Climate Predictability on Interannual to Decadal Time Scales: The Initial Value Problem.

Matthew Collins[1,2]

March 18, 2002

[1]Centre for Global Atmospheric Research, Dept. of Meteorology,
University of Reading, Reading, RG6 6BB, UK.

[2]Hadley Centre for Climate Prediction and Research, The Met. Office, London Road, Bracknell, RG12 2SZ, UK.

Corresponding Author: Matthew Collins

Centre for Global Atmospheric Research, Dept. of Meteorology,

University of Reading, Reading, RG6 6BB, UK.

Tel: +44 118 9316018, Fax: +44 118 9318316, matcollins@met.rdg.ac.uk

Abstract

Any initial value forecast of climate will be subject to errors originating from poorly known initial conditions, model imperfections, and by “chaos” in the sense that, even if the initial conditions were perfectly known, infinitesimal errors can amplify and spoil the forecast at some lead time. Here the latter source of error is examined using a “perfect model” approach whereby small perturbations are made to a coupled Atmosphere-Ocean General Circulation Model and the spread of near-by model trajectories, on time and space scales appropriate to seasonal-decadal climate variability, is measured to assess the lead time at which the error saturates. The study therefore represents an estimate of the upper limit of the predictability of climate (appropriate to the initial value problem) given a perfect model and near perfect knowledge of the initial conditions.

It is found that, on average, surface air temperature anomalies are potentially predictability on seasonal to interannual time scales in the tropical regions and are potentially predictable on decadal time scales over the ocean in the North Atlantic. For mid-latitude surface air temperature anomalies over land, model trajectories rapidly diverge and there is little sign of any potential predictability on time scales greater than a season or so. For mean sea level pressure anomalies, there is potential predictability on seasonal time scales in the tropics, and for some global scale annual-decadal anomalies, although not those associated with the North Atlantic Oscillation. For precipitation, the only potential for predictability is for seasonal time anomalies associated with the El-Niño Southern Oscillation. For the majority of the highly populated regions of the world, climate predictability on interannual to decadal time scales based in the initial value approach is likely to be severely limited by chaotic error growth. It is found however that there can be cases in which the potential predictability can be higher than average indicating that there is perhaps some utility in making initial value forecasts of climate in those regions which show low predictability on average.

1 Introduction

There are two distinct classes of climate prediction problem: (i) the initial value problem of predicting the evolution of the climate system given some estimate of its current state and (ii) the boundary value problem of assessing a change in climate due to some “external” forcing. Lorenz (1975) called these predictability problems of the first and second kind respectively. In this paper we focus on the initial value problem.

While there is no universally excepted precise definition of what “climate” is, a useful working definition is that climate is some measure of the average effects of weather events over some specified time interval (months, seasons, decades, centuries etc.) over some geographical region (at a point, over some region which may define some climate index or political or socio-economic boundary, or some weighted average such as an EOF). A climate forecast then becomes a prediction of the deviation in some aspect of the probability density function (PDF) of weather events from normal; a simple example being the mean of the PDF but potentially extending to higher order moments and to extremes.

An example of the boundary value or second kind prediction problem is that of anthropogenically induced climate change. The interest then is in looking for some (predictable) change in the PDF of weather events induced by a change in radiative forcing of the Earth system due to an excess of greenhouse gases produced by man. An example of the initial value or first kind climate predictability problem is the forecasting of El Niño Southern Oscillation (ENSO - e.g. Latif et al (1994)). Then the interest is in examining the current (and past) state of the ocean and atmosphere in the tropical Pacific region and trying to assess dynamical imbalances which may lead to the development of cold or warm SST anomalies. These SST anomalies can modify the weather PDF both locally and in remote regions around the globe.

The focus of this study is the initial value climate predictability problem which has many similarities with the weather forecasting problem. In weather forecasting observations of the state of the atmosphere are collected and assimilated into a numerical model and a forecast is produced by integrating the model forward in time. The main cause of forecast inaccuracy arises via the sensitive dependence on initial conditions whereby trajectories which are initially close can diverge and the forecast error can saturate after some time (although model error can also be important Orrell et al (2001)). The climate system is subject to similar sensitive dependence on initial conditions and it is the purpose of this paper to examine the extent to which this limits the possibility of predicting variations in climate.

In the initial value climate prediction problem, the key prognostic variables are those aspects of climate which may (i) evolve on time scales which are long in comparison to the time scale of individual weather events and (ii) are such that they can modify the PDF of weather events in some predictable fashion. The "slow" development of anomalies in the ocean circulation and associated sea surface temperature (SST) field which can effect the fluxes of heat, moisture and momentum at the lower boundary of the atmosphere is an obvious candidate for such a prognostic variable. The ocean initial conditions, and observations which may be used to constrain them, are key. Other variables, such as sea ice, soil moisture or vegetation however may also evolve on suitably "slow" time scales to be of use in prediction.

The key issue then is the extent to which small errors in the initial conditions of a climate forecast can grow and at some point render the forecast useless. We can isolate this limiting factor from other potential sources of error (e.g. model inaccuracies) using a numerical coupled atmosphere-ocean global circulation model (AOGCM) of the climate system and performing "perfect model" or "perfect ensemble" experiments. In such experiments model generated initial conditions have small perturbations added to them and the potential climate predictability can be assessed by measuring the ensemble spread as a function of "lead time". Because of the likely important role of the ocean, we use a coupled ocean-atmosphere model (the third version of the Hadley Centre Coupled Model - HadCM3 Gordon et al (2000)). The limits to predictability derived in this study are thus (model estimates) of the upper limit to predictability based on having a perfect model and having near perfect knowledge of the current state of climate system (principally the state of the ocean). Both these goals will, of course, never be achieved in practice; an important point to bear in mind while reading the rest of this paper. It is also worth noting that this study is much closer to the "full" problem of potential predictability in contrast to studies which use atmospheric GCMs forced by observed variations in SST to examine potential predictability (e.g. Rowell (1998)).

The number of studies that have examined the potential predictability of climate on interannual to decadal time scales using comprehensive coupled ocean-atmosphere models are few, presumably due to the computer resources required to run ensembles of such experiments. Griffies and Bryan (1997a) and Griffies and Bryan (1997b) examined the predictability in the North Atlantic region. They found that during periods when there were large variations in the thermohaline circulation in the model, the leading Empirical Orthogonal Functions (EOFs) of ocean dynamic topography were potentially predictable out to time scales of 10-20 years and the leading EOF patterns of SST were predictable out to time scales of 5-7 years. During periods of low thermohaline circulation variability, lead times were reduced. In a similar study, Grötzner et al (1999) found potential predictability on the order of a decadal for variables related to thermohaline circulation but only of the order of a year or so for North Atlantic surface variables. Grötzner et al (1999) also claim potential predictability for ENSO out to 2-3 years. Boer (2000) used both diagnostic and prognostic approaches in assessing the predictability of the Canadian Centre Climate Model (CCCma). We adopt a similar diagnostic approach in section 3 of this paper in analysing HadCM3. On interannual time scales, Boer (2000) found little potential predictability of surface air temperature (SAT) over land but some potential predictability in the tropical Pacific, the Southern Ocean and in the extra-tropical North Pacific. On longer time scales SAT over some land areas, the tropical Pacific, the Southern Ocean and the tropical Atlantic were all found to be potential predictable.

There is, of course, a vast literature on the real time prediction and predictability on seasonal time scales of ENSO and its associated climatic effects. Such studies use a wide variety of models ranging from simple statistical models (e.g. Burgers (1999)) to fully couple ocean-atmosphere models incorporating data assimilation systems (e.g. Stockdale et al (1998)). The experiments presented below are examined in Collins et al (2002) to assess the potential predictability of ENSO, which turns out, on average, to be 8 months using a measure based on anomaly correlation coefficient greater than 0.6 (see later). Operational ENSO forecasts have been one of the success stories of climate research over the recent years and have been of considerable value to society. One of the main motivations for this study is to examine (a model of) the climate system for other such predictable components which may be further exploited for societal benefit. We focus principally on interannual to decadal time scales because of the relative abundance of studies which have examined variability on these time scales and highlighted the possibility of predictability and prediction (e.g. Sutton and Allen (1997)).

2 The Model

We use the third version of the Hadley Centre Coupled Model - HadCM3. The atmospheric component of HadCM3 is a version of the UK Met. Office Unified Model (Cullen (1993)). The model dynamics and physics are solved on a $3.75^\circ \times 2.5^\circ$ longitude-latitude grid with 19 hybrid vertical levels. The oceanic component of the model is an updated version of that used in HadCM2 (Johns et al (1997)), which is a version of the Cox (1984) model, with a horizontal resolution of $1.25° \times 1.25°$ and 20 levels in the vertical. A significant improvement with respect to the previous version of the model (HadCM2 - Johns et al (1997)) is the elimination of the flux adjustments which were needed in HadCM2 to keep the model climate stable. HadCM3 has no flux adjustment term and has a stable climate in the global mean. More details of the formulation of HadCM3 and its mean climate can be found in Gordon et al (2000) and Pope et al (2000). A description of the climate variability simulated by the model is given by Collins et al (2001).

One criticism that may be levelled against any study of the potential predictability of climate which uses a coupled AOGCM is that the model may not accurately simulate climate and thus may seriously overestimate or underestimate predictability limits. We partially address this by examining diagnostically the autocorrelation structure of both the model and observed climate (next section) in order to expose any obvious flaws in the model simulation of climate variability. Also the reader is referred to Collins et al (2001) for a more extensive comparison of the model variability with the observed climate variability.

We use the control experiment of HadCM3 in which all concentrations of greenhouse gases etc. are fixed at pre-industrial values together with the four experiments of Stott et al (2000) which we use as “base experiments” for our ensemble runs. The Stott et al (2000) simulations are of the climate from 1870-1997 (which we extend to 2005) using HadCM3 forced with estimated changes in greenhouse gases, anthropogenic and volcanic aerosols, solar output and tropospheric and stratospheric ozone changes. (Post 1997, solar variations were extraplolated over one 11-year cycle and the volcanic forcing was assumed constant at 1997 values). The experiments capture much of the variability on decadal and large spatial scales of the observed surface temperature record. We include the variations in radiative forcing in our simulations in order to examine the boundary value, or second kind, predictability problem in a future study. For this study we remove the climate change signal from the experiments in order to focus cleanly on the initial value problem.

3 Diagnostic Measures of Predictability

Following Boer (2000) it is useful to first examine diagnostic measures of predictability of the model and the observations to validate the model in order to assess the likelihood of the model having similar predictability characteristics to the real climate system, and to aid in defining regions and variables to look at in the perfect model component of the study.

As in Boer (2000) we take the simplest statistical null hypothesis for climate variability, namely the model of Hasselmann (1976) and Frankignoul and Hasselmann (1977) which can be written

$$\frac{\partial X}{\partial t} = -\gamma X + \epsilon, \tag{1}$$

where t is time X is the climate variable in question (e.g. a time and space averaged meteorological variable such as surface air temperature), γ is a damping coefficient and ϵ is some (white) weather noise. Often X is interpreted as the mixed layer temperature of the ocean, but here we use it to mean any climate variable which may be potentially predictable due to some "slow" process which may be local or remote. Hence we do not strictly test the physical null hypothesis of a mixed-layer response to random atmospheric forcing. The discretised version of 1 can be written

$$X_{t+1} = a_1 X_t + a_0 z \tag{2}$$

where $a_1 = 1 - \gamma\Delta t$ is the autocorrelation coefficient at lag Δt, $a_0 = \mathtt{var}(\epsilon)$ and z is taken from a unit variance, zero mean, white noise process. `var` is the variance operator. Equation 2 is the familiar equation of a autoregressive process of order 1 (an AR(1) process).

The AR(1) process can be used to define a simple and incredibly cheap prediction model for climate variables. Boer (2000) shows that predictability characteristics, in terms of the Anomaly Correlation Coefficient (ACC) and Root Mean Squared Error (RMSE), for an AR(1) process are

$$\mathtt{ACC}(t) = a_1^{2t} \tag{3}$$

and

$$\mathtt{RMSE}(t) = \sqrt{2a_0^2(1 - a_1^{2t})}. \tag{4}$$

Hence by examining a_1 coefficients computed from the model and the observations we can get some idea of the likely predictability characteristics. Fitting the AR(1) model in the presence of secular trends can give unrealistically high a_1 coefficients because of the autocorrelation introduce by the trend. Of course, trends in relatively short climate records can be a consequence of natural modes of climate variability in which case we wish to know the predictability characteristics. For the AR(1) fits to the model we use the control simulation of HadCM3 (Gordon et al (2000), Collins et al (2001)) which has a stable surface climate and hence has no global secular trends. The previously stated work of Stott et al (2000) has shown that much of the large-scale multidecadal variability of surface temperature anomalies can be explained by changes in radiative forcing due to increasing greenhouse gases, sulphate aerosols, ozone, variations in solar output and volcanic eruptions. Such radiative forcing changes do not follow a linear trend, hence in order to remove the climate change signal from the observed surface air temperature (Jones et al (1997)), we remove a polynomial trend of order 3 at each grid box. There is a danger that this may either remove some low-frequency climate signal or leave some residual forced climate component that might bias the coefficients of the AR(1) fit. We do not therefore present the following results as accurate estimates of the observed autocorrelation structure but merely as approximate estimates with which to compare the model in order to reveal any obvious model errors or deficiencies. Boer (2000) defines potential predictability diagnostically by looking regions where an AR(1) process is not a good description of the model climate. Collins et al (2001) present evidence that for HadCM3 the tropical Pacific is a region in which the AR(1) null hypothesis can clearly be rejected but that elsewhere it provides a relatively good definition of (at least SAT) variability. We do not follow Boer (2000) by testing for the whiteness of the residuals as, again, we merely use the AR(1) model to highlight any severe model deficiencies. It is clear from the rest of this paper that the prognostic perfect model approach is a much more powerful method with which to identify potential predictability.

Figures 1 and 2 show a_1 coefficients for the model SAT and the detrended observed surface temperature for various averaging operators. For the model, there are large (i.e. > 0.7) values of monthly a_1 throughout the tropical oceans and modest (i.e. $0.3 < a_1 < 0.7$) values over tropical land areas. This is also reflected in the observed monthly a_1 values, although the spatial coverage is rather limited. There are only small values (i.e. < 0.3) of a_1 over mid-latitude land areas (away from the coasts) indicating poor potential predictability on monthly time scales in these regions. On seasonal time scales, both the model and the observed autocorrelations are dominated by the tropical Pacific region with the signal coming from ENSO. There are also values of $a_1 > 0.5$ over regions of both the North Atlantic, tropical Atlantic and North Pacific. On annual time scales the North Atlantic and North Pacific are also regions of modest coefficients together with some isolated regions of the Southern Ocean in the model. These regions also show up with modest a_1 values on pentadal and decadal time scales although there are very few grid boxes in the observations with enough data on which to fit the AR(1) model. On annual time scales and longer there is little sign of any high autocorrelation over land regions with the possible exception of modest correlations in Europe and the Eurasian continental interior on decadal time scales in the model. There is a relatively broad agreement with the similar analysis of Boer (2000). Plots of a_1 (not shown) for the model MSLP and precipitation show some modest values in tropical regions on monthly and seasonal time scales, but only small values on annual time scales and above.

The above analysis represents one of the simplest interpretations of the climate system i.e. damped persistence forced by white noise. It is suitable then to take this model, together with simple persistence, as null hypotheses in the measurement of potential climate predictability. That is, the AR(1) model, along with the simple idea of taking todays climate as a prediction of the future, are suitable benchmarks in identifying possible climate predictability. As we shall see later, the model has greater "memory" than implied by a fitted AR(1) process in many ocean regions but not over many land regions.

4 Perfect Model Experiments

4.1 The Method

We may proceed by following the method of Griffies and Bryan (1997a); Griffies and Bryan (1997b) and Grötzner et al (1999). The perfect model or perfect ensemble approach involves taking an initial ocean-atmosphere state generated by HadCM3, making a small perturbations to the initial conditions and running ensembles of experiments. The rate of divergence of some measure of the ensemble members can then be used to calculate the potential predictability given a perfect model and near-perfect initial conditions.

In operational weather forecasting, great effort is expended to design perturbations which sample the weather PDF according to some pre-defined strategy (with the use of singular (Molteni et al (1996)) and breeding (Toth and Kalnay (1993)) vectors). Often the strategy involves maximising ensemble spread in a particular region or prognostic variable and thus the perturbations point in the directions of instability of the model state space, optimised over geographic region and lead time. For this study we do not wish to focus on any particular geographic region, or variable, so we adopt a more pragmatic approach to perturbations.

The weather forecasting problem has shown us that (even infinitesimal) perturbations to initial conditions result in completely different weather patterns in nearby ensemble members after only a matter of a week or so (Lorenz (1982)). Hence by just making perturbations to the atmospheric component of the model, and keeping the ocean component the same, we are assured that the coupled system has some significant perturbation on climate time scales. We thus take atmospheric initial conditions from successive days from the start of the "control" member and oceanic initial conditions as the same in all the members of a particular ensemble. Taking successive days allows us to make perturbations to the atmosphere without encountering the problem of balance and initialisation which would be required to damp spurious gravity wave motions excited by perturbing atmospheric variables. It also forms the basis for a simple ensemble forecasting system whereby an ocean analysis is made every month and forecasts are then run every day using that days atmospheric analysis. By taking identical ocean initial states we are, in effect, assuming that we could make a perfect analysis of the ocean state at a particular instant. This is, of course, never likely to be possible given the small deformation radius and the problems of measuring the ocean. It is also clear that this system is in no way optimal in the sense of maximising ensemble spread, however it allows us to examine the potential predictability and thus provides a useful bench-mark for any future initial value operational climate forecasting systems.

For all the ensemble experiments used in this study we use ocean initial conditions from 1st of December and atmospheric initial conditions from the 1st, 2nd etc. of December. We perform twelve ensemble experiments each with five members, where each experiment is 10 years in duration. The NINO3 anomalies from the experiments are shown in fig. 4 and provide a good graphical depiction of the experimental set-up. The top four "rows" represent the four independent realisations of 20th century climate simulated by Stott et al (2000). For each row there are three ensemble experiments starting on 1st December 1975, 1985 and 1995 respectively, making twelve experiments in all. The number of experiments and ensemble members might be considered small in comparison with similar studies of weather predictability (e.g. Molteni et al (1996)) or ENSO predictability using intermediate models (e.g. Goswami and Shukla (1991)). The study does however represent over 600 years of coupled model simulation which took over a year on a supercomputer. We make careful use of statistical techniques to judge the significance of measures of potential predictability (see section 4.3 below).

4.2 Choice of “Climate” Variables.

In this study we examine various time series of climate variables to look for potential predictability. There are many ways to choose such time series e.g. taking indices of well known climate phenomena (such as the NINO3 ENSO index or the North Atlantic Oscillation Index (NAO)), taking averages over areas defined by socio-economic or political boundaries, or by optimising regions to find the best/worst predictability (e.g. by examining the autocorrelation coefficients calculated in section 3 or using techniques such as that described by Venzke et al (1999)). Here we use a mixture of techniques to give some broad overview of potential predictability.

A summary of all the indices used in this paper is given in table 1 and many are also represented graphically in fig. 3. The descriptions of the regions are rather loose and no attempt has been made to use precise geographic or political boundaries. The intention is to provide a broad overview of predictability characteristics.

4.3 Measures of predictability

We contrast several simple measures of predictability used in studies of weather forecasting and climate predictability. Let $x_{ij}(t)$ be the value of some climate variable at time t for the ith member of the jth ensemble experiment ($i = 1, 2, .., 5$ and $j = 1, 2, .., 12$ in this case - see above and fig. 4). We also require a "verification" or "truth" for this ensemble. As this is a perfect model study, any ensemble member may be chosen randomly as the truth and sample size can be increased by taking each member in turn. The simulations performed here include variations in radiative forcing due to increasing greenhouse gases, changes in natural and anthropogenic aerosols, ozone changes and variations in solar output (in order to examine the role of boundary conditions in interannual to decadal climate predictability in a future study and the calculation of initial value predictability measures (especially the ACC) can be contaminated by secular trends so we remove the climate change signal from all the time series before plotting and calculating measures of predictability. The signal is computed from the grand ensemble mean across all the members of each ensemble experiment in the relevant time interval (20 experiments in all). For here on we drop the ts for the sake of brevity but all predictability measures are a function of (lead) time.

The average Anomaly Correlation Coefficient (ACC) over all ensemble members is defined as

$$\mathtt{ACC} = \frac{\sum_{j=1}^{12}\sum_{i=1}^{5}\sum_{k\neq i}(x_{kj}-\bar{x})(x_{ij}-\bar{x})}{\sqrt{\sum_{j=1}^{12}\sum_{i=1}^{5}\sum_{k\neq i}(x_{kj}-\bar{x})^2\sum_{j=1}^{12}\sum_{i=1}^{5}\sum_{k\neq i}(x_{ij}-\bar{x})^2}}, \tag{5}$$

where

$$\bar{x} = \frac{1}{20}\sum_{j=1}^{4}\sum_{i=1}^{5}x_{ij} \tag{6}$$

is the ensemble mean climate change signal (as a function of time) over all the experiments. The ACC is 1 for a perfect forecast and zero when there is no skill. Forecast accuracy is a user defined concept, but a value of ACC > 0.6 is often taken (largely it seems for historical reasons (Hollingsworth et al (1980))) as a definition of "useful" forecast accuracy and is marked on all the ACC plots. Simple statistics tell us that a correlation between two independent variables of greater than $\frac{1}{\sqrt{2}}(\simeq 0.7)$ implies that one variable can be used to "explain" greater than half the variance of the other. It is useful also to compute the ACC for a persistence forecast and for a fitted AR(1) process and measure this against the predictability coming from the coupled model. As the sample size is relatively small, the statistical significance of an ACC being greater than zero is assessed using the student t-test. The effective number of degrees of freedom in the t-test is somewhere between 59 (the number of independent numbers in the ACC calculation at each lead time) and 240 (the number of variables in the summation of equ. 5). To avoid overestimating the effective degrees of freedom, which would increase the likelihood of falsely identifying a ACC as greater than zero when it is really indistinguishable from zero, the lower value of 59 is taken.

The average Root Mean Squared Error (RMSE) is defined as

$$\mathtt{RMSE} = \sqrt{\frac{1}{239}\sum_{j=1}^{12}\sum_{i=1}^{5}\sum_{k\neq i}(x_{kj} - x_{ij})^2} \qquad (7)$$

with some measure of forecast accuracy when the RMSE is less than $\sqrt{2}$ times the climatological RMS of x. The $\sqrt{2}$ factor comes from the definition of RMSE as the difference between two values of x. The statistical significance of the RMSE being less than (or greater than) the climatological RMS is judged using an F-test with the degrees of freedom estimated as before. When the RMSE saturates at the climatological value, we say that the potential predictability is zero.

We also use the measure of Collins and Allen (2002) defined as

$$I = 1 - \frac{1}{24}\sum_{i=1}^{5}\sum_{j=1}^{12}\frac{(x_{ij} - \bar{x})^2}{\mathtt{var}(x)} \qquad (8)$$

which is a normalised variance measure which is 1 for a perfect forecast and < 0 when less than half the climatological variance is predictable. The $\frac{1}{24}$ factor comes from the $\frac{2}{n(m-1)}$ factor of equ. 11 of Collins and Allen (2002) with $n = 12$ and $m = 5$ in this case. $\mathtt{var}(x)$

is the climatological variance computed from the control simulation. As for the RMSE, statistical significance is judged using an F-test. We do not use the trick of taking each member of the ensemble as the truth in calculating I in order to contrast this with the other measures of potential predictability.

There are a large number of more exotic measures of forecast skill used in weather and ENSO forecast verification, such as Brier Scores, Relative Operating Characteristics, Cost/Loss ratios etc. (Katz and Murphy (1997)). Many of these measures require relatively large ensembles sizes in order to compute high resolution estimates of the forecast PDF. Because of the relatively small ensembles size used in this study, it was not possible to accurately apply any of these techniques.

4.4 Potential Predictability

4.4.1 Interpretation of predictability indicators

We first examine the potential predictability of ENSO by re-examining the results of Collins et al (2002) in order to highlight the method. Figure 4 shows the seasonally averaged NINO3 anomalies from the ensemble experiments. A section of the observed NINO3 anomalies are also shown to indicate the relative realism of the model ENSO cycle (see Collins et al (2001) for more details). Clearly there is little correlation between the ensemble members at lead times greater than a few years. Thus even relatively small perturbations to the atmosphere can result in completely different ENSO states after a few years. The ACC predictability measure (fig. 5(a)) is 1 at a lead time of 0 seasons (by construction), drops below 0.6 after 3 seasons and is statistically indistinguishable from 0 after 5 seasons. The perfect model experiment does beat simple persistence and the AR(1) process for the initial 5 seasons, indicating that

ENSO is unlikely to be simply a mixed layer response to random atmospheric fluctuations with parameters derived from the AR(1) fit (an obvious point in retrospect) and that there is skill to be gained by constructing statistical and dynamical forecasting schemes for ENSO. (While an AR(1) process is probably not a good null hypothesis for the quasi-period ENSO cycle (see e.g. Burgers (1999)) we show it for consistency with other climate variables). The RMSE (fig. 5(b)) shows an increase from 0 and saturates at the climatological value after 3 seasons. There is a seasonal cycle in climatological RMS which is not evident in the other predictability measures. The I measure of Collins and Allen (2002) drops below zero after only 2 seasons indicating that less than half the climatological variance is predictable after this time. This is in agreement with the ACC measure which drops below 0.7 after 2 seasons (see above). We must re-iterate that forecast accuracy is a user defined concept and there is no universally defined skill standard. Hence there reader is free to interpret these measures of potential predictability as they wish. However, it is worth taking on board experience gained in weather forecasting which shows that a forecast with an ACC of less than 0.6 is of little use (e.g. Hollingsworth et al (1980)).

4.4.2 Predictability of global mean temperature

We next examine the potential predictability of global mean surface air temperature (SAT). While the utility of forecast of such global mean quantities is unclear, such forecasts are currently produced for the UK government (Folland and Colman (2000)). We assess the predictability of annual mean (December-November) values as it is inappropriate to examine seasonal values of a global mean quantity. The ACC for annual global mean SAT (fig. 6(a)) is 0.9 at a lead time of 1 year and persistence and the AR(1) process are clearly beaten by the perfect model. The ACC then drops below 0.6 but remains above 0 for 6 years, although a persistence forecast is just as skillful. We highlight here a potential problem with computing ACC-type measures of quantities that may include significant secular trends. If one is not careful in removing secular trends ACC measures can appear to be falsely large. In this case the RMSE measure (fig. 6(b)) which remains below the climatological RMS out to 6 years lead time and the I measure (fig. 6(b)) seems to confirm the conclusions from the ACC calculation. For both global mean land SAT and global mean ocean SAT (figures not show) the conclusions are similar but with the ocean predictability being slightly larger than the land predictability. This is perhaps surprising as one might have expected ocean temperatures to be significantly more predictable than land temperatures (however see Hall and Manabe (2000)). We see later that this is indeed the case for regional averages of SAT.

For pentadal averaged global mean SAT (table 2) the first 5 year average does show some potential predictability with ACCs around 0.6 for global mean, global mean land and global mean ocean SAT and RMSEs less than the climatological average. For decadal averages (table 3) the ACC ranges from approximately 0.5 to 0.6 for the different global indices. Hence it appears that annual mean global mean temperatures are potentially predictable 1 year in advance and that longer time averages are also marginally predictable 5 and 10 years in advance.

4.4.3 Predictability over ocean regions

We next examine the potential predictability of regional ocean temperatures using the regions defined in fig. 3 (the potential predictability of ENSO was examined above and in Collins et al (2002)). In the tropics it is well documented that SST anomalies can cause anomalous variations in local climate (e.g. Rowell (1998)) and anomalous variations in some mid-latitude regions via Rossby wave teleconnections (e.g. Hoskins and Karoly (1981)). In mid-latitudes there is not such a strong local causal link between SST anomalies and climate anomalies, although some studies have hinted at a potential ocean to atmosphere coupling on decadal time scales (Rodwell et al (1999), Grötzner et al (1998)).

The Tropical Atlantic region is influenced by the remote effects of ENSO and the NAO but also has variations which are confined locally (Sutton et al (2001)). It is also a region of relatively large a_1 coefficients at monthly to decadal time scales in both the model and the observations (figs. 1 and 2). The ACC predictability measure for the monthly averaged Tropical Atlantic SAT index (fig. 7) shows a drop off in skill after the first 3 months but then shows a return of skill in the following autumn and winter. This is also a feature of the RMSE and I measures. After the initial return of skill the ACC drops to zero. Taking the mechanisms of Sutton et al (2001) in turn, it is unlikely that this return of skill is due to some predictable component of the NAO (see later) nor is it likely that it is due solely to some lagged response of the predictable component of ENSO, as the tropical Atlantic index is on average correlated at a value of 0.6 at lags of 3-4 months with the NINO3 index which has an ACC of less than 0.5 at these lead times. Perhaps a more likely explanation is that the return of skill is some combination of the remote effects of ENSO and some other ocean dynamical effect. There are some significant values of ACC and values of RMSE lower than the climatological average for pentadal and decadal averages (tables 2 and 3) hence the reason for potential predictability in the Tropical Atlantic might be some decadal time scale oceanic variability arising from variations in the poleward heat transport associated with variations in the strength of the thermohaline circulation (Dong and Sutton (2001)).

In contrast, the ACC for the sub-tropical North Pacific index (fig. 7), which does have some similarly large values of a_1 in fig. 1, is just above 0.6 one season into the experiments but drops below 0.6 thereafter. There is a plateau of modest ACC values for around a year and this is also evident in RMSE values less than the climatological RMS. However, it appears that a simple persistence forecast is just as adequate as a perfect model forecast. While there is some potential predictability in the low-latitude North Atlantic on interannual time scales, the low-latitude North Pacific shows a relative lack of potential predictability.

Several studies have highlighted decadal time scale variations and possible predictability of Mid-latitude SST anomalies in both the Atlantic (Sutton and Allen (1997)) and the Pacific (Latif and Barnett (1994)). Again these are both regions which also show relative high values of a_1 in both the observations and the model (figs. 1 and 2) so would seem good candidates for potential predictability beyond the seasonal-annual time scale. The ACC skill measure for the seasonally averaged North Atlantic SAT index (fig. 7) remains above or around 0.6 out to a lead time of 8 seasons (although this is matched by the simple persistence forecast) and the remains above zero for nearly all of the 10 years of the ensemble experiments. This extended range potential predictability is also shown by the RMSE measure which is significantly below the climatological RMS for the entire 10 years of the ensemble experiments. The I measure is intriguing as it shows potential decadal skill only during the summer months (although not statistically significant). Pentadal and decadal averages (tables 2 and 3) also show significant ACCs > 0.6 and low RMSEs confirming the long-lead time potential predictability seen in the seasonal averages.

This high level of potential predictability can be clearly seen by examining the time series of the North Atlantic (fig. 8). While there is little correlation between ensemble members on seasonal time scales, there are clear decadal variations which are common in several of the ensemble experiments. These plots are similar to those shown by Griffies and Bryan (1997b) for their leading EOF of SST (see fig. 16 of that paper), although no such decadal predictability in this region was found by Grötzner et al (1999) or Boer (2000). The large disparity between the ACC and RMSE for this index and the corresponding ACC and RMSE for an AR(1) process (fig. 7), indicates that it is unlikely that these decadal fluctuations are simply a mixed layer response to random atmospheric noise. A mechanism involving re-emergence of subducted mixed-layer anomalies during the winter season (Alexander et al

(1999)) is also unlikely as the predictability persists through the seasons and is at a maximum during the summer. Dong and Sutton (2001) have shown that the dominant mechanism for decadal variability in North Atlantic ocean heat transport in HadCM3 involves a mixed thermohaline/gyre mode driven by variations in buoyancy flux and windstress curl. This is perhaps the most likely candidate for the decadal potential predictability, although this does imply that there must be some ocean-to-atmosphere coupling in order to set the correct pattern of buoyancy flux and windstress curl. An examination of the mechanisms of predictability will be the subject of future work.

Again, in contrast to the Atlantic, for the North Pacific index, there appears to be less potential predictability with the ACC dropping below 0.6 after 2 seasons but with a plateau of modest ACC values for 2 years after this. As was the case for the Sub-tropical Pacific index, a persistence forecast is just as skillful as a perfect model forecast. This is also reflected in the relatively lower predictability indicators for pentadal and decadal averages (tables 2 and 3).

4.4.4 Predictability of SAT over land regions

We next turn our attention to the potential predictability of regional variations of climate over land. While predictions over ocean regions may be of some practical use (e.g. for fisheries) unless anomalies in ocean circulation manifest themselves as variations in climate over land then they will be of little use to the majority of society. As explained above, we examine the potential predictability over rather arbitrarily selected broad geographic regions simply to give a flavour for the potential predictability of regional land climate. Future work could involve more precise geographic definitions or optimal definitions of patterns of potential predictability.

Predictability measures for three northern hemisphere SAT indices are shown in fig. 9. For each index there is a rapid drop-off of ACC with values below 0.6 even for the first season of the experiments. This lack of potential predictability is also evident in the rapid saturation of RMSE and variance. The unpredictability is perhaps not surprising given the small values of autocorrelation seen in fig. 1 although the lack of potential predictability in North Western Europe is perhaps the most disappointing given the long lead time potential predictability of the SSTs in the North Atlantic. It would seem that much of the variability of SAT in mid-latitudes land regions is simply the average of unpredictable weather events which are little effected by the sea surface anomalies over which they travel. The potential predictability of pentadal and decadal averages (tables 2 and 3) is similarly low for the three measures of Northern Hemisphere Land SAT, although for the large area Eurasian Land SAT index there are some significant but modest values of ACC and some RMSE values significantly lower than climatology. There may be the potential for longer time scale prediction of decadal mean temperatures beyond the 10 year limit used here.

SAT in sub-tropical and tropical regions (fig. 10) are more potentially predictability than the northern hemisphere mid-latitude SAT indices. African Land SAT shows potential predictability out to several seasons and in tropical South America, there is potential predictability out to a lead time of about a year. This is presumably due to the influence of ENSO and Tropical Atlantic SSTs (both of which are similarly predictable) on climate in this region (e.g. Rowell et al (1995)). Asian Land SAT behaves more like the mid-latitude land SAT indices with perhaps marginal skill a one season but very little hope of predictability (on average) beyond this.

For the southern Hemisphere mid-latitudes, both the Southern South American land SAT index and Australian land SAT show some potential predictability a lead times of a season or two (fig. 11). Both indices are only modestly correlated with NINO3 (correlations of 0.2-0.3

at a lag of 3 months) so the slightly enhanced skill in comparison with the northern hemisphere land SAT indices is unlikely to be entirely due to some predictable teleconnection from ENSO. It is possible that it is a consequence of it being summer in the southern hemisphere and that there is some seasonality to seasonal potential predictability. Unfortunately, to adequately test this hypothesis requires performing further ensemble experiments starting at different times of the year. Constraints on computer time mean that such experiments have were not possible. They should, however, be a priority for the future for those interested in southern hemisphere predictability.

In contrast then to the potential predictability in ocean regions, it would seem that there is little potential predictability of seasonal mean SAT beyond a season or two, particularly for the northern mid-latitude regions.

4.4.5 Other climate variables

Climate variables other than SAT were also examined for potential predictability and a small section of the results are presented here. The model NAO index (as defined by the leading EOF of MSLP in the region 90°W-90°E, 10°N-90°N) shows little potential predictability (fig. 12) with only modest ACC values after one season and essentially zero ACC thereafter (we define the NAO in winter only). There is no indication of any potential predictability for the pentadal and decadal averaged NAO index (tables 2 and 3). HadCM3 does show some decadal time scale variations in the NAO (see fig. 15 of Collins et al (2001)) but it would appear that these are essentially unpredictable. Rodwell et al (1999) and Hoerling et al (2001) have suggested a role for SST anomalies in forcing NAO variations on decadal time scales. Hoerling et al (2001) suggest that tropical SST anomalies attributable to changes in radiative forcing are the likely source of such decadal SST variations. The model simulations performed here include estimates of changes in the radiative forcing of climate, but they do not reproduce the observed upward trend or decadal variability in the NAO which is seen in the observations.

Land precipitation anomalies are less predictable than their SAT equivalents (a selection of which are shown in fig. 13 and tables 2 and 3) with only the tropical South American region showing some seasonal potential predictability due, presumably, to the influence of either ENSO or tropical Atlantic SSTs. The message is similar to that for land SAT anomalies; their appears to be little potential predictability on average.

4.4.6 Maps of potential predictability

Since the selection of geographic regions is somewhat arbitrary we finally examine maps of the potential predictability measures computed at each model grid-point. Although we have noted that climate variables are perhaps better defined over spatial areas, the examination of such maps should give some indication of the choice of such spatial averages.

Only the ACC measure of potential predictability is plotted to limit the number of figures. The ACC is shaded when (i) it is statistically significantly greater that zero and (ii) the perfect model ACC is greater than that for a persistence forecast. Maps of the potential predictability of SAT are shown in figs. 14 and 15 for various different lead times and averaging periods. For seasonal means there are large regions of high ACC over the ocean and in tropical regions at a lead time of one season. At two seasons lead time and beyond the tropics retain potential predictability in some regions and there are some small areas of the N. Pacific and N. Atlantic which show significant ACCs. Over much of the mid-latitude land areas ACCs of seasonal means are indistinguishable from zero. The same is broadly true for annually averaged SAT; the tropics the N. Pacific and the N. Atlantic show relatively large ACCs out to 1 year and similarly reduced ACCs out to 2 years. There are also some regions of the Southern Ocean which shows potential predictability on annual time scales. Maps for longer time averages show some potential predictability for tropical regions on pentadal time scales. The only regions which shows potential predictability on decadal time scales are the Atlantic region in the northern hemisphere (as was seen in the analysis of Atlantic indices) and some regions of the Southern Ocean. As was found in the analysis of climate indices, there appears to be little or no multi-annual to decadal average potential predictability for land areas.

Maps of ACC for MSLP (Fig. 16) show seasonal mean MSLP in the the tropics and sub-tropics to be potentially predictable out to one season. For mid-latitude regions though there is little sign of any potential predictability of seasonal MSLP. Intriguingly, there are some large regions of annual-decadal potential predictability of MSLP. However, these regions do not have similarly predictable variations in temperature and precipitation. ACCs for precipitation itself (fig. 17) show that the ENSO region is the only area for which there are significant ACCs for seasonal mean precipitation at a lead time of one season. At a lead time of beyond one season and for longer time scale averages there is little indication of any potential predictability, confirming the results of the analysis with geographic averages.

5 State Dependent Predictability

The above analysis represents an estimate of the *average* potential predictability of various climate variables. There will be states of (the model) climate which are more predictable and less predictable than this average. It is possible that the states which may be more predictable than average will be those climate states which are unusual (e.g. in the tails of the distribution) and thus have the largest impact on society. Figure 18 shows an example of North West European temperatures from one of the ensemble experiments used to compute the average potential predictability. In this region (fig. 9) there was very little average skill with an ACC of less than 0.6 even in the first season (DJF). In fig. 18 the number of ensemble members for this particular experiment has been increased from 5 to 12 for the initial 1.5 years of the experiment. The experiment is interesting as the initial conditions show a relatively large negative anomaly in SAT and nearly all of the ensemble members lie below the zero line after the first season and many of the members lie above the zero line after 6 seasons. The ACC computed for this ensemble only is greater than 0.6 at a lead time of one season and the RMSE is smaller than the case for the average predictability. Hence there can be “useful” potential predictability for individual cases even when there is little hope of predictability on average. The caveat to this particular example is that even though the perfect model ensemble shows potential predictability, the skill as measured by the ACC, is lower than that for a simple persistence forecast. We expect there to be cases when this is not true and there may be useful skill to be gained using a coupled AOGCM.

6 Conclusions

We have examined a coupled ocean atmosphere global circulation model (HadCM3) to look for aspects of the variability of climate which might be predictable using the initial value technique.

The simplest null hypothesis for climate variability is the damped persistence with white noise forcing of Hasselmann (1976) and Frankignoul and Hasselmann (1977) which can be discretised as an AR(1) process. By fitting an AR(1) process to the model surface air temperature (SAT) and the observed surface temperature we firstly showed that the model has similar variability to that observed and hence the potential predictability derived from the model is likely to be similar to that of the real climate system. The AR(1) process, together with the idea of simple persistence, are suitable bench-marks to measure potential predictability against.

We have used the perfect model or perfect ensemble approach to assessing predictability by taking HadCM3 and running ensembles of experiments with small perturbations to the initial conditions. By measuring in the rate of divergence of near-by ensembles members using simple statistical measures, namely anomaly correlation coefficients (ACCs), root mean squared error (RMSE) and a variance measure derived by Collins and Allen (2002) the potential predictability can be found. The experimental set-up was such that a relatively large number (12) of ensemble experiments were performed with only a small number of ensemble members (5) in each case. The potential predictability is measured over all these members and hence the limits should be interpreted as the average potential predictability limits.

Annual mean global mean SAT (over both land and ocean together as well as separately) is potentially predictability one year ahead, but there is only low predictability thereafter. Longer averages of global mean SAT (i.e. 5 and 10 year averages) are also potentially predictable.

SAT over regional ocean areas show a wide range of potential predictability. In the tropical Pacific, NINO3 anomalies are potentially predictable out to a lead time of 3 seasons based on an ACC > 0.6 and a RMSE less that the climatological average. In the tropical Atlantic, there is a similar level of potential predictability on monthly-seasonal time scales, but also there is some indication of potential predictability on multi-annual to decadal time scales. In northern mid-latitudes, the North Pacific is region of relatively low predictability even on seasonal time scales. The North Atlantic region in the model is the region which shows the longest lead-time potential predictability for seasonal through to decadal averages. Longer predictability experiments are required to further investigate this source of predictability. In all cases of ocean area-averaged SAT, the model shows a longer lead-time predictability limit than that derived from the simple fitted AR(1) null hypothesis, indicating that there is more "memory" in the system than would be expected from a simple model of a mixed layer ocean forced by white noise in which the model parameters are precisely those of the fitted AR(1) process. The tougher bench-mark to beat is the idea of simple persistence which can, in some cases, be at least as skillful as a perfect model.

The story is very different for SAT over land areas and for MSLP and precipitation. While there is some seasonal-annual time scale potential predictability in the tropics, in much of the mid-latitude regions, and in particular the highly populated northern land regions, there is very little sign of any average potential predictability beyond seasonal lead times. For the 3 indices of northern hemisphere land averaged SAT, is also seems that the AR(1) process is an adequate model of climate variability. It would seem that in mid-latitudes, much of the variability in climate (that is not forced by "external" factors such as increasing greenhouse gases) is simply due to the averaging of unpredictable weather events.

Grötzner et al (1999) present a simple conceptual model which may be used to interpret the results of this and other predictability studies. The model is that of a coupled ocean-atmosphere oscillator with a representation of ocean "memory", a response of the atmosphere to the ocean, and a feedback of the atmosphere on the ocean. It is shown by Grötzner et al (1999) that the whiteness of the spectra of land SAT limits the predictability on inter-annual to decadal time scales, whereas the redness of ocean variables allows for some long lead-time predictability. In their model, decadal "modes" of variability do not necessarily lead to any potential predictability; signals can easily be swamped by atmospheric noise.

This study represents a broad look at average initial value potential climate predictability on interannual to decadal time scales. No attempt has been made to examine mechanisms for predictability or the lack of it and this will be the focus of future studies.

7 Discussion

7.1 Towards real climate prediction

It is important to re-iterate that the predictability limits derived in this study are those assuming a perfect model and near perfect knowledge of the initial conditions. In practice, models of the climate system will always use descretised (and thus approximate) versions of the equations of motion and will probably rely on some empirical parametrisation schemes for the unresolved scales. Thus models will never be perfect. Perhaps a more important uncertainty though is in any initial state that we require. The small deformation radius in comparison with the atmosphere and the problems involved in observing the ocean below the surface, mean that we may never obtain ocean analyses with similar levels of accuracy to those currently available for the atmosphere. The ensemble method used in this study could provide some clues as to the most appropriate regions of the ocean to observe in order to provide ocean analyses which are accurate enough only in the regions which can then lead to the most skillful, or most valuable, climate predictions. By examining ensemble experiments which show potential predictability and identifying the key oceanic regions and variables, it would be possible to target observations for climate forecasting.

7.2 State dependent predictability

An example was given for land SAT in the North West European region which showed that there are some states of the climate system which are more predictable than on average. This state-dependence is well know in the weather forecasting literature (e.g. Molteni et al (1996))) where skillful lead times can vary from days to weeks depending on the current weather regime. Either many more ensemble experiments are required in order to adequately sample the “climate attractor” and to identify the states which show high potential predictability, or techniques need to be developed (such as the singular vectors of weather forecasting) which provide a cheaper (and arguably more elegant) estimate of predictability. There has been some success in the case of ENSO in applying singular vectors (e.g. Chen et al (1997)) but clearly there is more work to be done.

7.3 Ensemble size

Computer time severely limited the ensemble size used in this study and hence only simple measures of predictability were used to assess ensemble spread. As is the case for ensemble weather forecasting, it is expected that increasing the size of the ensemble, and thus better resolving the forecast PDF, can provide more information to the user to aid in their decision making process. Increased ensemble size in a study of this nature would allow more complex (and perhaps more informative) measures of ensemble spread to be calculated and lead to more useful estimates of potential predictability.

7.4 How model dependent are these results?

One criticism which may be levelled at a study of this nature is that the climate model may not be faithful representation of the real climate system. We addressed this problem partially by examining the autocorrelation characteristics of the model SAT in comparison with the observed SAT autocorrelation. Also the reader is referred to Gordon et al (2000) and Collins et al (2001) for further information about the mean climate and variability of HadCM3. Model-to-model comparisons can be of some use and it appears that the unpredictability of land SAT is also a feature of the studies of Grötzner et al (1999) and Boer (2000). There do appear thought to be inter-model differences in the predictability characteristics of North

Atlantic SSTs which show the potential for interannual to decadal predictability in this study and that of Griffies and Bryan (1997a); Griffies and Bryan (1997b) but show little predictability in the studies of Grötzner et al (1999) and Boer (2000). While there are clearly differences between the models that may lead to different physical mechanisms and coupling strengths, because of the relatively small number of experiments carried out in all coupled model studies of potential predictability, the state-dependent nature of climate predictability could explain such differences.

It is worth noting however that even a AOGCM is a simplification of the real system with a large, but finite, number of degrees of freedom. There are many known (and perhaps unknown) processes which are either not included in the model, or are represented in a highly simplistic manner, but which are active in the real climate system. It is impossible to say what the difference in potential predictability would be if the model was improved (e.g. by increasing the resolution of the ocean such that ocean eddies are permitted or even resolved) without performing such a study. However, a climate model such as HadCM3 does represent many of the important physical processes responsible for variations in climate and it is hard to see how subtle improvements to the model might result in an order of magnitude increase in the potential predictability limits derived here. If anything, improving the model will almost certainly lead to an increase in complexity and thus an increase in the degrees of freedom of the model leading to a reduction in the potential predictability limits. The possibility that greater predictability limits would arise in a more realistic model cannot, however, be ruled out.

7.5 Other factors which may lead to potential climate predictability

There are “external” factors which can effect climate, for example increasing greenhouse gases and explosive volcanic eruptions. Such changes in radiative forcing can be exploited to make predictions in climate. For example, the increase in greenhouse gases over the 20th century, and the relatively slow uptake of the corresponding excess heat by the ocean, means that we are already “committed” so some climate change (Wetherald et al (2001)) and this can be used to constrain future climate forecasts (Allen et al (2000)). While it is currently impossible to predict explosive volcanic eruptions, once they have occured their (cooling) effect on climate can be used to make predictions. The boundary value problem in climate prediction is discussed further in Collins and Allen (2002) and will be the subject of a future study.

Acknowledgements

Thanks to Dave Frame, Bablu Sinha and Chris Wilson who performed some of the model experiments. This work was funded by the Natural Environmental Research Council under the Coupled Ocean-Atmosphere Processes and European Climate programme and by the PREDICATE project (EVK2-CT-1999-00020) of the European Union Framework 5 programme.

References

Alexander MA, Deser C, Timlin MS (1999) The reemergence of SST anomalies in the North Pacific ocean. J Climate 12(8):2419–2433.

Allen MR, Stott PA, MItchell JFB, Schnur R, Delworth TL (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature 407:617–620.

Boer G (2000) A study of atmosphere-ocean predictability on long time scales. Climate Dynamics 16:469–472.

Burgers G (1999) The El-Niño stochastics oscillator. Climate Dynamics 15:521–531.

Chen YQ, Battisti DS, Palmer TN, Barsugli J, Sarachik ES (1997) A study of the predictability of tropical Pacific SST in a coupled ocean-atmosphere model using singular vector analysis: The role of the annual cycle and the ENSO cycle. Mon Wea Rev 125:831–845.

Collins M, Allen MR (2002) On assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. J Climate In Press.

Collins M, Tett SFB, Cooper C (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics 17:61–81.

Collins M, Frame D, Sinha B, Wilson C (2002) How far ahead could we predict El Niño. Geophys Res Letts In Press.

Cox MD (1984) A primitive equation, three dimensional model of the ocean. Ocean Group Technical Report 1, GFDL Princeton.

Cullen MJP (1993) The unified forecast/climate model. Meteorological Magazine 122:81–94.

Dong BW, Sutton RT (2001) The dominant mechanisms of variability in Atlantic ocean heat transport in a coupled ocean-atmosphere GCM. Geophys Res Letts 28:2445–2448.

Folland CK, Colman AW (2000) Empirical prediction of the global temperature anomaly for 2000. In Kirtman B, editor, Experimental Long Lead Forecast Bulletin volume 10.

Frankignoul C, Hasselmann K (1977) Stochastic climate models, Part II: Application to sea-surface temperature anomalies and thermocline variability. Tellus 29:289–305.

Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics 16:147–168.

Goswami BN, Shukla J (1991) Predictability of a coupled ocean-atmosphere model. J Climate 4:3–22.

Griffies SM, Bryan K (1997a) Predictability of North Atlantic multidecadal climate variability. Science 275:181–184.

Griffies SM, Bryan K (1997b) A predictability study of simulated North Atlantic multidecadal variability. Climate Dynamics 13:459–488.

Grötzner AM, Latif M, Barnett TP (1998) A decadal climate cycle in the North Atlantic ocean as simulated by the ECHO coupled GCM. J Climate 11:831–847.

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

Hall A, Manabe S (2000) Effect of water vapour feedback on internal and anthropogenic variations of the global hydrologic cycle. J Geophys Res 105(D5):6935–6944.

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

Hoerling MP, Hurrell JW, Xu T (2001) Tropical origins for recent North Atlantic climate change. Science 292:90–92.

Hollingsworth A, Arpe K, Tiedtke M, Capaldo M, Savijarvi H (1980) The performance of a medium range forecast model in winter - impact of physical parameterizations. Mon Wea Rev 108:1736–1773.

Hoskins BJ, Karoly DJ (1981) The steady linear response of a spherical atmosphere to thermal and orographic forcing. J Atmos Sci 38:1179–1196.

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. Climate Dynamics 13:103–134.

Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. Journal of Climate 10:2548–2568.

Katz RW, Murphy A (1997) Economic value of weather and climate forecasts. Cambridge University Press.

Latif M, Barnett TP (1994) Causes of decadal variability over the North Pacific and North America. Science 266:634–637.

Latif M, Barnett TP, Cane MA, Flügel M, Graham NE, von Storch H, Xu JS, Zebiak SE (1994) A review of ENSO prediction studies. Climate Dynamics 9:167–179.

Lorenz EN (1975) The physical bases of climate and climate modelling. In Climate predictability number 16 in GARP Publication Series 132–136. WMO.

Lorenz EN (1982) Atmospheric predictability experiments with a large numerical model. Tellus 34:505–513.

Molteni F, Buzzia R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: Methodology and validation. Q J Roy Met Soc 122:73–119.

Orrell D, Smith LA, Barkmeijer J, Palmer TN (2001) Model error in weather forecasting. Nonlin Proc Geophys In press.

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model – HadAM3. Climate Dynamics 16:123–146.

Rodwell MJ, Rowell DP, Folland CK (1999) Oceanic forcing of the winter North Atlantic Oscillation and European climate. Nature 398:320–323.

Rowell DP, Folland CK, Maskell K, Ward MN (1995) Variability of summer rainfall over tropical North Africa (1906-92): Observations and modelling. Quart J Roy Meteor Soc 121:669–704.

Rowell DP (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11:109–120.

Stockdale TN, Anderson DLT, Alves JOS, Balmaseda MA (1998) Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. Nature 392.

Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ (2000) External control of 20th century temperature by natural and anthropogenic factors. Science 290:2133–2137.

Sutton RT, Allen MR (1997) Decadal predictability of north Atlantic sea surface temperature and climate. Nature 338:563–567.

Sutton RT, Jewson SP, Rowell DP (2001) The elements of climate variability in the tropical Atlantic. J Climate 13:3261–3284.

Toth Z, Kalnay E (1993) Ensemble forecasting at NMC: the generation of perturbations. Bull Amer Met Soc 74:2317–2330.

Venzke S, Allen MR, Sutton RT, Rowell DP (1999) The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. J Climate 12(8):2562–2584.
Wetherald RT, Stouffer RJ, Dixon KW (2001) Committed warming and its implications for climate change. Geophys Res Letts 28:1535–1538.

Short Description	Region	Land/Ocean	Notes
Global Mean	Global	Both	
Global Mean Land	Global	Land	
Global Mean Ocean	Global	Ocean	Includes sea-ice regions
NINO3	150°W-90°W, 5°S-5°N	Ocean	
Tropical Atlantic	70°W-20°W, 0°-20°N	Ocean	
North Atlantic	50°W-10°W, 40°N-60°N	Ocean	
North Pacific	160°W-120°W, 30°N-50°N	Ocean	
Sub-tropical Pacific	120°E-110°W, 10°N-30°N	Ocean	
North West Europe	10°W-50°E, 30°N-70°N	Land	
Eurasian Land	0°W-180°E, 30°N-70°N	Land	
Asia	60°E-120°E, 10°N-40°N	Land	
North America and Canada	160°W-50°W, 20°N-70°N	Land	
Australia	100°E-160°E, 10°S-40°S	Land	
Tropical South America	90°W-30°W, 15°S-15°N	Land	
Southern South America	90°W-30°W, 60°S-15°S	Land	
Africa	20°W-60°E, 40°S-30°N	Land	
NAO index	90°W-90°E, 10°N-90°N	Both	Leading EOF of MSLP

Table 1: Definitions of geographic regions and other climate indices used in this study. The regions are defined rather arbitrarily to give a flavour of the potential predictability and no attempt has been made to use precise political boundaries or to optimise the regions to show high or low potential predictability.

		ACC		RMSE		*I*	
Index	Season	1-5 years	6-10 years	1-5 years	6-10 years	1-5 years	6-10 years
Global Mean SAT	All	**0.62** > >	0.13 < >	**0.08** < <	0.09 < <	0.19 >	-0.24 >
Gobal Mean Land SAT	All	**0.54** > >	0.15 > >	**0.13** < <	**0.12** < <	-0.03 >	-0.01 >
Global Mean Ocean SAT	All	**0.65** > >	0.11 > >	**0.06** < <	0.08 < <	0.3 >	-0.4 >
Tropical Atlantic SST	All	**0.5** > >	-0.09 > <	**0.14** < <	0.16 < <	-0.27 >	-0.49 >
	DJF	**0.45** > >	0.06 > >	0.15 < <	**0.13** < <	-0.42 >	0.03 >
	MAM	**0.37** > >	-0.04 < <	**0.17** < <	**0.19** < <	-0.11 >	-0.17 >
	JJA	**0.46** > >	-0.09 < <	**0.16** < <	0.19 < <	-0.28 >	**-0.65** >
	SON	**0.54** > >	-0.17 < <	**0.14** < <	0.17 < <	-0.25 >	-0.59 >
N Atlantic SST	All	**0.82** > >	0.02 > >	**0.23** < <	**0.38** < <	0.61 >	-0.02 >
	DJF	**0.78** > >	0.16 > >	**0.28** < <	**0.43** < <	0.53 >	-0.05 >
	MAM	**0.82** > >	**0.25** > >	**0.25** < <	**0.37** < <	0.65 >	0.26 >
	JJA	**0.81** > >	0.03 > >	**0.25** < <	**0.37** < <	0.55 >	0.08 >
	SON	**0.74** > >	0.06 > >	**0.26** < <	**0.37** < <	0.45 >	-0.1 >
N Pacific SAT	All	**0.29** > >	0.1 > >	**0.23** < <	0.3 < <	0.11 >	-0.4 >
	DJF	**0.31** > >	0.04 > >	**0.24** < <	0.28 < <	-0.06 >	-0.35 >
	MAM	**0.28** > >	0.09 > >	**0.25** < <	**0.3** < <	0.04 >	-0.33 >
	JJA	**0.23** > >	0.16 > >	**0.27** < <	0.34 < <	0.15 >	-0.3 >
	SON	0.21 < >	-0.01 > <	**0.27** < <	0.32 < <	0.01 >	-0.39 >
NW Euro Land SAT	All	**0.22** < >	**0.34** > >	0.34 < <	0.33 < <	-0.37 >	-0.21 >
	DJF	0.12 > >	0.11 > >	**0.58** < <	0.74 < <	-0.14 >	**-0.71** >
	MAM	0.06 < >	**0.27** > >	**0.44** < <	**0.41** < <	-0.52 >	-0.21 >
	JJA	0.05 < >	0.08 > >	0.37 > >	0.39 > >	**-1.21** <	**-1.41** <
	SON	**0.22** > >	**0.45** > >	0.5 < <	**0.32** < <	**-0.87** >	0.24 >
Eurasian Land SAT	All	**0.49** > >	**0.44** < >	**0.22** < <	**0.19** < <	-0.08 >	0.17 >
	DJF	**0.41** > >	**0.23** < >	**0.43** < <	0.48 < <	-0.11 >	-0.4 >
	MAM	0.1 < >	**0.28** < >	0.36 < <	0.34 < <	**-0.57** >	-0.36 >
	JJA	**0.29** < >	0.26 > >	**0.14** < <	0.19 < <	-0.19 >	**-0.89** >
	SON	**0.34** > >	0.24 < >	0.29 < <	0.3 < <	-0.6 >	**-0.74** >
N American Land SAT	All	0.03 < >	**0.38** > >	**0.26** < <	**0.18** < <	-0.26 >	0.35 >
	DJF	-0.05 < <	0.02 > >	0.53 < <	**0.44** < <	**-0.63** >	-0.16 >
	MAM	-0.06 > <	**0.3** > >	0.39 < <	**0.32** < <	**-0.56** >	0.01 >
	JJA	-0.02 < <	0.21 > >	0.19 < <	0.19 < <	**-0.63** >	-0.5 >
	SON	-0.03 < <	-0.15 < <	0.34 < >	**0.25** < <	**-0.89** <	0.07 >
Trop. S Amer. Land SAT	All	**0.26** > >	-0.07 < <	0.32 < <	0.3 < <	**-0.62** >	-0.6 >
	DJF	**0.4** > >	0.05 > >	0.31 < <	0.33 < <	-0.26 >	**-0.6** >
	MAM	**0.27** > >	0.04 > >	0.29 < >	0.34 > >	**-0.85** >	**-1.54** <
	JJA	0.21 > >	-0.12 < <	0.38 > >	0.37 > >	**-1.57** <	**-1.43** <
	SON	0.11 > >	-0.08 < <	0.46 > >	0.4 < <	**-1.23** <	**-0.9** >
NAO Index	DJF	0.11 > >	0.14 > >	**0.01** < <	0.02 < <	0.42 >	**-0.73** >
NW Euro Land Precip.	All	0.17 > >	-0.05 > <	**0.04** < <	0.04 < <	-0.2 >	-0.31 >
	DJF	0.01 < >	0.07 < >	0.07 < <	**0.06** < <	-0.41 >	-0.08 >
	MAM	-0 < <	-0.12 < <	0.07 > >	0.07 > >	**-1.42** <	**-1.16** <
	JJA	0.07 > >	0.18 > >	**0.08** < <	0.09 < <	-0.21 >	-0.47 >
	SON	0.05 > >	-0.18 < <	**0.07** < <	0.08 < <	-0.29 >	**-0.62** >

Table 2: Potential predictability measures for 5 year averaged climate indices. Averages are performed over all seasons (denoted All) or over individual seasons (denoted DJF, MAM etc.). Bold numbers indicate statistically significance at greater than the 5% level. For the ACC the > and < symbols indicate the relationship of the ACC to persistence and an AR(1) process respectively. The > and < symbols in the RMSE columns indicate the relationship of the RMSE to the control RMS and to an AR(1) process. The > and < symbols in the I column indicate the relationship to an AR(1) process.

		ACC	RMSE	*I*
Index	Season	1-10 years	1-10 years	1-10 years
Global Mean SAT	All	**0.57** > >	**0.05** < <	0.39 >
Gobal Mean Land SAT	All	**0.55** > >	**0.08** < <	0.25 >
Global Mean Ocean SAT	All	**0.5** < >	**0.04** < <	0.36 >
Tropical Atlantic SST	All	**0.4** > >	0.1 < <	-0.38 >
	DJF	**0.44** > >	**0.1** < <	-0.29 >
	MAM	**0.3** > >	**0.12** < <	-0.26 >
	JJA	**0.37** > >	0.12 < <	-0.72 >
	SON	**0.41** > >	**0.09** < <	-0.06 >
N Atlantic SST	All	**0.64** > >	**0.23** < <	0.4 >
	DJF	**0.56** > >	**0.27** < <	0.32 >
	MAM	**0.66** > >	**0.24** < <	0.53 >
	JJA	**0.64** > >	**0.24** < <	0.39 >
	SON	**0.46** > >	**0.19** < <	0.52 >
N Pacific SAT	All	0.18 > >	**0.2** < <	0.06 >
	DJF	**0.32** > >	**0.17** < <	0.16 >
	MAM	**0.31** > >	**0.18** < <	0.26 >
	JJA	0.09 > >	**0.24** < <	-0.04 >
	SON	0.09 < >	**0.18** < <	0.36 >
NW Euro Land SAT	All	0.25 < >	0.28 < <	-0.48 >
	DJF	0.13 > >	0.47 < <	-0.13 >
	MAM	0.18 < >	**0.28** < <	0.02 >
	JJA	0.16 < >	0.3 > >	**-1.31** <
	SON	**0.29** < >	0.28 < <	-0.18 >
Eurasian Land SAT	All	**0.48** < >	**0.17** < <	-0.06 >
	DJF	**0.43** < >	**0.33** < <	-0.24 >
	MAM	**0.27** < >	**0.27** < <	-0.5 >
	JJA	0.26 < >	0.14 < <	-0.55 >
	SON	**0.29** < >	**0.19** < <	-0.09 >
N American Land SAT	All	0.26 > >	**0.15** < <	0.21 >
	DJF	0.04 > >	**0.36** < <	-0.53 >
	MAM	0.02 > >	0.29 < <	-0.41 >
	JJA	0.17 > >	**0.13** < <	0.01 >
	SON	-0.15 < <	**0.16** < <	0.39 >
Trop. S Amer. Land SAT	All	0.04 > <	0.2 < <	-0.55 >
	DJF	**0.23** > >	0.22 < <	-0.56 >
	MAM	0.13 > >	0.2 > >	**-1.19** <
	JJA	0 > <	0.22 > >	**-1.28** <
	SON	-0.05 < <	0.25 < <	**-0.91** >
NAO Index	DJF	0.09 < >	**0.01** < <	-0.08 >
NW Euro Land Precip.	All	0.05 > >	0.03 < <	-0.28 >
	DJF	0.06 < >	0.04 < <	-0.3 >
	MAM	-0.12 < <	0.05 < <	**-1.27** <
	JJA	0.21 > >	0.07 < <	**-1.22** <
	SON	-0.1 > <	0.06 < <	-0.45 >

Table 3: As in table 2 but for 10 year averaged potential predictability measures.

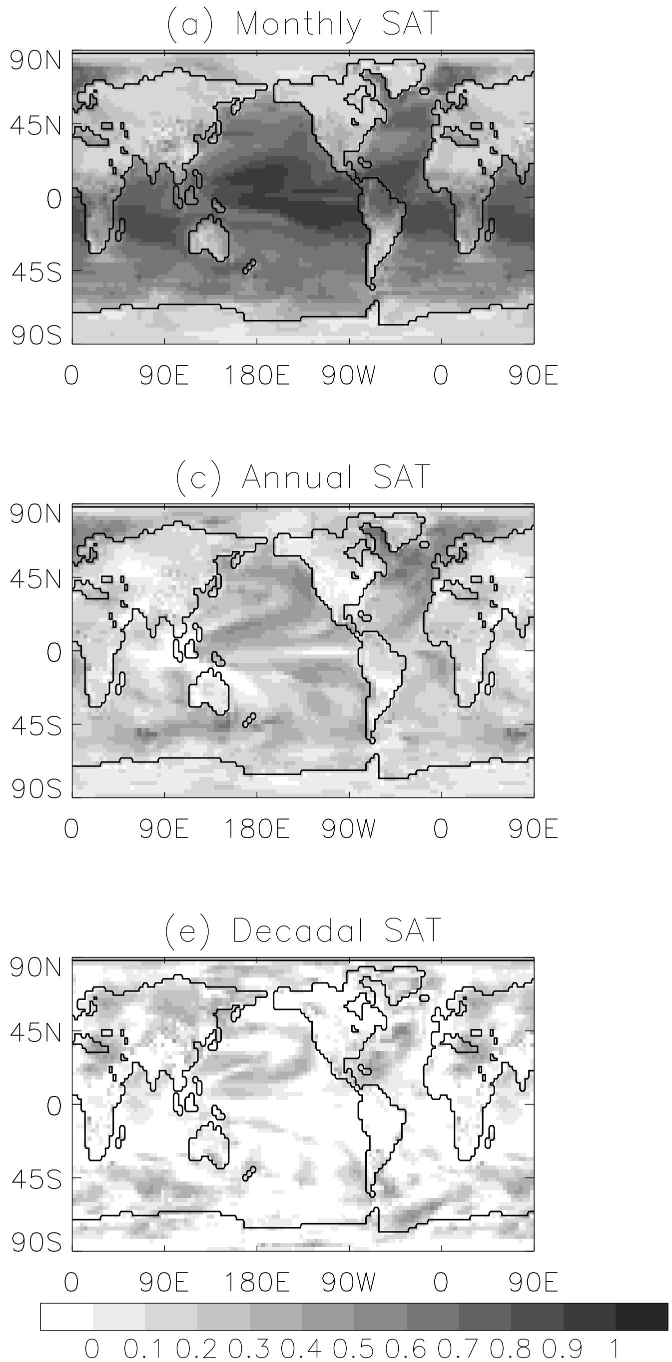

Figure 1: a_1 coefficients of an AR(1) process (see equation 2), equivalent to lag-1 auto-correlations) computed from fits to surface air temperature from 300 years of the control integration of HadCM3 in which concentrations of greenhouse gases etc. are held fixed. Before fitting the AR(1) model the model output was averaged into (a) monthly averages, (b) seasonal averages, (c) annual averages, (d) pentadal (i.e. 5 year) averages and (e) decadal averages. (f) shows the a_1 coefficients computed for global mean, global mean land and global mean ocean temperatures for monthly (M) means, seasonal (S) means, etc.

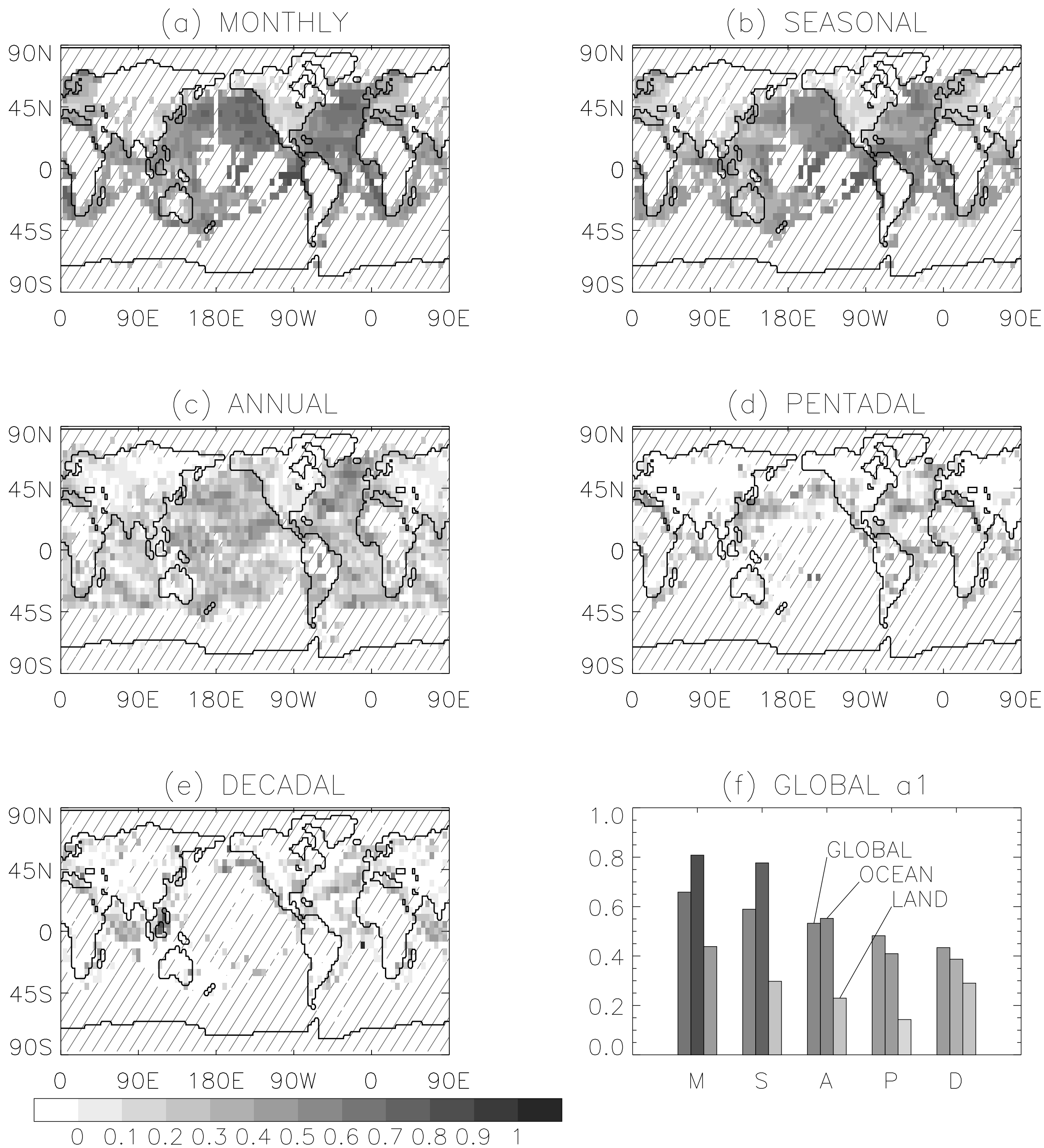

Figure 2: As in fig. 1 but computed from the observed land surface air temperatures and sea surface temperatures of Jones et al (1997). A third order polynomial trend was removed at each grid point prior to fitting the AR(1) model in order to remove the influence of climate change. Diagonal shading denotes regions in which there is not enough data to perform the AR(1) fit.

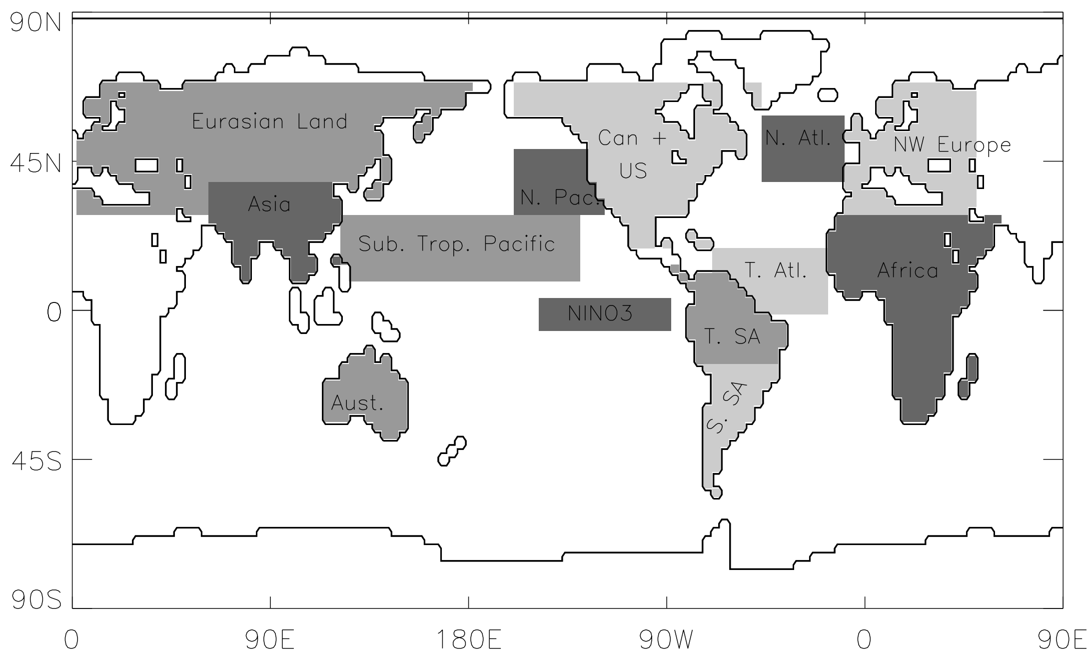

Figure 3: Graphical definitions of some of the regions used to define climate indices. See table 1 and section 4.2 for more details.

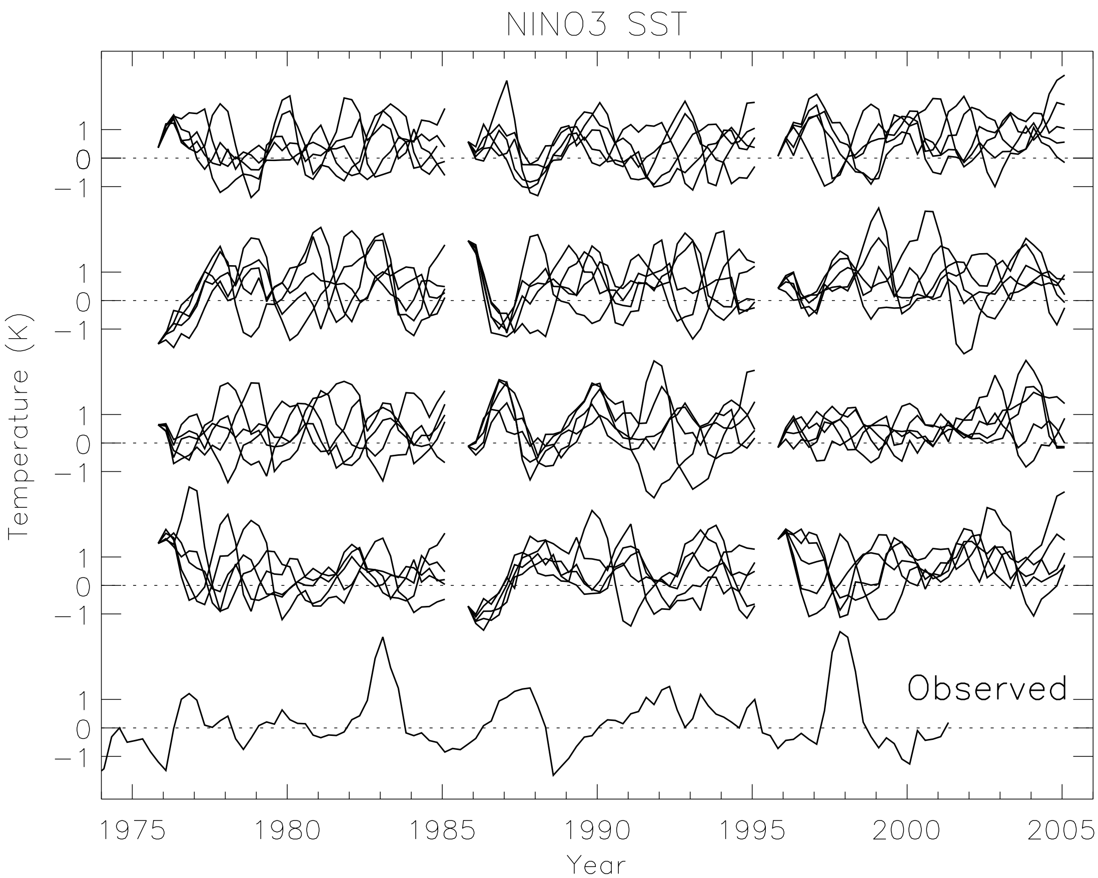

Figure 4: NINO3 anomalies from the ensemble experiments used in this study. Each of the 12 ensemble experiments has 5 members in which the ocean initial states are identical and atmospheric initial states differ by one day. The observed NINO3 anomalies are shown only to indicate the relative realism of the model ENSO cycle.

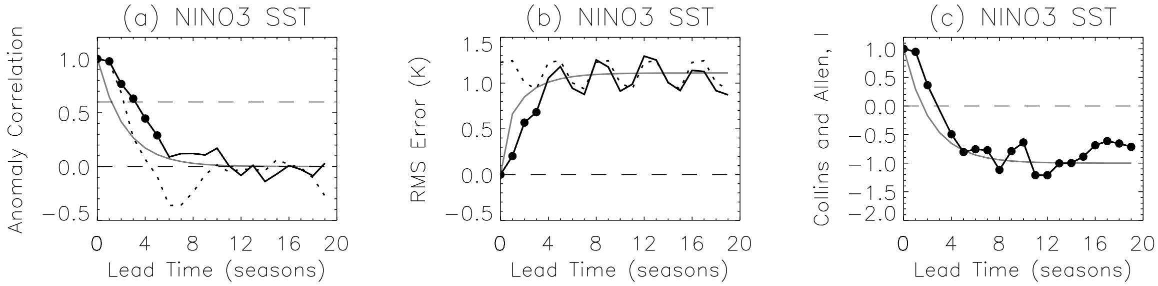

Figure 5: Measures of potential predictability of seasonal mean NINO3 anomalies averaged over all ensemble experiments. (a) Anomaly correlation coefficients for the perfect model ensembles (black solid line), for a persistence forecast (black dotted line) and for a fitted AR(1) process (grey solid line). (b) Root Mean Squared Error for the perfect model ensembles (black solid line) and for a fitted AR(1) process (grey solid line). The dotted line shows the climatological RMS computed from the model control experiment. (c) the predictability measure of Collins and Allen (2002) for the perfect model ensembles (solid black line) and a fitted AR(1) process (solid grey line). The black dots indicate statistical significance for the particular measure at the greater than 5% level.

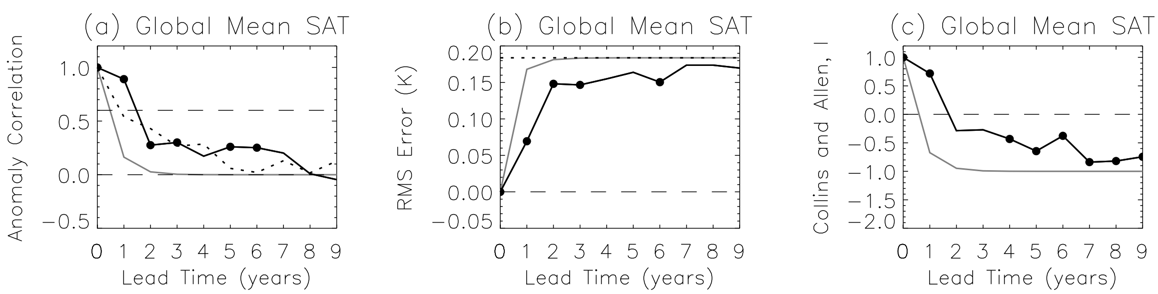

Figure 6: As in fig. 5 but for annually averaged global mean surface air temperature.

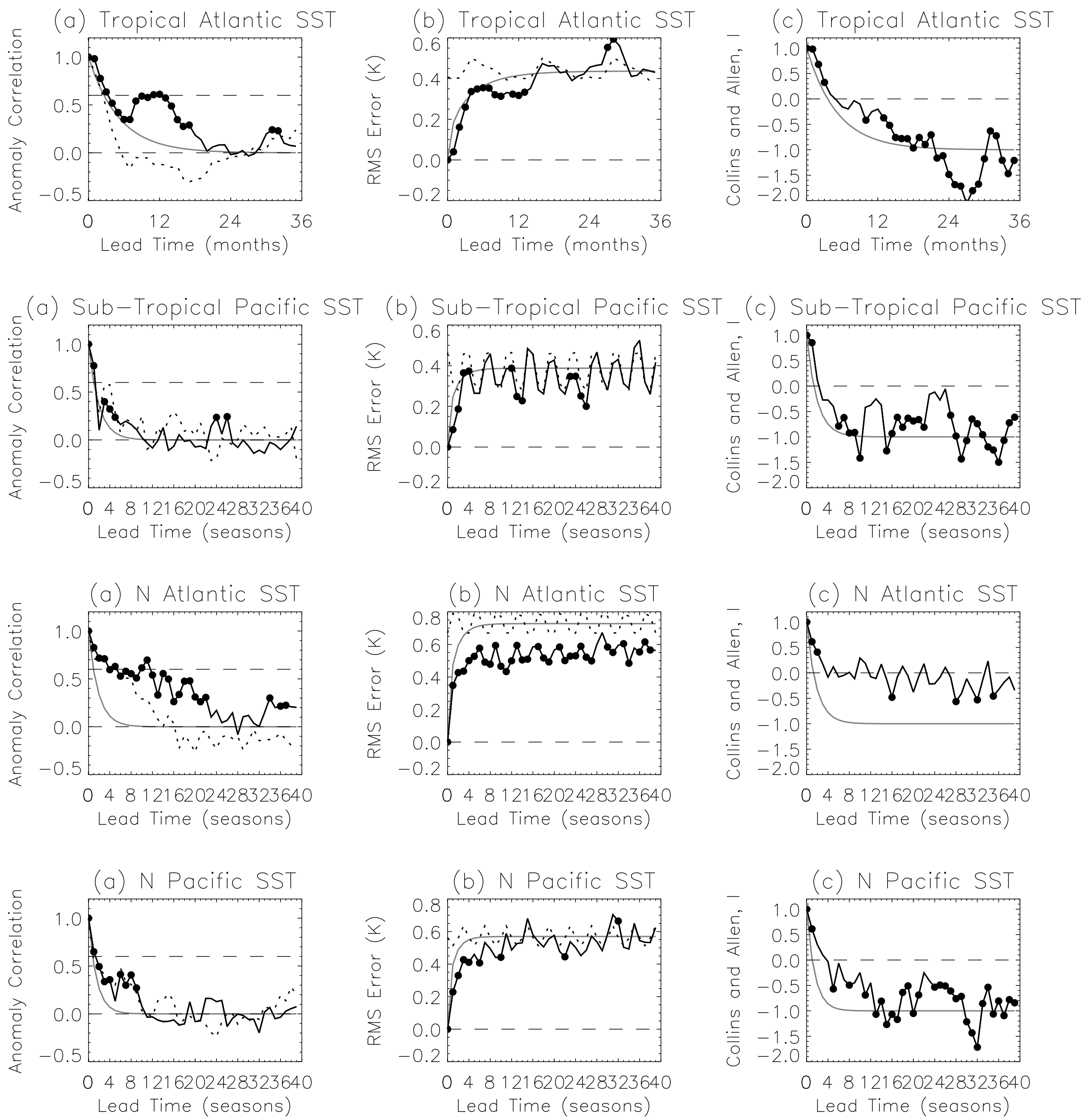

Figure 7: As in fig. 5 but for seasonally averaged SAT over the tropical Atlantic Ocean (top row), the sub-tropical Pacific Ocean (second row), the North Atlantic Ocean (third row) and the North Pacific Ocean (bottom row). See table 1 and fig. 3 for definitions of these indices.

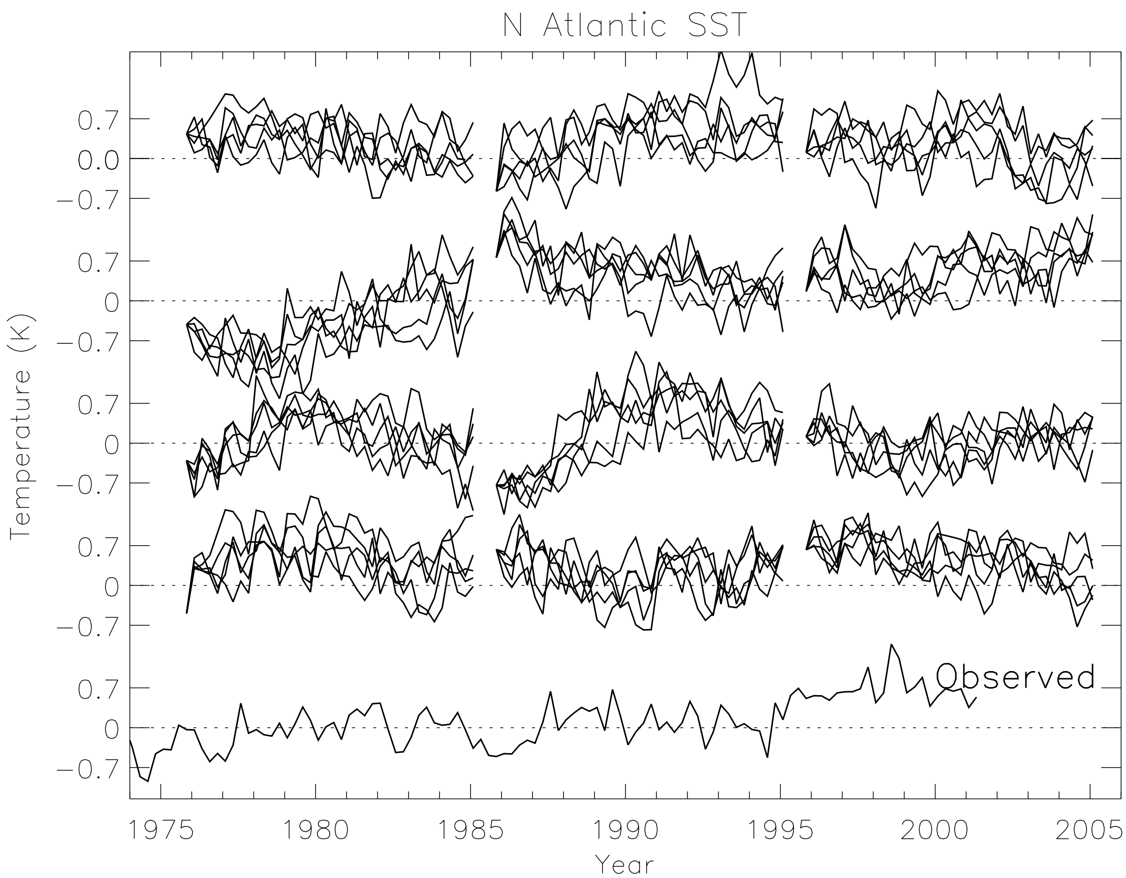

Figure 8: As in fig. 4 but for Surface Air Temperature anomalies in the region 50°W-10°W, 40°N-60°N in the North Atlantic Ocean.

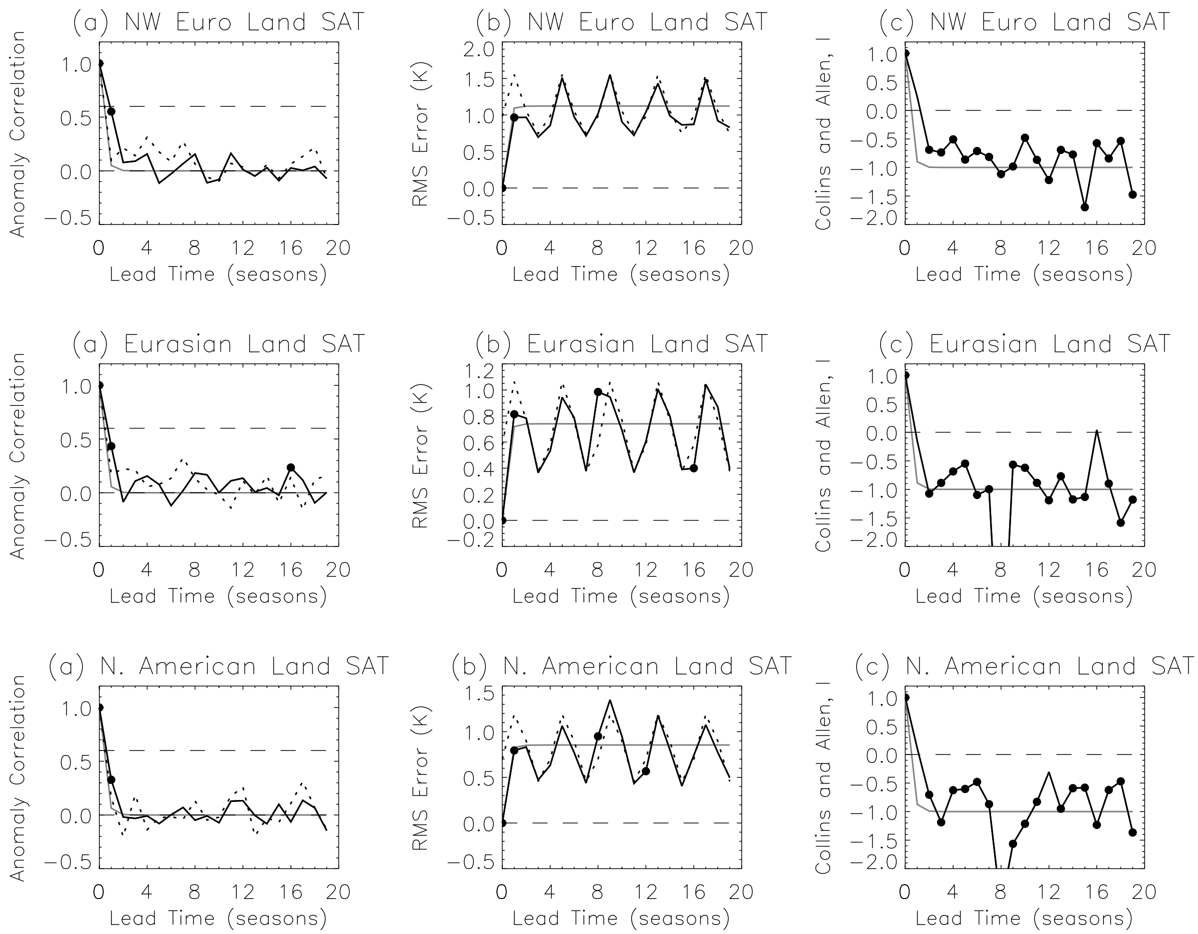

Figure 9: As in fig. 5 but for seasonally averaged SAT over the North Western European land area, the Eurasian continent and the United States and Canada region.

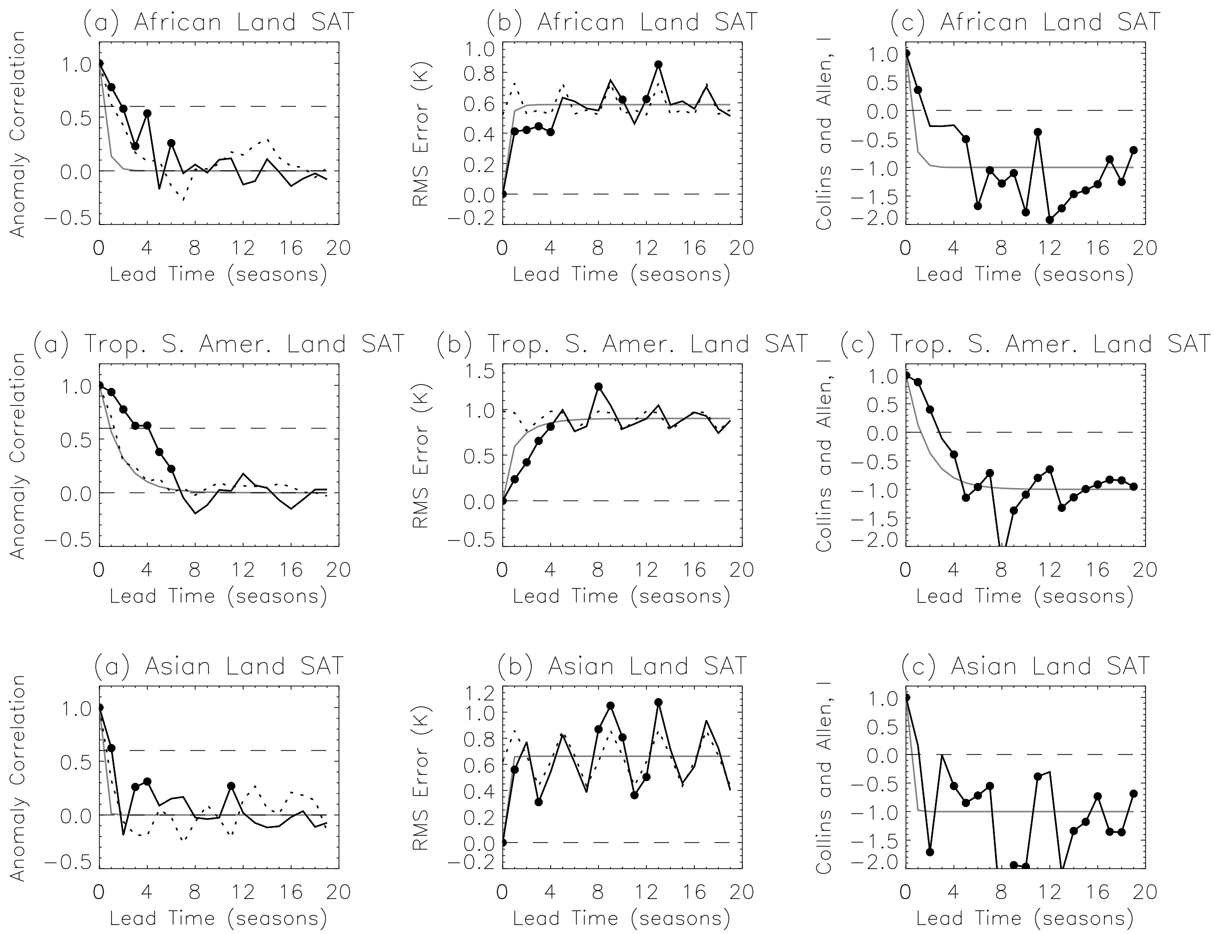

Figure 10: As in fig. 5 but for seasonally averaged SAT over the African continental region, the tropical South American land region and the Asian land region.

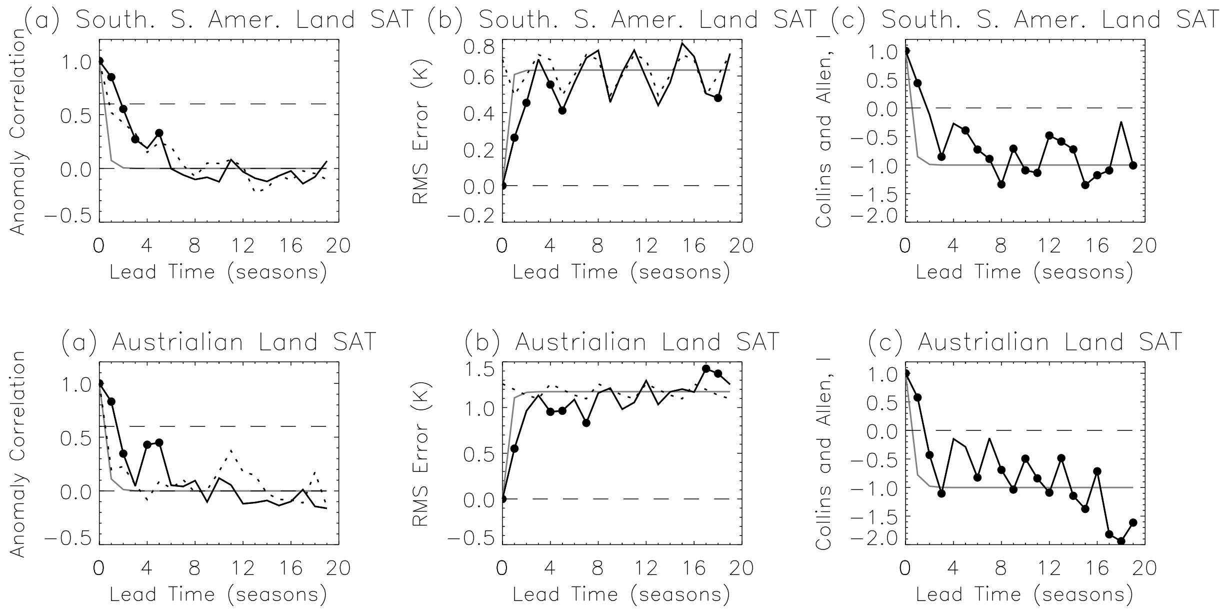

Figure 11: As in fig. 5 but for seasonally averaged SAT over the Southern South American region and over Australia.

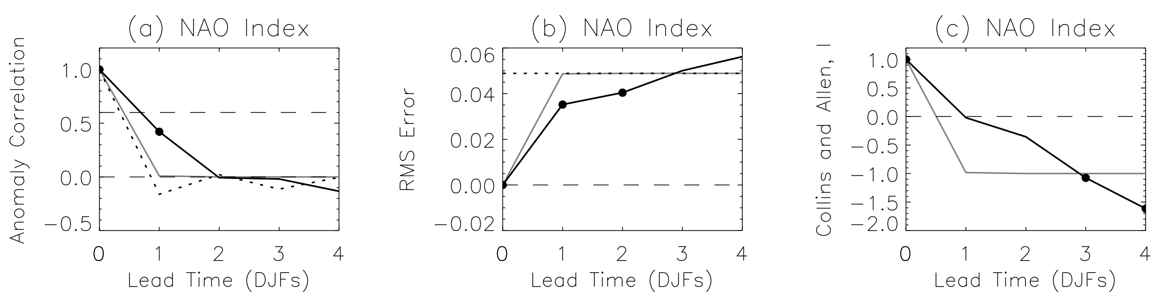

Figure 12: As in fig. 5 but for the December-February EOF index of the North Atlantic Oscillation.

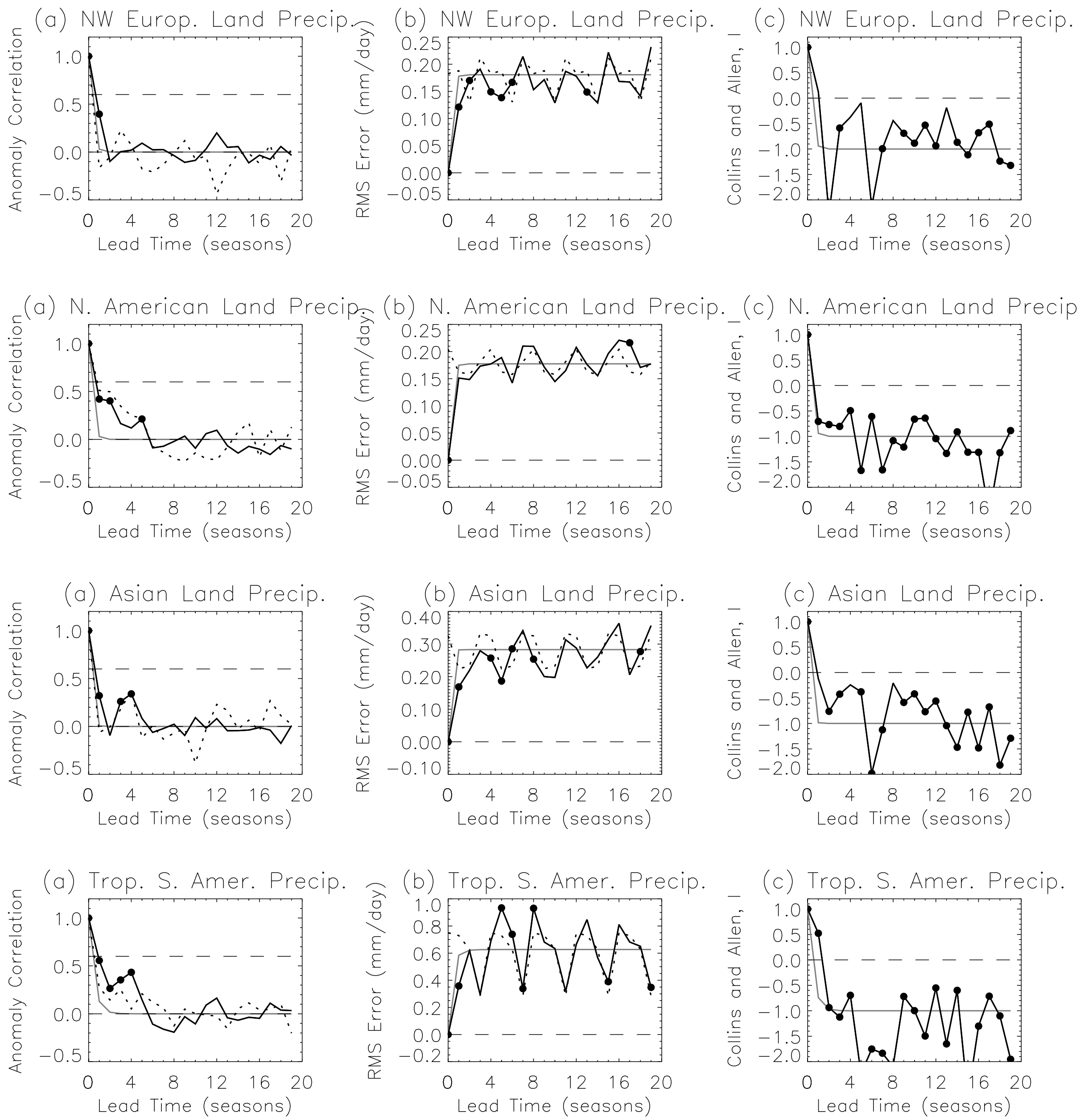

Figure 13: As in fig. 5 but for monthly averaged precipitation over North Western Europe land, United States and Canada, the Asian land area and the Tropical South American region.

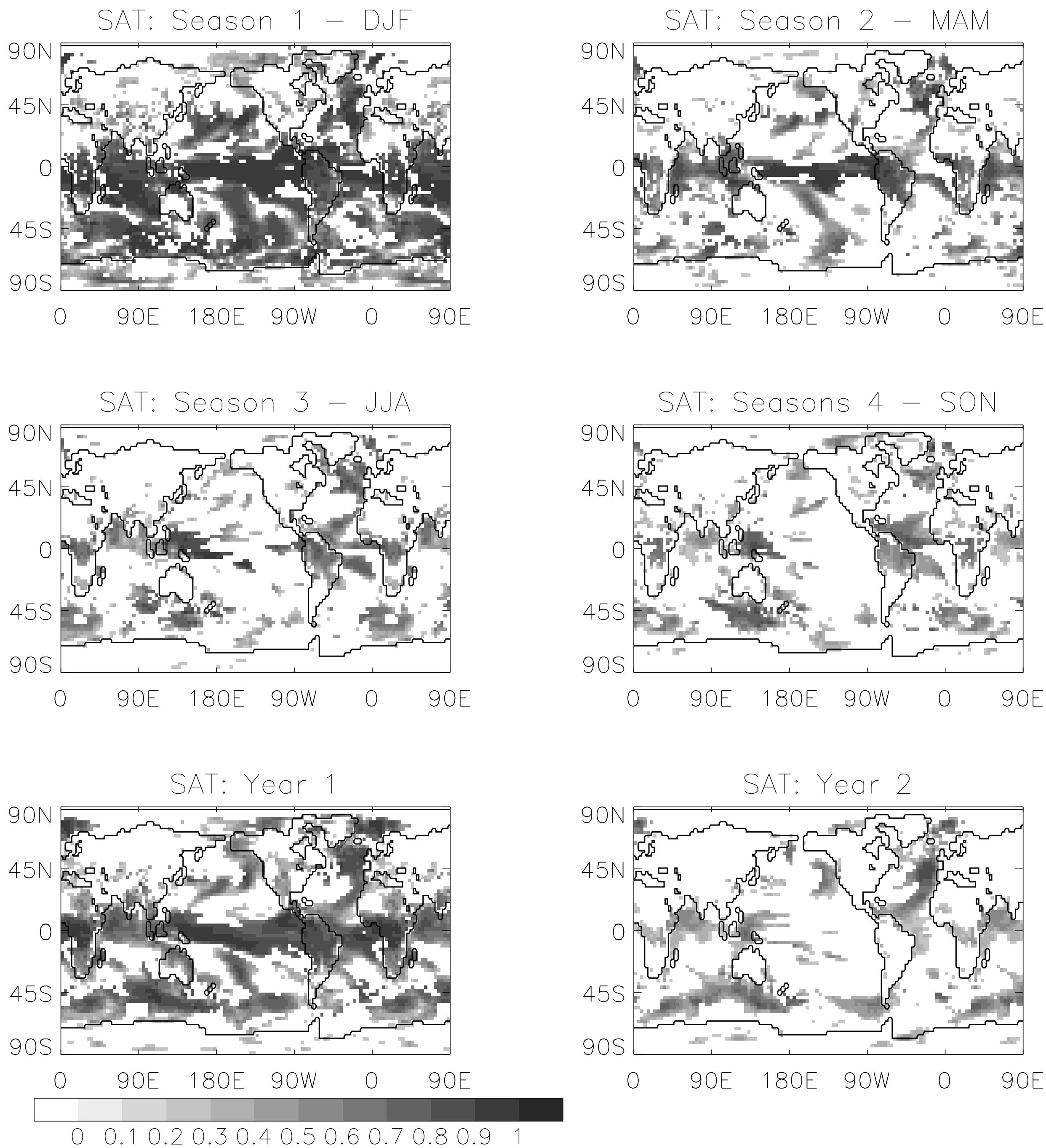

Figure 14: Anomaly correlation coefficients computed over all ensemble experiments at each model grid point for average SAT. Values are only shaded when they are statistically different from zero and the ACC is greater than that for a persistence forecast.

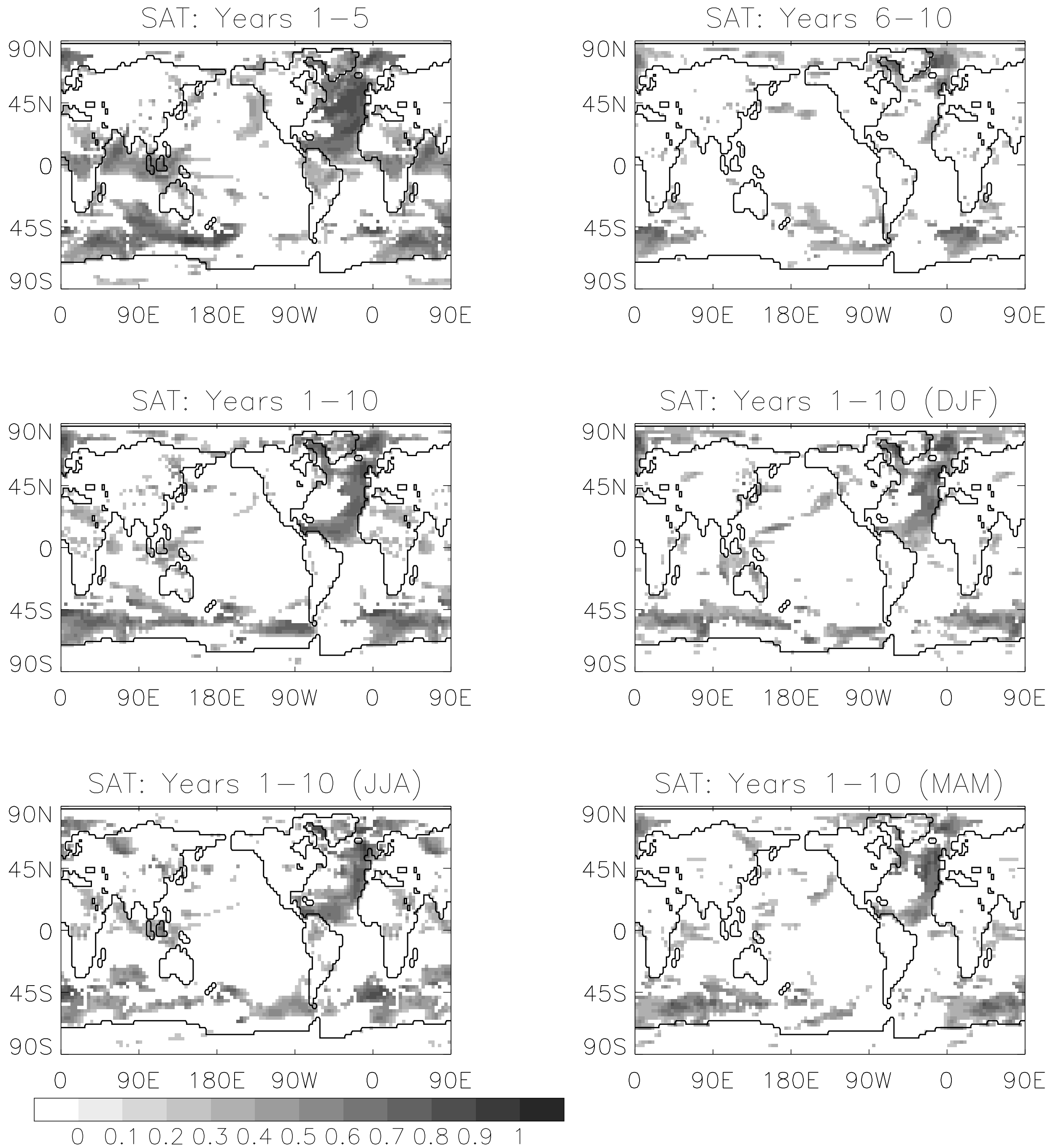

Figure 15: As in fig. 14 for different averaging periods.

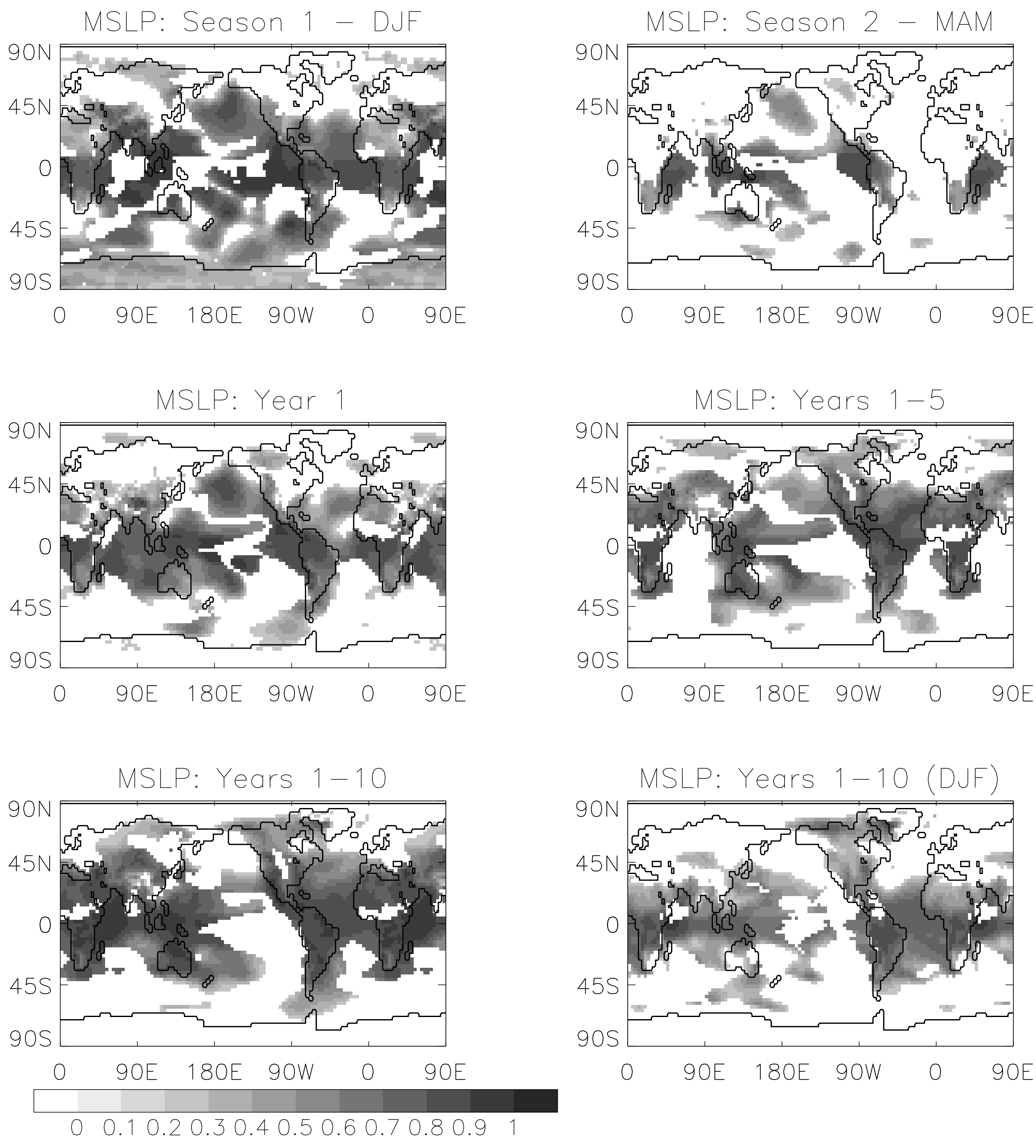

Figure 16: As in fig. 14 but for mean sea level pressure.

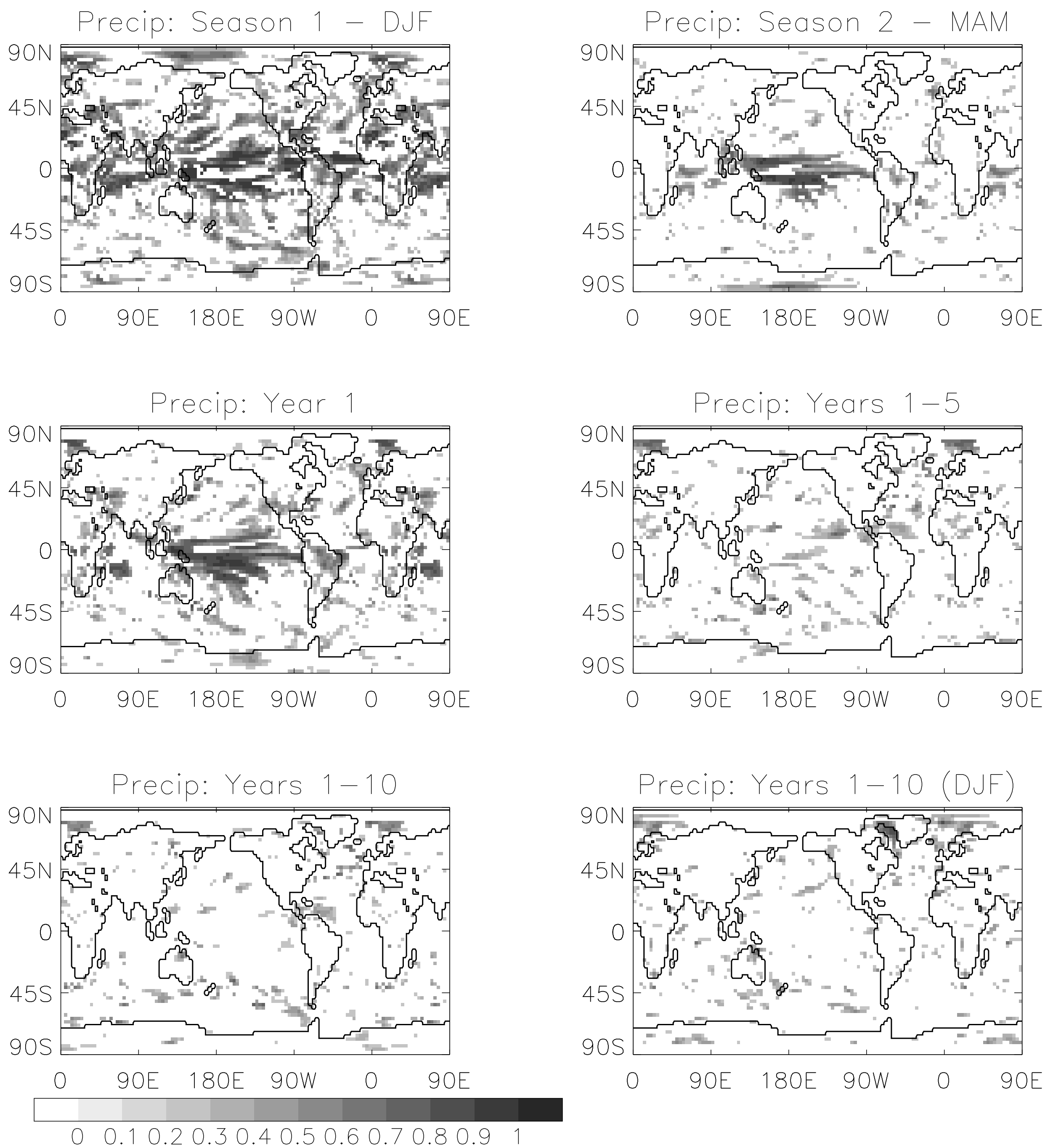

Figure 17: As in fig. 14 but for precipitation.

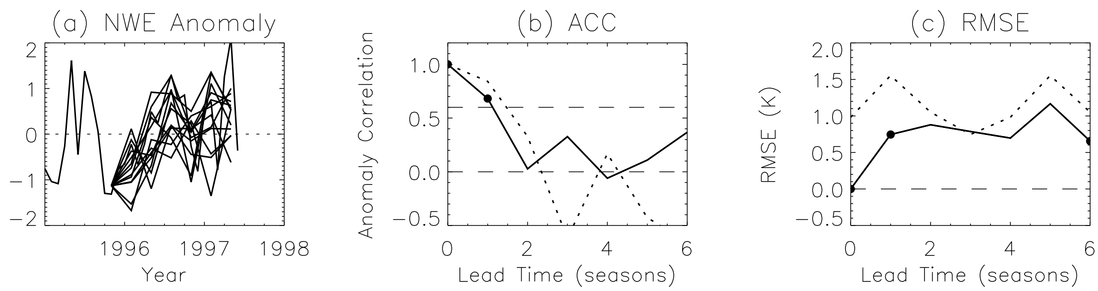

Figure 18: An example of state-dependent potential climate predictability. (a) North West European SAT anomalies for an ensemble of 12 simulations in which the initial conditions are anomalously cold. (b) ACC and (c) RMSE measures for this ensemble experiment only. The potential predictability is greater than the average potential predictability shown in fig. 9.