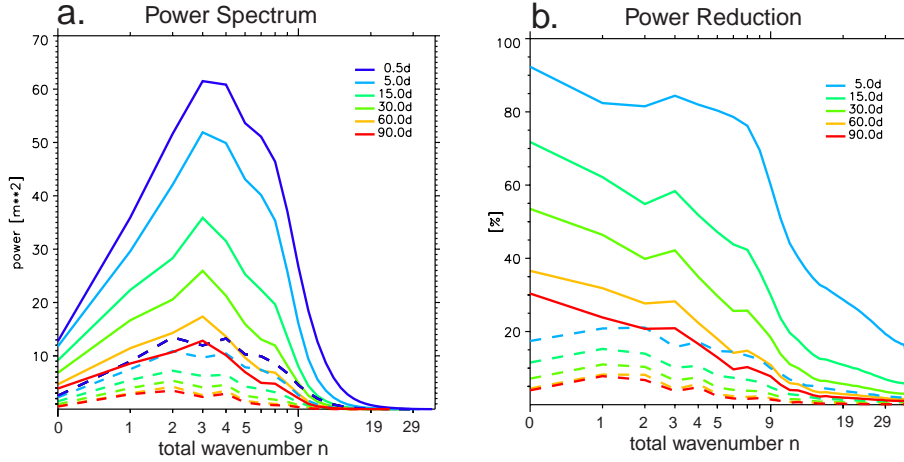The effects of ensemble averaging can be best seen from a

Fig. 2. (a) Wavenumber spectra derived from the 500 hPa fields of simulation CC for various temporal averaging intervals. Dashed lines show spectra for ensemble and time averaged data. **(b)** Power reduction relative to unaveraged spectrum without (continuous) and with (dashed) ensemble averaging. Note logarithmic scale for x-axis.
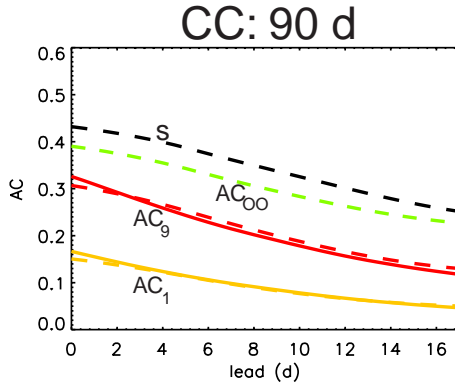

Fig. 3. Seasonally averaged AC of simulation CC for different ensemble sizes as function of lead time. AC_1 (yellow) and AC_9 (red) were measured from global 500 hPa heights during DJFM. Dashed red (AC_9) and dashed yellow (AC_1) are calculated from measured AC_1 and AC_9, respectively. AC_∞ in green was calculated from AC_1. Black is signal to noise ratio s derived from AC_9 and AC_1.

direct comparison with the average skill of individual members. Figure 3 shows for simulation CC the time evolution of global seasonally averaged skill, which was calculated with and without ensemble averaging. Curves in yellow show AC_1 from individual members, and the red curves show AC_9 using 9 member ensemble averaging. At all leads, the skill of AC_9 is much higher than AC_1, demonstrating the positive effects of ensemble averaging. We can compare this increase in skill by ensemble averaging with conceptual models which have established relationships between the anomaly correlation and ensemble size (e.g. Brankovic et al., 1990; Kumar and Hoerling, 2000; Kumar et al., 2001; Sardeshmukh et al., 2000; Rowell, 1998). In general, the increase in skill by the ensemble technique is a function of the ensemble size n and the signal to noise ratio s. Maximum efficiency is achieved for large n and intermediate s values. If s is too small or too

large, then ensemble averaging does not change much the either very low or very high correlations. Sardeshmukh et al. (2000), for example, showed that the ensemble averaged anomaly correlation for a perfect model can be expressed as

$$AC_n = s^2 / \sqrt{(s^2 + 1)(s^2 + n^{-1})}, \qquad (13)$$

where

$$s^2 = \frac{\langle \mathbf{x} \rangle \cdot \langle \mathbf{x} \rangle}{\langle \mathbf{x}' \cdot \mathbf{x}' \rangle} = \frac{|\text{ensemble mean anomaly}|^2}{|\text{ensemble mean spread}|^2} \qquad (14)$$

is a measure of the signal to noise ratio. n is the ensemble size, $\langle \mathbf{x} \rangle$ is the ensemble mean anomaly state vector, and $\mathbf{x}'$ are the variations of the anomaly state vector around $\langle \mathbf{x} \rangle$. If AC_n for some n is known, then the AC for any other n can be derived from Eq. (13). For example, the theoretical upper predictability limit for an infinite ensemble size is given by

$$AC_\infty = \sqrt{AC_1}. \qquad (15)$$

The dashed green curves in Fig. 3 show AC_∞ which was calculated using Eq. (15). As expected, AC_∞ is everywhere larger than AC_9. However, both curves are close together, which indicates that there is not much room for skill improvements by further increasing the ensemble size. From Eq. (13) one can also diagnose AC_9 from AC_1 and vice versa. The result is shown by the dashed red and yellow curves in Fig. 3. Measured and calculated ACs are in good agreement, and the small differences can be explained as coming from limited samples.

The signal to noise ratio s is an important characterization for the potential predictability of the underlying data. One can see from Eq. (13) that s can be determined if AC for a given n is know. s can actually be derived from either AC_1 or AC_9. Using both has the advantage of giving more stable results. Solving Eq. (13) for s gives
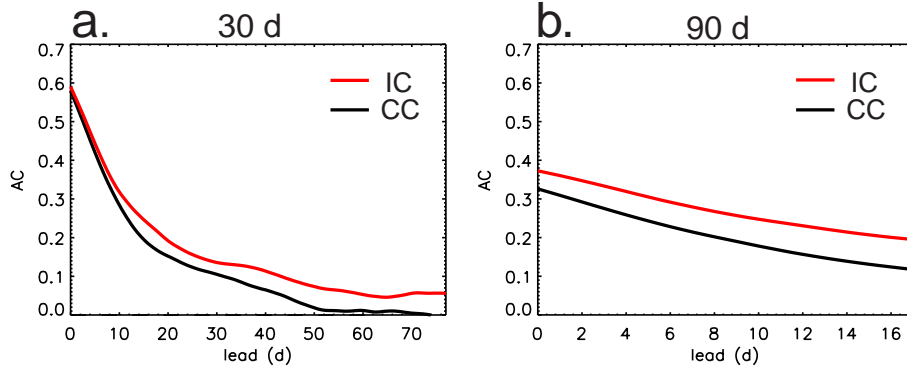
$$s = \pm \sqrt{\frac{1 - a/n}{a - 1}}, \qquad (16)$$

Fig. 4. Lag evolution of global 500 hPa AC for simulation IC (red) and CC (black), verified against itself under the perfect model assumption. The curves show composite ensemble mean skill over 21 winter seasons (1980–2000): **(a)** from 30 days time averaged data; **(b)** from 90 days time averaged data.
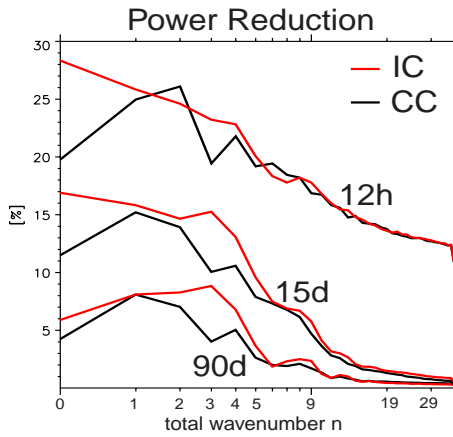

Fig. 5. Power reduction of global 500 hPa heights due to ensemble and time averaging relative to unaveraged data. Results are for ensemble averages from simulation IC (red) and CC (black) for 12 h, 15 days and 90 days time averaged data taken from all years (1979–2000).

where a is

$$a = \left(\frac{AC_n}{AC_1} \right)^2. \qquad (17)$$

The result for simulation CC is shown in Fig. 3 by the black dashed curves. s ranges between values of 0.4 and 0.3, which is actually quite small. This indicates that noisy unpredictable components greatly exceed the predictable components. This should not be very surprising since s was determined from experiment CC, which has seen only climatological boundary forcing. The values derived here represent therefore the lower limit for the signal to noise ratio. More s values from the more realistic experiment ICBC will be discussed later for various regions, averaging periods and years.

5.2 Initial conditions

It is clear from the previous section that a purely climatologically forced run can produce considerable forecast skill

simply from initial conditions alone. To further investigate the influence of initial conditions, we examined if it makes any measurable difference whether climatological or anomalous initial conditions are used, but still keeping the boundary conditions at their climatological values. This was examined by comparing the results from the two unforced simulations, CC and IC. Remember, the initial conditions for IC were derived from BASE-O, and they are adjusted to the observed anomalous boundary forcing at the initialization time. In analogy to the anomalous boundary forcing, we call them "anomalous initial conditions". In contrast, the initial conditions for CC came from a climatological base run, so that we will term them as "climatological initial conditions".

Figure 4 shows the time evolution of forecast skill as measured by the AC of the global 500 hPa height for the two simulations. Shown are composites over 21 winters from 1980 to 2000. Each simulation was verified against itself. Figure 4a illustrates forecast skill from 30 days averaged data. During the first 10 days, the skill of both simulations decreases rapidly. After that, the skill approaches slowly zero. This indicates that certain components of the atmospheric flow are predictable for several weeks, simply because of the influence of initial conditions. Note that simulation IC shows larger skill than simulation CC at all lead times. Statistically, the skill of IC always remains greater than zero, whereas that of CC reaches zero after 50 days.

The seasonally averaged skill in Fig. 4b shows a similar behavior. The difference in skill between IC and CC increases with lead time, and is everywhere statistically significant at the 5% error level. This proves that different initial conditions have different predictability. Our results thus show that anomalous initial conditions are more predictable than climatological initial conditions. The difference in the seasonally averaged AC is 0.1 to 0.2 at maximum lead time.

To better understand the difference between anomalous and climatological initial conditions, we calculated the power spectra from global 500 hPa fields for both simulations. Figure 5 shows the reduction in power due to ensemble and time averaging for different averaging intervals. The reduction
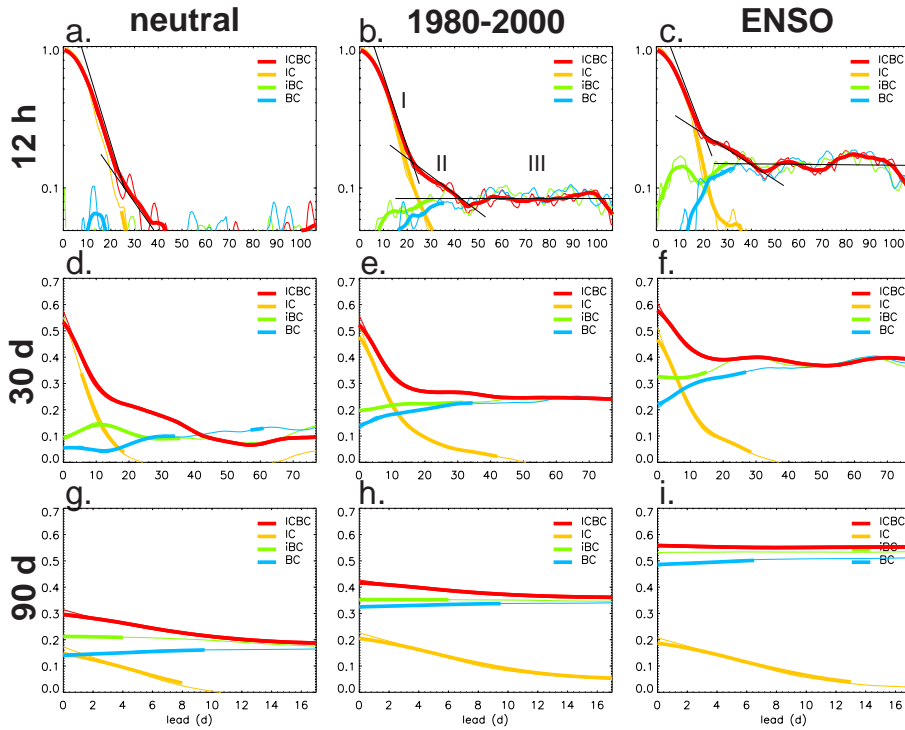
Fig. 6. Lag evolution of the spatial AC of the 500 hPa surface over the Northern Hemisphere (20° N–90° N) from 15 December to 31 March. All experiments have been verified against experiment ICBC. Results are based on all ensemble members of the indicated years. Top panels show AC from unaveraged 12 h data, middle panels are from 30 days time averaged data, and bottom panels are from 90 days time averages. Thick lines have been smoothed with a 10-day running mean filter, and are only shown where the AC is significantly (5% error) different from zero and from the reference experiment ICBC. Y-axis denotes the AC. Note the logarithmic y-axis for the top panels (range: 0.05–1.0).

measures the degree to which individual wave number components decorrelate in time and therefore lead to a decrease in skill. Generally, IC shows a smaller reduction than CC, in particular at the larger scales. Since the larger scales carry more energy, this explains the difference in skill between the two simulations. We hypothesize that boundary forcing leads to excited large-scale low-frequency modes, which are then included in the anomalous initial conditions of IC. These modes remember themselves longer than the unexcited modes from climatological initial conditions. In other words, the atmosphere remembers through its initial conditions the past history of boundary forcing.

5.3   Boundary forcing

How does boundary forcing alone and in combination with initial conditions affect predictability? Let us now compare the three boundary forced experiments (ICBC, iBC, BC) with the unforced simulation IC. Figure 6 shows the time evolution of their ACs calculated from 500 hPa heights over the Northern Hemisphere for different time averaging periods and years. Displayed are mean ACs from verifying the ensemble mean forecast against 20 different realizations of the reference experiment ICBC. The time averaging period is instantaneous (top row), 30 days (middle row), and 90 days (bottom row). The middle column shows averages over all 21 years (1980–2000). Thin lines denote unsmoothed skill, and

thick lines represent 10 day running mean smoothed results. Smoothed results are only drawn where the skill is significantly greater than zero and significantly different from the verification forecast ICBC. The significance is derived from a t-test at the 5% error level.

First, consider the skill of simulation ICBC, which is shown in red. This simulation represents the upper predictability limit for this model. ICBC is being forced with observed ocean boundary conditions, and initialized from correct initial conditions (BASE-O). Therefore, initial and boundary conditions are perfectly balanced. The instantaneous forecast skill of ICBC (Fig. 6b) is at all lead times positive. Note that we have chosen a logarithmic scale to depict instantaneous forecast skill so that we can better focus on the small skill values. The AC curve shows three different sections which are indicative for specific error growth rates: During the first 20 days, the AC drops quickly to values of around 0.2. Then, the slope flattens, and after day 40 or so, the correlations remain more or less constant at about 0.1. The three sections are marked with roman numbers in Fig. 6b. In each section, the skill decreases approximately linearly, as indicated by the black lines. Considering the logarithmic y-axis, this corresponds to an exponential decrease in AC. We will show later that the initial rapid decrease is related to the error growth of the synoptic scales. After 20 days or so, when the error spectrum of the synoptic scales is

presumably saturated, the slower error growth of the larger planetary scales is noticeable. After 40 days or so, the overall error growth finally saturates, resulting in constant skill thereafter. The division of skill in three sections – initial rapid growth, intermediate slow growth, and final saturation – can also be found in other simulations and over other areas. Qualitatively, this agrees with the study of Dalcher and Kalnay (1987), who also found an increase in error growth with wave number.

Let us compare ICBC with the unforced simulation IC (yellow curves, Fig. 6). Both simulations start from identical initial conditions and therefore show similarly high skill at the beginning. The unaveraged skill of both simulations start to differ from each other after 10 days or so, and they are significantly different after roughly 16 days. IC verified against ICBC reaches zero skill after 30 days.

Next, we consider the skill of simulation BC, which starts from climatological initial conditions, and which is then forced with observed boundary conditions (blue curves, Fig. 6). BC can be regarded as the complementary experiment to simulation IC. Both experiments measure how quickly the atmosphere responds to changing boundary conditions. In the case of BC, anomalous boundary conditions are imposed onto a neutral initial state, and the evolution of skill tells us how quickly the atmosphere adjusts from neutral to forced boundary conditions. From IC we learn how quickly the atmosphere forgets its anomalous initial state when no anomalous forcing from the boundaries is subsequently applied. The unaveraged skill of BC starts to be significantly different from zero after 10–20 days, which is comparable to the time for simulation IC to be significantly different from ICBC. An interesting time scale is given by the crossover point between IC and BC, which indicates how long initial conditions are more important than boundary forcing. From the unaveraged skill for all years (Fig. 6b), the crossover is at 20–30 days. After this time, boundary forcing is more important for forecast skill than the memory of the initial state.

The skill of simulation iBC is represented by the green curves in Fig. 6. Again, this experiment is a one year extension of ICBC. Its initial conditions are comparable to that of an AMIP-type simulation, and do not show the deterministic predictability regime at the beginning of the forecast. Without seasonal variations of the signal to noise ratio, the skill of iBC would be constant in time, since there are no influences from the initial conditions, and since this simulation is perfectly adjusted to its boundary forcing. However, iBC starts (in December) with lower skill than at later times. This is an expression of seasonality in signal and noise, and therefore in predictability. Comparing instantaneous skill of iBC with ICBC indicates that it takes 30–40 days until both simulations reach similar skill. This time is much longer than the roughly 20 days it takes ICBC to saturate its error spectrum for the synoptic and smaller scales. The difference is in the larger scales. We conclude that having perfect large scale initial conditions improves predictability out to 5 weeks, and that good initial conditions are therefore important on sub-

Table 2. Classification of strong ENSO and neutral years (year of January)

event	years
strong ENSO	1983, 1985, 1987, 1989, 1992, 1998, 1999, 2000
neutral	1981, 1982, 1986, 1990, 1994, 1997

seasonal to seasonal time scales.

The effects of initial conditions on forecast skill are more evident for 30-day averaged skill (Fig. 6e). Note that the range of lead time decreases with increasing averaging interval. As expected, ICBC is the best forecast at all lead times. From the differences between ICBC and BC one can see that for monthly forecasts, initial conditions are important out to 4–5 weeks lead time. iBC shows somewhat higher skill than BC, as one might expect from the better initial conditions. For seasonally averaged forecasts (Fig. 6h), initial conditions are important at short lead times. At zero lead, the difference in seasonal forecast skill between ICBC and BC is about 0.1. This difference is statistically significant, and proves that initial conditions impact even seasonal forecasts for short lags. Similar conclusions hold for simulation iBC and ICBC. The differences between the three boundary forced simulations (ICBC, BC, iBC) become statistically insignificant after 6 to 8 days.

How do the previous results change with the strength of ENSO forcing? To answer this, we subsampled the data from our experiments and included either only years from major ENSO events, or years where the conditions over the tropical Pacific were near neutral. The classification of strong ENSO and neutral years stems from the Climate Prediction Center at NCEP (published on the internet: www.cpc.ncep.noaa.gov), which provides a season-by-season breakdown of the conditions of the tropical Pacific. Table 2 shows the years which are classified as near neutral or as strong ENSO years. During the period 1980–2000, there were 8 major cold or warm events, and 6 years could be classified as being near neutral.

The time evolution of global skill during the eight major ENSO events for the previously discussed experiments is illustrated in the right column of Fig. 6, and the skill during neutral years is shown in the left column. Since this analysis is based on less data, the scatter and the statistical uncertainty are higher than the previous results, which used all years. During ENSO (Fig. 6c), the unaveraged AC for the boundary forced simulations levels off close to 0.2, as compared to 0.1 for all years (Fig. 6b), and much below 0.1 during neutral years (Fig. 6a). This difference illustrates a well known enhancement in predictability with tropical forcing. According to the changes in instantaneous skill, the 30 and 90 days averaged skill of ICBC during ENSO is much higher than during neutral years. During ENSO, the seasonally averaged ACs of the three boundary forced experiments ICBC, BC and iBC differ only insignificantly, and they are nearly
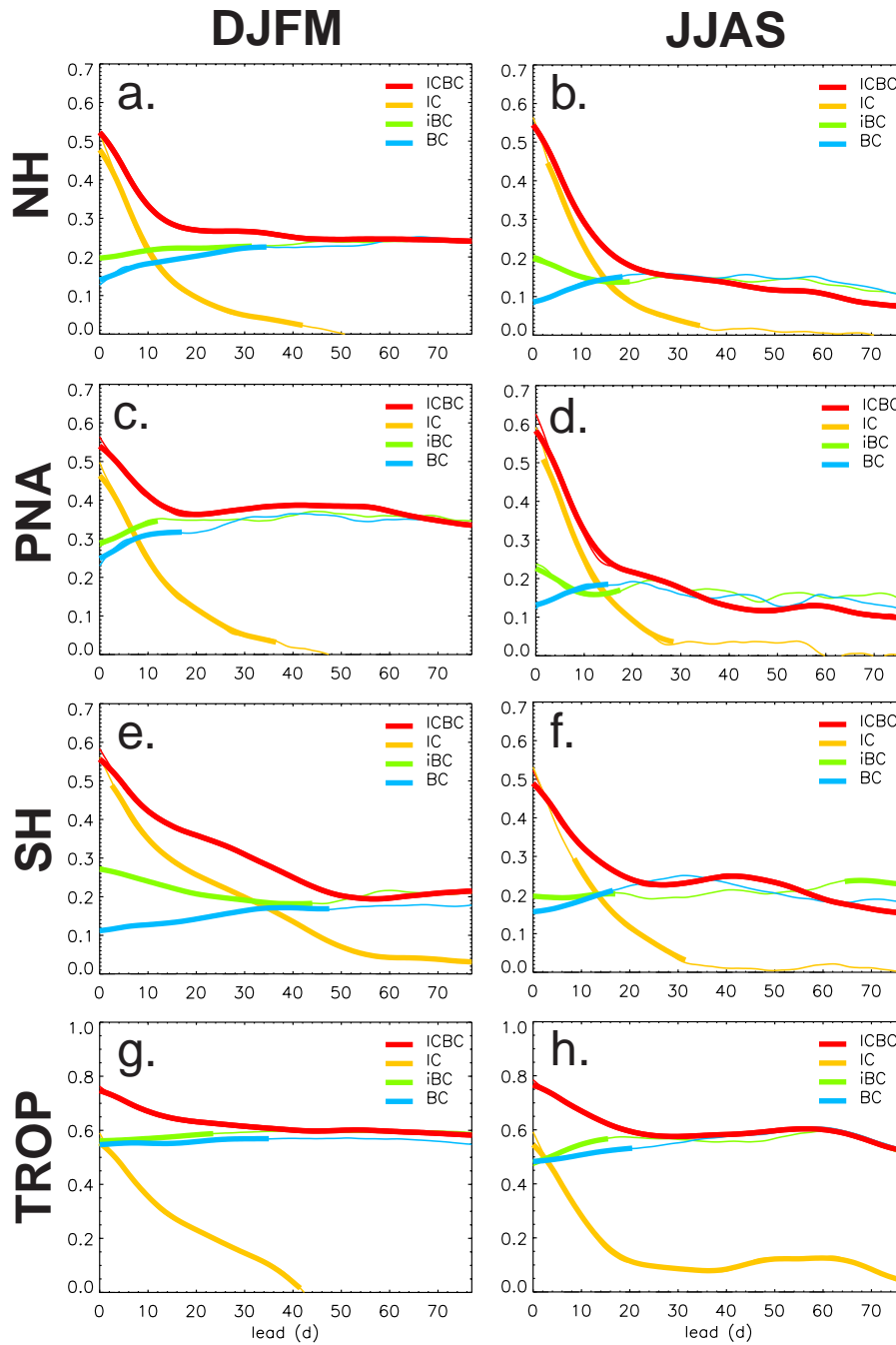
Fig. 7. Lag evolution of the spatial *AC* averaged over all years (1980–2000) from 30 days time averaged data. Panels **(a)**–**(b)** show the *AC* over the Northern Hemisphere (20° N–90° N), **(c)**–**(d)** for the PNA region (180° W–60° W, 15° N–70° N), **(e)**–**(f)** for the Southern Hemisphere (20° S–90° S), and **(g)**–**(h)** for the tropics (20° M–20° S). The *AC* has been calculated from the 500 hPa surface except for the tropics, where the 200 hPa surface has been used. Left panels are from 15 December to 31 March, and right panels are from 15 June to 31 September. All experiments were verified against ICBC under the perfect model assumption.

constant in time. This implies that the effect of initial conditions is negligible in the presence of strong boundary forcing. Again, the opposite is true during neutral years: The skill of ICBC decreases considerably with lead time. The skill of iBC and BC, which start from inferior initial conditions, is significantly lower than that of ICBC. Comparing the skill of BC and IC for neutral years (Fig. 6g) shows that even for

seasonal averages, the effect of initial conditions is as important as boundary conditions at zero lead time. The various time scales derived for all years are generally shorter during ENSO years, and longer during neutral years. For example, the crossover between BC and IC advances from roughly 20 days during ENSO years, to roughly 30 days during all years. This is consistent with initial conditions becoming less im-

portant with increasing strength of boundary forcing.

5.4 Regional and seasonal variations

In the previous section, forecast skill was analyzed over the Northern Hemisphere during DJFM. How do these results vary for different regions and seasons? Figure 7 compares the time evolution of 30 day time averaged skill (1980–2000) for the Northern Hemisphere, the PNA region, the Southern Hemisphere, and the tropics. The left column is for DJFM, and the right column is for JJAS.

Consider first the DJFM season. As expected, the overall skill over the PNA region (Fig. 7c) is higher than for the whole Northern Hemisphere. Skill values of the three boundary forced experiments level off between 0.3 and 0.4, as compared to 0.2 to 0.3 for the whole hemisphere. The heightened sensitivity of the PNA region to boundary forcing is illustrated by the shorter timescales for the crossover of IC and BC, and for BC reaching the same level of skill as ICBC. The ACs for the unforced experiment IC are almost identical for the PNA region and the Northern Hemisphere. Over the Southern Hemisphere (Fig. 7e), the influence of initial conditions is most important. This is evident from the slow decrease in skill of ICBC and IC. Note that the skill of IC is at all lead times significantly larger than zero. The aforementioned three periods with nearly linear decrease in skill can also be seen very clearly in Fig. 7e. The crossover between IC and BC for the 30 days time averaged data is reached after 30–40 days, and it takes almost 50 days for BC to reach the same skill as ICBC. The final skill of ICBC over the Southern Hemisphere is somewhat smaller than over the Northern Hemisphere. The opposite is the case during JJAS (Figs. 7b and 7f). This may be related to seasonal variations in the strengths of the subtropical jets. Strong westerly flows are very effective for Rossby wave forcing of the midlatitude wave train in relationship with ENSO (e.g. Sardeshmukh and Hoskins, 1988; Held and Kang, 1987). Over the tropics (Fig. 7g), the overall skill is, as expected, very high. However, initial conditions still play a role, as can be seen by the slow approach of BC to ICBC, and the relative slow decrease of IC to zero skill.

During JJAS, the skill over the Northern Hemisphere (Fig. 7b) and over the PNA region (Fig. 7d) are very similar and much smaller than during DJFM. This is consistent with a weaker subtropical jet and weaker tropical forcing during this season. From the delayed crossover between IC and BC it must be concluded that initial conditions are somewhat more important during JJAS than during DJFM. Over the Southern Hemisphere (Fig. 7f), the final skill of ICBC is larger than over the North. Interestingly, there is no such strong influence of initial conditions as during DJFM. Perhaps, the strong baroclinicity during Southern Hemispheric winter decreases predictability, although there could be other reasons as well. Over the tropics (Fig. 7h), the differences between the two seasons are small. Interestingly, IC shows a very slow approach to zero skill, which indicates that even over the tropics initial conditions must not be neglected.

Table 3. Signal to noise ratios s of experiment ICBC during DJFM for different regions, averaging periods and years. First number is for neutral years, second number is for all years, and third number is for ENSO years

	NH	PNA	SH	TROP
12 h	0.2 0.3 0.3	0.3 0.4 0.5	0.3 0.3 0.3	0.3 0.6 0.9
30 d	0.4 0.5 0.6	0.4 0.7 0.9	0.5 0.6 0.7	0.7 1.0 1.5
90 d	0.5 0.7 0.8	0.4 0.9 1.4	0.7 0.8 0.9	0.7 1.2 2.3

Table 3 shows the signal to noise ratio s calculated after Eq. (16) for simulation ICBC for different regions and years. The values were determined after error saturation of the initial 30 days. Generally, the signal to noise ratio and thus the potential predictability increase with increasing averaging period. According to the importance of boundary forcing, s is largest during ENSO years, over the PNA region, and the tropics, where s values greater than 1 (i.e. signal exceeds noise) can be found. In a previous section, the lower limit of s determined from simulation CC was found to be 0.25 at maximum lead (Fig. 3). Corresponding values from ICBC exceed this limit during all years and over all regions, indicating that even weak boundary forcing and anomalous initial conditions contribute to a higher signal.

5.5 Real world skill

So far, all predictability results were performed under perfect model assumptions, meaning that model output was used for the verification. The advantage of this method was that model related errors did not negatively affect the sometimes very subtle signals. Moreover, various ensemble members could be used for the verification, which greatly reduced sampling problems. However, can similar signals be seen in actual forecast experiments verified against standard reanalysis? To this purpose, experiment ICBC-r was conducted for the DJFM season. Again, each run was initialized by breeding from NCEP/DOE reanalysis-2. To bring the simulations closer to real observations, daily fields of reanalysis-2 soil moisture and snow cover were also prescribed.

The red curves in Fig. 8 present the results for ICBC-r for the same four regions as in the previous section. 30 days averages are shown in the left column, and 90 day averages in the right column. The basic features between real world skill (ICBC-r) and perfect model skill (ICBC) are the same. The major differences of ICBC-r are: (a) the results are more noisy, since only one member was used for verification, and (b) the overall skill is lower because of the introduction of model errors. More robust results would have been achieved by larger ensembles and longer simulation periods. To test the effect of initial conditions on the real world forecast, Fig. 8 shows the skill of experiment iBC and BC verified against reanalysis. The skill of ICBC-r is almost everywhere larger than iBC or BC. Statistically, these differences are not
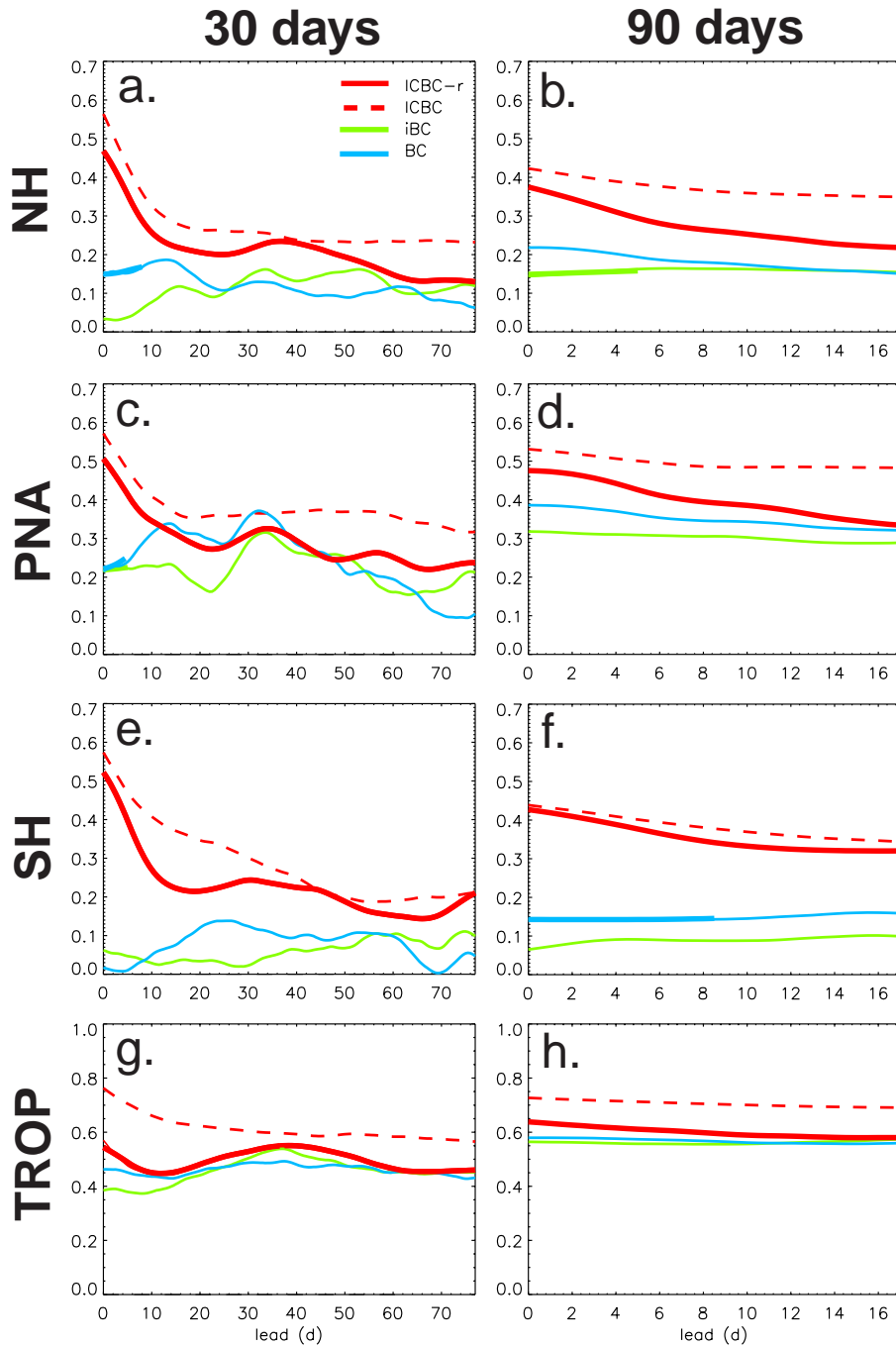
Fig. 8. As Fig. 7 but for experiment ICBC-r verified against NCEP/NCAR reanalysis. For DJFM, and for 30 and 90 days averaged data. Dashed red lines show for comparison the skill of ICBC under the perfect model assumption.

always significant, as shown by the very short thick lines. This, however, is mostly related to the higher noise level in the data. The 90 days averaged data on the right column show similar but smoother results. It must be concluded that also for a real world forecast initial conditions are important on sub-seasonal to seasonal time scales. Note that for the Southern Hemisphere, the skill of ICBC is much higher than that of iBC or BC, as it has been the case in the previous section for the perfect model verification. Over the tropics, the influ-

ence of initial conditions seems to be least important. This is in contradiction with the perfect model verification, and may indicate that (a) the quality of observed initial conditions over the tropics is not very good, or that (b) the model's response to anomalous tropical initial conditions, e.g. the intraseasonal oscillation, is not realistic enough.

A good estimate of the effects of model errors on skill can be achieved by comparing ICBC-r verified against reanalysis with ICBC verified against itself. The skill of ICBC un-
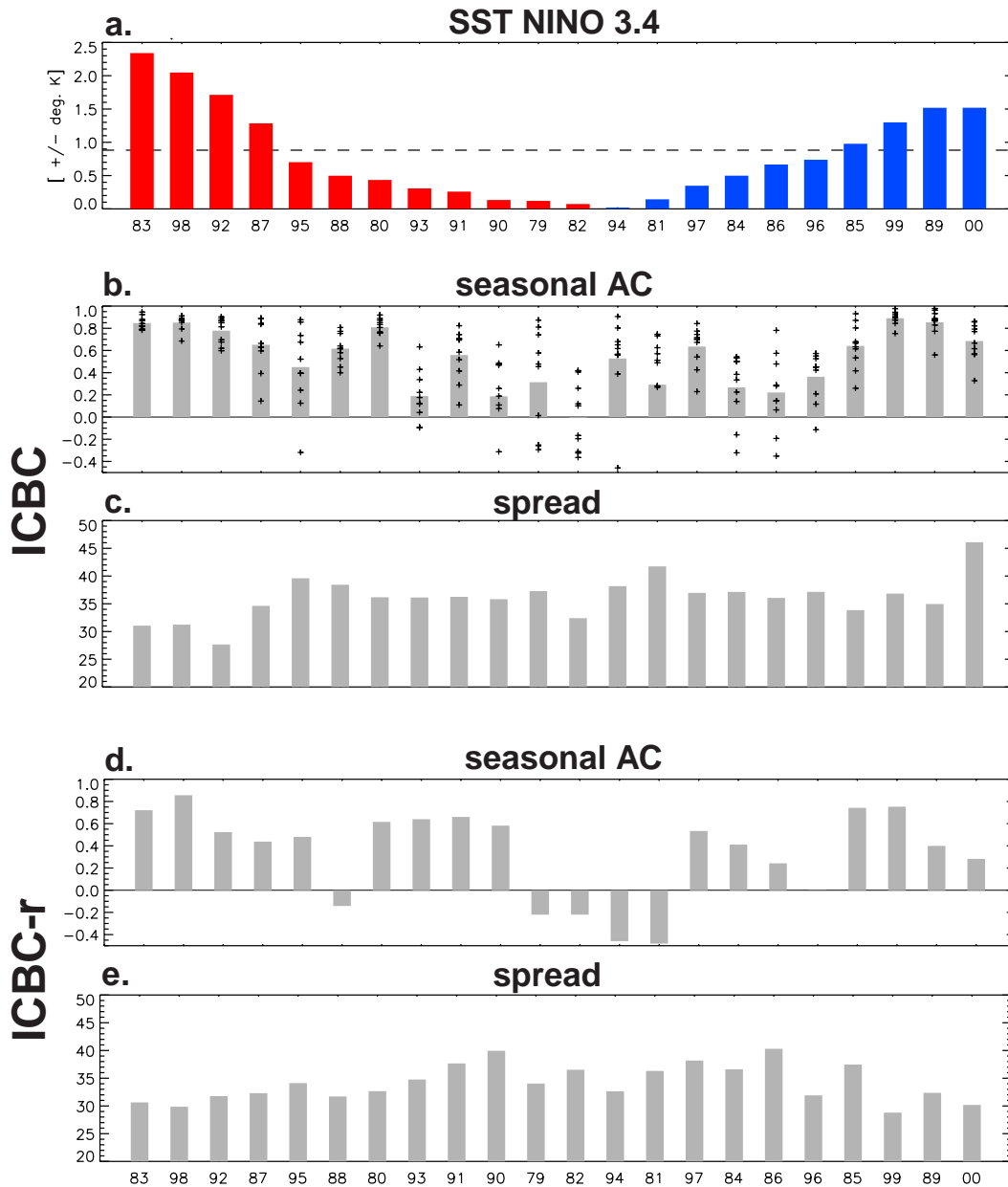
Fig. 9. (a) JFM mean SST anomalies over the Niño 3.4 region, ranked according to the strength of the SST anomalies from warm events (red) to cold events (blue). **(b)** JFM seasonal mean AC from 500 hPa heights over the PNA region for ICBC, verified under the perfect model assumption. Crosses indicate ACs from individual verification members, and columns denote the mean over all ten verification members. **(c)** Spread among ensemble members for ICBC (units in m). **(d)** Seasonal AC for ICBC-r verified against reanalysis. (*e*) Spread for ICBC-r.

der perfect model verification is shown in Fig. 8 by the red dashed curves. As expected, the perfect model skill is in all cases higher than the real world skill. From the seasonally averaged skill one can see that the difference between perfect and real world skill tends to increase with time, owing to the model drift toward an erroneous state when initialized with reanalysis. Over the PNA region, for example, the seasonal forecast skill at zero lead for the real world is 0.38, which is rather close to 0.42 of the perfect world. The difference of 0.04 increases to about 0.15 at maximum lead time. Over the Southern Hemisphere, the difference between real and per-

fect world is only small, although this might be related to the sparse observations over this part of the world, which tend to make reanalysis data more similar to the model solution.

5.6 Interannual variations

Variations in the strength of ENSO related boundary forcing are the main reason why some years are better predictable than others. To some extent, variations in skill are also random, caused by the internal variability of the model. This inherent uncertainty is noticeable on all time

scales, even on seasons or longer, and can be ameliorated by the ensemble approach. It may also be that initial conditions influence interannual variations in predictability. As was discussed previously, the atmosphere may be better predictable from some initial conditions than from others. Our results showed that this assumption may even hold on seasonal time scales, where there were significant differences in predictability merely because of differences in initial conditions. Initial conditions that are associated with atmospheric persistence are probably more likely to increase predictability. For example, initial conditions may contain or develop into stable blocking situations, which in turn lead to higher predictability, given the ability of the GCM to reproduce this event correctly. It may also be that stable natural eigenmodes of the atmosphere are contained in initial conditions, which then persist through much of the forecast.

In Fig. 9 the strength of the tropical ENSO forcing is compared with the interannual variations of seasonal forecast skill over the PNA region during January, February and March (JFM). For the seasonally averaged skill, the sampling uncertainty of individual years is sufficiently small, so that stable results are achieved without taking composites over many years. Fig. 9a depicts the strength of the ENSO signal for each winter of the period 1979–2000. This signal was calculated from the magnitudes of SST anomalies over the Niño 3.4 region during JFM relative to the climatology 1950–2000. Individual years were arranged according to the strength of the SST anomalies, with the warmest years in red on the left-hand side, and the coldest years in blue on the right-hand side. The graphs below show the corresponding ACs for simulation ICBC verified under the perfect world assumption (Fig. 9b), and for ICBC-r verified against NCEP/NCAR reanalysis (Fig. 9d). The ACs were calculated from seasonally averaged (JFM, i.e. day 17–106) 500 hPa heights over the PNA region, and arranged according to the ranked SST anomalies. Crosses show individual results for each of the ten verifying members, and the gray bars denote the mean over all 10 verification members.

ICBC (Fig. 9b) shows considerable year-to-year variability in skill, but in every year the skill is positive. There is a fairly good relationship between the strength of the ENSO forcing and seasonal forecast skill; larger ACs occur at the two extremes of the graphs than for neutral years at the middle. The correlation between the Fisher z-transformed ACs and the magnitude of the SST anomaly over the Niño 3.4 region is 0.76. There are also exceptions: During the winter 2000, which was a strong La Niña event, ICBC has relatively modest skill. In contrast, during 1980, ICBC shows surprisingly good skill compared to the weakness of the tropical forcing.

The skill of the real world forecast ICBC-r (Fig. 9d) is in most cases positive. Exemptions are 1979, 1982, 1994, and 1981, which were all neutral ENSO years. Two years show very negative skill. This is probably a consequence of ensemble averaging, which makes poor skill of individual forecasts even worse (e.g. Brankovic et al. 1990). The skill of ICBC-r is clearly lower than that of ICBC. The mean over all years

for ICBC-r is 0.33 as compared to 0.53 for ICBC. This difference must be mainly due to model errors, since boundary and initial conditions were in both forecasts similar. ICBC-r has also the disadvantage of a larger sampling uncertainty, since only one member – the real atmosphere – was available for the verification. The correlation between skill and ENSO forcing for ICBC-r is 0.56, which is also smaller than for ICBC. It is interesting to compare the real world skill of this model with that of other models which participated with the DSP project. The mean AC of all DSP models over the PNA region and over the years 1983–1993 was 0.47 (Shukla et al., 2000), which is slightly smaller than 0.48 for this model. This means that the performance of this GCM is comparable to the average of all DSP models.

There is also a modest relationship between spread SP and skill AC. Such a relationship would mean that one could have more confidence when ensembles have low spread (Anderson et al., 1999). The spread among ensemble members for the two simulations from 500 hPa heights over the PNA region are shown in Figs. 9c and 9e. The correlation is -0.27 for ICBC, and -0.12 for ICBC-r. Although a modest relationship is indicated, these levels are statistically insignificant, given the small number of years.

5.7 Scale variations

In the previous section, three distinctive regimes of error growth were briefly discussed: An initial fast error growth during the first 20 days or so, an intermediate moderate error growth from days 20–40, and error saturation thereafter. It was speculated that this could be related to the scale dependence of error growth. Intuitively, it is clear that smaller scales have a faster error growth than larger scales, simply because smaller length scales are more sensitive to deformation by advective errors. Earlier studies of predictability confirm that errors in prediction of the smallest scales grow the fastest (e.g. Dalcher and Kalnay, 1987; Roads, 1987). In fact, Lorenz (1969) proposed that errors in the smallest scales gradually propagate to larger scales until the whole spectrum is saturated. From the analysis of our experiments, however, errors do not seem to fully saturate at all length scales.

Figure 10 shows relative error energy S_{nm}^2 as a function of lead time and 2-dimensional wavenumber for the three experiments ICBC, IC and BC. Note that $S_{n(-m)}^2 = S_{nm}^2$, so that two-sided spectral plots ($-n < m < n$) are redundant, and one-sided plots ($0 < m < n$) contain all necessary information. The quantity shown is time averaged instantaneous error energy scaled by the maximum saturation variance. It was calculated using Eq. (11). All data were taken from daily global 500 hPa fields. As for the ACs, ensemble mean forecasts were verified against individual members of the verifying experiment ICBC. The spectra for ICBC (Fig. 10, top row) show that small scale errors grow faster than larger ones. During days 20–40, errors for wavenumbers greater than 10 are fully saturated. During days 40–100, only a few specific large wavenumbers are not completely saturated. These scales are maintained by boundary forcing,
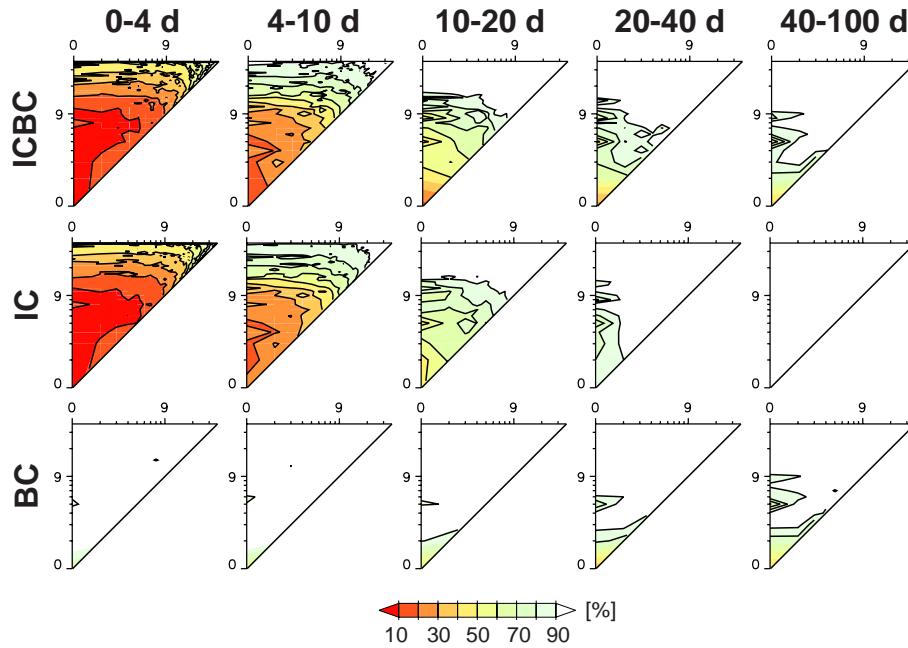
Fig. 10. 2-dimensional error spectra of global 500 hPa fields for zonal wavenumber m (x-axis) and total wavenumber n (y-axis) at different lags. Shown are time averages of instantaneous error spectra over different averaging periods. Ensemble means of each simulation were verified against individual members of ICBC. Errors are expressed relative to the theoretical maximum error variance.

and they must be responsible for the non-zero instantaneous skill at long leads. These wavenumbers are mainly located along zonal wavenumber zero, which means that error energy is quite homogeneously distributed in the zonal direction. Apart from wave number zero (global mean), there is a secondary maximum for wavenumber 4,0 (n, m). The characteristic bow shape of the isolines indicates that for a given total wavenumber, errors for higher zonal wavenumbers grow faster than for lower ones. This is probably related to the fact that zonal wind speeds are on average higher than meridional ones, so that error energy spreads zonally faster than meridionally.

The spectra of simulation IC (Fig. 10, middle row) during days 0–4 and 4–10 are very similar to ICBC, which indicates that effects from the boundaries are not very important at this early stage. From day 10 on, errors for IC grow much faster than for ICBC. After day 40, the errors for IC are fully saturated at all scales. This is consistent with IC having zero skill after day 40. During days 20–40, ICBC has much better skill than from initial conditions (simulation IC) or boundary conditions (simulation BC) alone. This means that the combined effect of boundary forcing and initial conditions is particularly important for good skill at this lag. The error spectra for simulation BC show a progression from larger to smaller scales. Boundary effects are first noticeable in the largest scales, and then gradually affect smaller scales. At the maximum lead time interval (days 40–100), the spectra of ICBC and BC are almost identical. This is consistent with ICBC and BC having the same skill, and initial conditions information being completely lost.

A one-dimensional representation of the error energy pro-

vides additional insight into the scale dependence of error growth. In Fig. 11 errors are shown for every time step, but only for total wave number n. During the first 20 days (regime I), simulation ICBC shows rapid error growth at scales with $n > 5$. Between day 10 and 20, scales with $n < 5$ also contribute to some error growth. Let us define scales with $n > 5$ as "small", and scales with $n < 5$ as "large". After day 20 (regime II), when the small scales are completely saturated, the error growth in the large scales continues at its own slower rate. This explains the slow decrease in skill after day 20 noted before. Mainly wavenumbers 2, 3, and 5 contribute to error growth at this stage. After day 40 or so (regime III), the error energy for ICBC and BC remains almost constant in time.

Small scales contribute to most of the error growth at shorter lead times. From band pass filtered (2–5 days) wave number spectra (not shown) it is found that the synoptic energy maximum is located at $n = 11$, so that the "small" scales are mostly associated with synoptic and smaller scales. The "large" planetary waves, on the other hand, have a slower error growth rate, and do not saturate completely at long lead times. This can also be shown by replacing time averaging, which filters out unpredictable small scale structures, with a more direct spectral filtering. This method has the advantage that the evolution of skill can be calculated with better temporal resolution. The data were spectrally filtered by keeping either only large scales, all scales, or small scales. Figure 12 shows the effects of spatial filtering on instantaneous global forecast skill from ICBC, IC, iBC and BC. Skill from large scale filtered data is much higher than from data with all or only the small scales. For simulation ICBC, the large scale
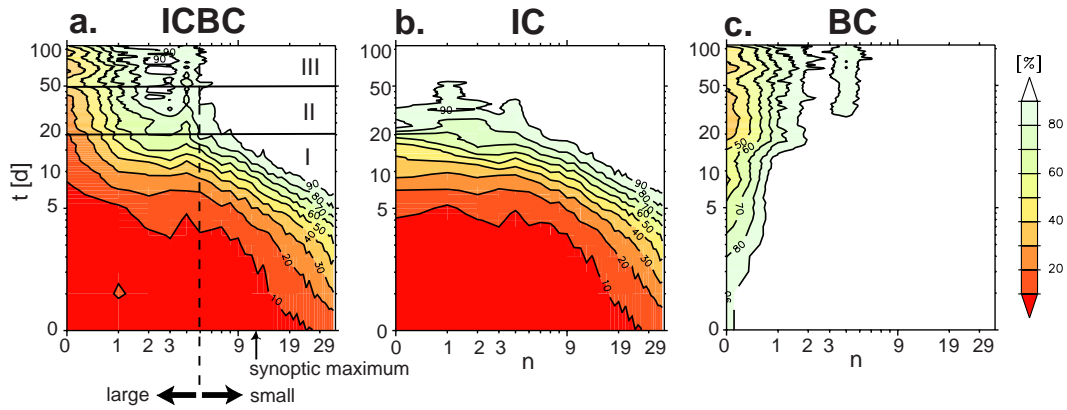
Fig. 11. Error energy spectrum as a function of total wavenumber n and lead time t for simulation ICBC **(a)**, IC **(b)** and BC **(c)**. 500 hPa height ensemble mean fields of each simulation were verified against individual members of ICBC. Error energy is expressed in percent of the maximum saturation error.
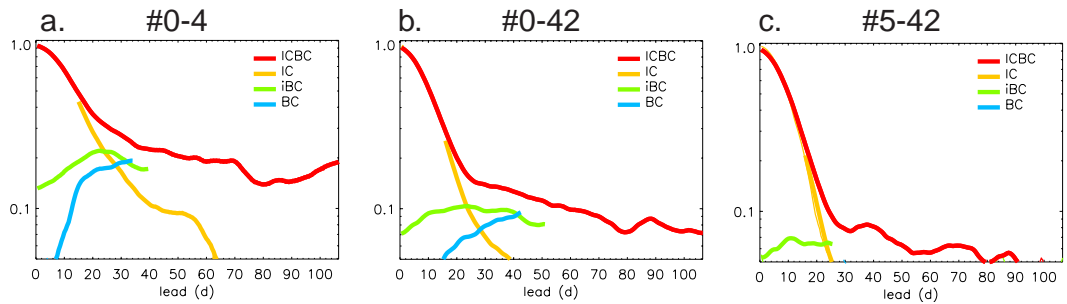

Fig. 12. Lag evolution of global instantaneous AC of simulation ICBC (red), IC (yellow), iBC (green), and BC (blue) for all years. Only wavenumbers 1–4 **(a)**, 0–42 **(b)**, or 5–42 **(c)** were retained. Skill is only shown where significantly different from ICBC.

correlations level off between 0.3 and 0.2, as compared to 0.1 for all scales, and 0.05 for the small scales. Note that the small scales show some measurable instantaneous skill at long lead times. This suggests that the influence of boundary forcing is not completely confined to the largest scales. The size of instantaneous large scale correlations is comparable to 30 day time averaged correlation using all scales. Simulation IC (yellow curves) shows that the memory of the initial conditions is short for the small scales and long for the large scales. For small scales, skill from initial conditions alone is completely lost after 20 days, but for the large scales it lasts for almost two months. This confirms again that the long term memory of the initial conditions resides in the largest scales.

The initial loss in predictability of the large scale filtered data for ICBC is much slower than that of the unfiltered data. The difference in skill between large scale filtered and unfiltered data is shown in Fig. 13 for three composite periods. The difference grows rapidly at the beginning, owing to the fast error growth of the small scales. After 20 days or so, the difference remains almost constant in time. Small scale errors are saturated now, and further error growth is only due to larger scales. The gain in skill by large scale filtering is bigger during ENSO years than during neutral years, indicating that ENSO forcing is mainly affecting large scales.

6 Summary and conclusions

The goal of this study was to determine the importance of initial and boundary conditions for atmospheric predictability on all time scales out to a season. The NCEP global spectral model was used to conduct a series of seasonal predictability experiments, which were forced with different combinations of boundary and initial conditions. Initial conditions for the ensemble experiments were perturbed using the breeding method. The 22 year long simulation period covered both the cold and warm season. Predictability was examined over several target regions using the perfect model assumption as well as reanalysis.

First, predictability was investigated from purely climatologically forced runs. Initial conditions, together with their instabilities and nonlinear interactions, remember themselves for 50 to 60 days. Interestingly, anomalous initial conditions lead to higher predictability than climatological initial conditions. It must therefore be concluded that excited low frequency modes of the atmosphere have a higher persistence than unexcited modes. The effects of boundary forcing are detectable after about 10 days and lead to instantaneous skill of 0.1 at maximum lead time over the Northern Hemisphere. This small amount of instantaneous skill increases to about 0.35 for seasonally averaged skill over the same re-
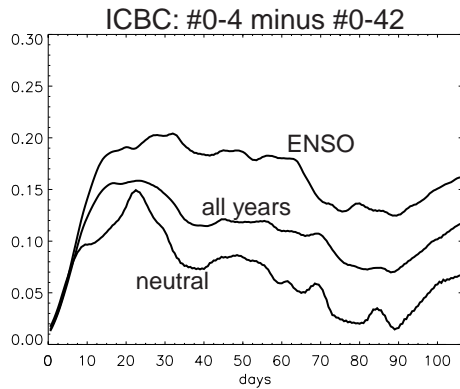
Fig. 13. Difference in global instantaneous skill of ICBC between large scale filtered and unfiltered data for ENSO years, all years (1980–2000), and neutral years.

Table 4. Time scale for influence of initial conditions measured by the intersection between the skill curves from experiments IC and BC from instantaneous data

	NH	PNA	SH	TROP
DJFM	27	24	46	23
JJAS	28	26	28	22

gion. From a detailed analysis of 30 days averaged skill significant regional and seasonal differences in skill were found. Under the same boundary forcing, true initial conditions led in all cases to better forecasts than random climatological or AMIP-type initial conditions. The improvement was detectable out to 5–6 weeks, and cannot be explained alone from the effects of high deterministic predictability during the first 2–3 weeks. This means that initial conditions also improve predictability for the low-frequency large scale components of the flow. The combined effect of boundary and initial conditions was particularly important for forecasts on a monthly lead time.

The real world verification of the simulations led to the same conclusion. In this case, good initial conditions were particularly important at short leads, when the impact of model errors are still relatively small. From the difference between perfect and real world skill the effect of model errors on predictability could be discussed. The difference generally increases in time. At maximum lead, it is about 0.15 correlation of seasonally averaged AC over the Northern Hemisphere during DJFM. This clearly demonstrates the large potential for better forecasts if models could be improved.

A detailed year to year breakdown of seasonally averaged forecast skill revealed a good relationship between the strength of ENSO forcing and skill. The correlation between both is around 0.8. Unfortunately, the relationship between spread and skill is quite weak. Spectral error analysis showed that large planetary scales have a much slower error growth rate than small length scales. Basically, the long term effect of boundary forcing is to maintain predictability for a few specific spatial wavenumbers. After day 40, total wavenumbers 0 and 4 contribute to most of the predictability, while error energy for wavenumbers larger than 9 is almost completely saturated after that time.

How important initial conditions on the longer time scales are depends on factors such as lead time, averaging period, region, level, season, verification method, variable, and how susceptible the specific atmospheric state is. The role of initial conditions also varies with the strength of boundary forc-

ing. During ENSO years, the effect of initial conditions is smaller than during neutral years. Table 4 summarizes our findings with a simple quantity, τ, that measures the importance of initial conditions. τ is the time scale at which the skill from the unforced simulation IC, which starts from good initial conditions, intersects with that from the forced simulation BC, which comes from random initial conditions. Thus, it represents the time at which boundary forcing is equally important as initial conditions. τ was determined for different regions and seasons, using the perfect model skill of unaveraged data over all years.

Over the Northern Hemisphere, it takes about 4 weeks until boundary forcing reaches the same significance as initial conditions. This 4 week time range highlights the special situation for forecasts on sub-seasonal time scales (2–6 weeks). On the one hand, initial conditions are important, but their influence is too small to give useful deterministic skill. On the other hand, effects from the boundaries are important, but they are hard to detect over the short 4 week time range. The consequence for monthly averaged forecasts is that the dominance of initial conditions is limited to the first 1–2 weeks, and for seasonal averages, τ is actually negative, meaning that boundary conditions are more important than initial conditions even at zero lead. Seasonal differences of τ over the Northern Hemisphere are small. During periods with strong ENSO forcing, τ reduces to about 3 weeks, meaning that boundary forcing becomes more important. Over the PNA region and the tropics, τ is somewhat shorter. Over these areas, boundary forcing is more important, but this does not mean that initial conditions can be neglected.

Over the Southern Hemisphere, there exists a remarkable difference in τ between summer and winter. During winter (JJAS), τ is in the 4 week time range, like over the Northern Hemisphere. During summer (DJFM), τ increases to 6–7 weeks, showing that the importance of initial conditions is dramatically increased. One might argue that this is an model artifact caused by the perfect model assumption, but the verification with reanalysis leads to very similar results. In fact, Trenberth (1985) found over Antarctica during summer indications for atmospheric persistence which is greatly in excess of that found over the Northern Hemisphere. The conditions which contribute to this long term memory may be related to the weak baroclinicity and the more zonally symmetric structure during the Southern Hemisphere summer. Also, major modes, like the Antarctic Circumpolar Wave (White and Peterson, 1996) and the Antarctic Oscillation (Thompson and

Wallace, 2000) may play a role. These aspects need to be addressed further.

From this study it is obvious that the contribution of initial conditions to predictability beyond 4 weeks is, on average, smaller than that from boundary conditions. However, we must conclude that the effects of initial conditions are not negligible, in particular when the boundary forcing is weak. The consequence for current sub-seasonal to seasonal forecasts is that good atmospheric initial conditions are needed.

References

Anderson, J., van den Dool, H., Barnston, A., Chen, W., Stern, W., and Ploshay, J.: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction, Bull. Amer. Meteorol. Soc., 80, 1349–1361, 1999.

Anderson, J. L. and Ploshay, J. J.: Impact of initial conditions on seasonal simulations with an atmospheric general circulation model, Quart. J. Roy. Meteorol. Soc., 126, 2241–2264, 2000.

Barker, T. W. and Horel, J. D.: The Impact of Climatology and Systematic Errors Upon the Skill of DERF Forecasts, Mon. Weather Rev., 117, 2835–2842, 1989.

Barnett, T. P.: Monte Carlo Climate Forecasting, J. Climate, 8, 1005–1022, 1995.

Barsugli, J. J., Whitaker, J. S., Loughe, A. F., Sardeshmukh, P. D., and Toth, Z.: The effect of the 1997/1998 El Niño on individual large-scale weather events, Bull. Amer. Meteorol. Soc., 80, 1399–1411, 1999.

Boer, G. J.: A spectral analysis of predictability and error in an operational forecast system, Mon. Wea. Rev., 112, 1183–1197, 1984.

Brankovic, C., Palmer, T. N., Molteni, F., Tibaldi, S., and Cubasch, U.: Extended-Range Predictions With ECMWF Models – Time-Lagged Ensemble Forecasting, Quart. J. Roy. Meteorol. Soc., 116, 867–912, 1990.

Buizza, R. and Palmer, T. N.: The Singular-Vector Structure of the Atmospheric Global Circulation, J. Atmos. Sci., 52, 1434–1456, 1995.

Buizza, R.: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system, Mon. Weather Rev., 125, 99–119, 1997.

Caplan, P., Derber, J., Gemmill, W., Hong, S. Y., Pan, H. L., and Parrish, D.: Changes to the 1995 NCEP operational medium-range forecast model analysis-forecast system, Weather and Forecasting, 12, 581–594, 1997.

Chang, Y., Schubert, S. D., and Suarez, M. J. Boreal winter predictions with the GEOS-2 GCM: The role of boundary forcing and initial conditions, Quart. J. Roy. Meteorol. Soc., 126, 2293–2321, 2000.

Chou, M.-D. and Suarez, M. J.: An efficient thermal infrared radiation parameterization for use in general circulation models, NASA Tech. Memo. 104606 [NTIS N95-15745.], 85 pp., 1994.

Chou, M. D., Lee, K. T., Tsay, S. C., and Fu, Q.: Parameterization for cloud longwave scattering for use in atmospheric models. J. Climate, 12, 159–169, 1999.

Dalcher, A. and Kalnay, E.: Error growth and predictability in operational ECMWF forecasts. Tellus, 39A, 474–491, 1987.

Dirmeyer, P. A.: Using a global soil wetness dataset to improve seasonal climate simulation, J. Climate, 13, 2900–2922, 2000.

Fels, S. B. and Schwarzkopf, M. D.: The simplified exchange approximation: A new method for radiative transfer calculations, J. Atmos. Sci., 32, 1475–1488, 1975.

Fennessy, M. J. and Shukla, J.: Impact of initial soil wetness on seasonal atmospheric prediction, J. Climate, 12, 3167–3180, 1999.

Grell, G. A.: Prognostic Evaluation of Assumptions Used By Cumulus Parameterizations, Mon. Weather Rev., 121, 764–787, 1993.

Held, I. M. and Kang, I. S.: Barotropic models of the extratropical response to El Niño, J. Atmos. Sci., 44, 3576–3586, 1987.

Hoffman, R. and Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting. Tellus, 35A, 100–118, 1983.

Hong, S. Y. and Pan, H. L.: Impact of soil moisture anomalies on seasonal, summertime circulation over North America in a Regional Climate Model, J. Geophys. Res.-Atmos., 105, 29 625–29 634, 2000.

Hong, S. Y.: Sensitivity of regional climate simulation to land-surface model, 3rd annual RSM conference, Taipei, Taiwan, 2001.

Hoskins, B. J. and Simmons, A. J.: A multi-layer spectral model and the semi-implicit method, Quart. J. Roy. Meteor. Soc., 101, 637–655, 1975.

Kalnay, E., Kanamitsu, M., and Baker, W. E.: Global Numerical Weather Prediction At the National-Meteorological-Center, Bull. Amer. Meteorol. Soc., 71, 1410–1428, 1990.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., et al.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Amer. Meteorol. Soc., 77, 437–471, 1996.

Kanamitsu, M.: Description of the NMC global data assimilation and forecast system, Weather and Forecasting, 4, 335–342, 1989.

Kanamitsu, M., Mo, K. C., and Kalnay, E.: Annual Cycle Integration of the NMC Medium-Range Forecasting (MRF) Model, Mon. Weather Rev., 118, 2543–2567, 1990.

Kanamitsu, M., Alpert, J. C., Campana, K. A., et al.: Recent changes implemented into the global forecast system at NMC, Weather and Forecasting, 6, 425–435, 1991.

Kanamitsu, M., Ebisuzaki, W., Woolen, J., Potter, J., and Fiorino, M.: An overview of NCEP/DOE reanalysis-2, 2nd international conference on reanalysis, Wokefield Park, United Kingdom, 1999.

Kanamitsu, M., Kumar, A., Schemm, J., et al.: NCEP dynamical seasonal forecast system 2000, Bull. Amer. Meteorol. Soc., (submitted), 2002.

Kistler, R., Kalnay, E., Collins, S., Saha, S., et al.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bull. Amer. Meteorol. Soc., 82, 247–267, 2001.

Kumar, A. and Hoerling, M. P.: Prospects and Limitations of Seasonal Atmospheric Gcm Predictions, Bull. Amer. Meteorol. Soc.,

76, 335–345, 1995.

Kumar, A. and Hoerling, M. P.: Annual cycle of Pacific North American seasonal predictability associated with different phases of ENSO, J. Climate, 11, 3295–3308, 1998.

Kumar, A. and Hoerling, M. P.: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction, Bull. Amer. Meteorol. Soc., 81, 255–264, 2000.

Kumar, A., Barnston, A. G., and Hoerling, M. P.: Seasonal predictions, probabilistic verifications, and ensemble size, J. Climate, 14, 1671–1676, 2001.

Landman, W. A. and Mason, S. J.: Change in the association between Indian Ocean sea-surface temperatures and summer rainfall over South Africa and Namibia, Int. J. Climatol., 19, 1477–1492, 1999.

Lau, N. C. and Nath, M. J.: A Modeling Study of the Relative Roles of Tropical and Extratropical Sst Anomalies in the Variability of the Global Atmosphere Ocean System, J. Climate, 7, 1184–1207, 1994.

Leith, C. E.: Theoretical Skill of Monte Carlo Forecasts, Monthly Weather Rev., 102, 409–418, 1974.

Lorenz, E.: A study of the predictability of a 28-variable atmospheric model, Tellus, 17, 321–333, 1965.

Lorenz, E.: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289–307, 1969.

Mason, S. J., Goddard, L., Graham, N. E., Yulaeva, E., Sun, L. Q., and Arkin, P. A.: The IRI seasonal climate prediction system and the 1997/1998 El Niño event, Bull. Amer. Meteorol. Soc., 80, 1853–1873, 1999.

Moorthi, S. and Suarez, M. J.: Relaxed Arakawa-Schubert – a Parameterization of Moist Convection For General Circulation Models, Mon. Weather Rev., 120, 978–1002, 1992.

NMC-Development-Division: Documentation of the research version of the NMC medium range forecasting model, 504 pp., 1988.

Palmer, T. N., Brankovic, C., Molteni, F., Tibaldi, S., Ferranti, L., Hollingsworth, A., Cubasch, J., and Klinker, E.: The European-Centre-For-Medium-Range-Weather-Forecasts (ECMWF) Program On Extended-Range Prediction, Bull. Amer. Meteorol. Soc., 71, 1317–1330, 1990.

Palmer, T. N.: Extended-Range Atmospheric Prediction and the Lorenz Model, Bull. Amer. Meteorol. Soc., 74, 49–65, 1993.

Palmer, T. N., Brankovic, C., and Richardson, D. S.: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations, Quart. J. Roy. Meteorol. Soc., 126, 2013–2033, 2000.

Pan, H. and Mahrt, L.: Interaction between soil hydrology and boundary layer developments, 38, 185–202, 1987.

Reynolds, R. W. and Smith, T. M.: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation, J Climate, 7, 929–948, 1994.

Roads, J. O.: Forecasts of Time Averages with a Numerical Weather Prediction Model, J. Atmos. Sci., 43, 871–892, 1986.

Roads, J. O.: Predictability in the extended range, J. Atmos. Sci., 44, 3495–3527, 1987.

Roads, J. O.: Lagged average predictions in a predictability experiment, J. Atmos. Sci., 45, 147–162, 1988.

Roads, J. O., Chen, S. C., and Fujioka, F.: ECPC's weekly to seasonal global forecasts, Bull. Amer. Meteorol. Soc., 82, 639–658, 2001.

Robinson, W.: Review of WETSThe Workshop on ExtraTropical SST anomalies, Bull. Amer. Meteorol. Soc., 81, 137–144, 2000.

Rowell, D. P.: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations, J. Climate, 11, 109–120, 1998.

Sardeshmukh, P. D. and Hoskins, B. J.: The generation of global rotational flow by steady idealized tropical divergence, J. Atmos. Sci., 45, 1228–1251, 1988.

Sardeshmukh, P. D., Compo, G. P., and Penland, C. C.: Changes of probability associated with El Niño, J. Climate, 13, 4268–4286, 2000.

Shukla, J.: Dynamical predictability of monthly means, J. Atmos. Sci., 38, 2547–2572, 1981.

Shukla, J.: Predictability in the midst of chaos: A scientific basis for climate forecasting, Science, 282, 728–731, 1998.

Shukla, J., Anderson, J., Baumhefner, D., et al.: Dynamical seasonal prediction, Bull. Amer. Meteorol. Soc., 81, 2593–2606, 2000.

Slingo, A. and Slingo, J. M.: Response of the National Center For Atmospheric Research Community Climate Model to Improvements in the Representation of Clouds, J. Geophys. Res.-Atmos., 96, 15 341–15 357, 1991.

Thompson, D. W. J. and Wallace, J. M.: Annular modes in the extratropical circulation. Part I: Month-to-month variability, J. Climate, 13, 1000–1016, 2000.

Toth, Z. and Kalnay, E.: Ensemble Forecasting at NMC – the Generation of Perturbations, Bull. Amer. Meteorol. Soc., 74, 2317–2330, 1993.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP and the breeding method, Mon. Weather Rev., 125, 3297–3319, 1997.

Trenberth, K. E.: Persistence of daily geopotential heights over the Southern Hemisphere, Mon. Weather Rev., 113, 38–53, 1985.

Tribbia, J. and Baumhefner, D.: Estimates of the predictability of low-frequency variability with a spectral general circulation model, J. Atmos. Sci., 45, 2306–2317, 1988.

White, W. B. and Peterson, R. G.: An Antarctic Circumpolar Wave in Surface Pressure, Wind, Temperature and Sea-Ice Extent, Nature, 380, 699–702, 1996.

Wilks, D. S.: Statistical methods in the atmospheric sciences: an introduction, International Geophysics Series, v. 59., Academic Press, xi, 467 pp., 1995.