Climate Dynamics (1998) 14:517–528

F. W. Zwiers · V. V. Kharin

Intercomparison of interannual variability and potential predictability: an AMIP diagnostic subproject

Received: 7 July 1996 / Accepted: 8 January 1998

Abstract. The inter-annual variability and potential predictability of 850 hPa temperature (T_{850}), 500 hPa geopotential (ϕ_{500}) and 300 hPa stream function (ψ_{300}) simulated by the models participating in the Atmospheric Model Intercomparison Project (AMIP) are examined. The total inter-annual variability is partitioned into a potentially predictable component which arises from the forcing implied by the prescribed SST and sea-ice evolution, or from sources internal to the simulated climate, and an unpredictable low frequency component induced by "weather noise". There is wide variation in the ability to simulate observed inter-annual variability, both total and weather-noise induced. A majority of models under simulate seasonal mean ϕ_{500} variability in DJF and JJA and over simulate ψ_{300} variability in JJA. All but one model simulates less T_{850} total inter-annual variability than in the analysed data. There is little apparent connection between gross model characteristics and the corresponding ability to simulate observed variability, with the possible exceptions of resolution.

1 Introduction

There has recently been substantial interest in using general circulation models (GCMs) to make multi-season dynamical climate forecasts (Dix and Hunt 1995; Rowell 1998; Zwiers 1996; Kumar et al. 1996; Kumar and Hoerling 1995; Stern and Miyakoda 1995; Anderson and Stern 1996). Such seasonal predictions are justified by the fact that much of the variability of the climate system on seasonal and longer time scales arises from interactions with the land, sea and ice covered surface of the Earth. Slowly varying boundary conditions, such as sea surface temperature, or sea-ice extent, may produce predictable variations in seasonal mean quantities. Also, the internal dynamics of the atmosphere, or slowly varying surface properties such as soil moisture and snow cover, may generate potentially predictable interannual variability. Additional variability on seasonal and longer time scales arises from daily weather which cannot be predicted in detail longer than two weeks in advance. Thus the inter-comparison of the inter-annual variability from these sources may provide us with useful insights into the predictability of the climate system on seasonal to interannual time scales.

It is therefore of general interest to examine the ability of current GCMs to simulate variability from potentially predictable and non-predictable sources, and to examine the relative magnitude of these sources of variability. Climates in which a large fraction of the interannual variance of seasonal means originates from sources other than high frequency weather (also known as weather noise) are often described as being "potentially predictable" (Madden 1976; Zwiers 1996).

The purpose of this study is to inter-compare the inter-annual variability and potential predictability of seasonal means of the climates simulated by the models participating in the Atmospheric Model Intercomparison Project (AMIP) (Gates 1992). The AMIP experimental protocol describes a 10-y climate simulation which is forced with "observed" sea-surface temperature and sea ice extent for the January 1979 to December 1988 period. We will focus primarily on 850 hPa temperature (T_{850}), 500 hPa geopotential (ϕ_{500}) and 300 hPa stream function (ψ_{300}) in this study. The 850 hPa surface is reasonably near the surface and variations in T_{850} are indicative of thermal variability generated either locally or in response to circulation changes. The 500 hPa geopotential is the classical

F. W. Zwiers (✉) · V. V. Kharin[1]
Canadian Centre for Climate Modelling and Analysis,
Victoria, B.C., Canada

Also affiliated with:
[1] Department of Mathematics and Statistics,
University of Victoria, Victoria, B.C., Canada V8W 2Y2

mid-tropospheric variable that represents the mid-tropospheric flow in the extra-tropics. In the tropics ϕ_{500} responds to variations in the distribution of atmospheric mass and to the mean temperature of the lower troposphere. Various mechanisms which determine the atmospheric response to SST anomalies in the tropics and extratropics are discussed by Webster (1981) and many others. The atmosphere is dynamically most active and eddy kinetic energy is greatest at about 300 hPa. The 300 hPa stream function ψ_{300} measures the intensity of the rotational flow at this level and carries roughly the same information as geopotential in the extra tropics. In the tropics, ψ_{300} represents the rotational dynamics. It responds to variations in boundary forcing that change the vertical motion and thereby induce rotational flows such as the anticyclonic and cyclonic dipoles associated with anomalies in the ascending and descending branches of the Walker circulation during ENSO episodes (Boer 1985; Gill 1980).

By developing suitable analysis techniques we have been able to restrict our data needs to only the standard monthly mean AMIP data products and to daily analyses of the observed climate system for the AMIP period. Consequently, we are able to inter-compare the simulations submitted by all 30 AMIP participants. The methodology used is described in detail by Zwiers (1996, cited as Z96 from here on).

The observed data used in this study consist of daily NMC analyses of T_{850}, ϕ_{500} and the 300 hPa winds for the AMIP decade, and monthly means of ECMWF analysed T_{850} for 1980–88. Daily ψ_{300} was derived from the analysed winds. Monthly means of NMC analysed ϕ_{500} and ψ_{300} were computed from the daily analyses. The monthly means of NMC analysed T_{850} were not used because they have inflated interannual variances. A variety of changes were made to the NMC analysis system in late May, 1986 (Ropelewski, personal communication) which resulted in a change in global mean T_{850} of approximately 2 °C. The data are described more fully in Z96.

The calculations for this study have been performed in a number of steps.

1. All data sets were interpolated to the 96×48 Gaussian grid used by the Canadian Climate Centre 2nd generation general circulation model (CCC GCM2; McFarlane et al. 1992) so that local data manipulation, diagnostic and graphics tools could be used.
2. A high terrain mask was derived for T_{850} by computing the union of masks provided by groups that supplied masks with their data. In essence, this derived mask represents the spatial resolution of the lowest resolution grid point model. The mask was used in all calculations involving T_{850}.
3. The annual cycle was removed separately from each of the 31 data sets used in the analysis of each variable (i.e. the observations and the 30 AMIP simulations).
4. The inter-annual variance of seasonal mean T_{850}, ϕ_{500} and ψ_{300} was computed in the standard way for each data set using the standard definitions of the seasons (i.e. DJF, MAM, JJA and SON). We refer to these variance estimates as *total variances* to distinguish them from the "weather noise" induced variance component discussed below. No data collected or simulated prior to June, 1979 is used in these, or any other, calculations. That is, the first five months of every simulation is discarded to allow the land surface processes to equilibrate with the prescribed SSTs and sea-ice boundaries. Even so, the equilibration process may not have completed within 5 months (Z96 and Robock et al. 1996), particularly in the Northern Hemisphere where the initial distribution of frozen soil moisture can persist for several months in some places.
5. Seasonal mean inter-annual variance arises from a number of sources (see Z96), including the forcing implied by the prescribed evolution of SST and sea-ice, sources internal to the simulated climate (such as a simulated atmosphere's internal dynamics and the interaction between it and the land surface), and low frequency variability induced by day-to-day changes in the simulated "weather". The latter is often referred to as "weather noise" (Madden 1976; Zwiers 1987). We estimated the weather noise induced inter-annual variance component using a technique that will be briefly described in Sect. 2. We will refer to this variance component as the weather noise *induced inter-annual variance* to distinguish it from the total inter-annual variance of the seasonal mean.
6. Global means of the induced and total inter-annual variances for DJF and JJA were computed. These variance indices were intercompared and attempts were made to relate index values to model characteristics.
7. The ratio between the total and induced variance is an indicator of the potential predictability of a climate. This ratio was computed and intercompared. The weather noise induced variance is inherently non-predictable. However, when the total variance is significantly greater than the estimated weather noise induced variance, the excess variance may be predictable. Z96 shows that much of the excess variance simulated by CCC GCM2 is attributable to the prescribed SST/sea-ice forcing.

Sample size imposes an important constraint on this study. AMIP samples are small, consisting of 9 northern winters and 10 summers, and only one realisation per model was available. Consequently we expect substantial sampling variability of the variance estimates. Nonetheless, we believe that the analysis is useful and revealing.

The outline of the remainder of this work is as follows. The method used to estimate the weather noise induced variance is described briefly in Sect. 2 as is the

test which is used to determine whether the total variance contains a potentially predictable component. The global averages of total and induced variance are discussed in Sect. 3. Potential predictability is discussed in Sect. 4. Conclusions are drawn in Sect. 5.

2 Estimating the weather noise induced interannual variance

The weather noise induced component of the interannual variance of a seasonal mean is estimated by means a time domain analysis of variance (see Z96; Zwiers 1987; or Madden 1976) which we now briefly describe.

First, to establish notation, let $\mathbf{X}_{yt}$ represent anomalies from the annual cycle for the variable of interest where $y = 1, \ldots, Y$ indexes years and $t = 1, \ldots, T$ indexes time within a specific season. Previous "potential predictability" studies (e.g. Madden 1976; Zwiers 1987) have used daily data to compute the weather noise induced interannual variance. Here we wish to use monthly mean data for this purpose to avoid the daunting task of obtaining daily data from each of the 30 AMIP simulations. For now, however, we leave the unit of time unspecified except to note that t marks time *within* a specific season, such as DJF.

The statistical model which is implicit in potential predictability studies is a 1-way ANOVA model which may be written as

$$\mathbf{X}_{yt} = \mu + \beta_y + \varepsilon_{yt}. \tag{1}$$

Here μ represents the long term seasonal mean, the β_y's represent year-to-year variations in the level of $\mathbf{X}$ that are potentially predictable, and the ε_{yt}'s represent within season variations that are presumably not predictable on seasonal and longer time scales.

It is necessary to make several assumptions about this model. First we state three assumptions that are implicit in all potential predictability studies of this type, even when the ANOVA model (1) is not explicitly stated. An additional fourth assumption is required in this study because daily data are not available.

1. The random variables β_y, $y = 1, \ldots, Y$, represent potentially predictable climate signals generated from both external sources and internal sources. The external sources (e.g. SST/sea-ice forcing which is prescribed identically in every AMIP simulation) may generate a potentially predictable signal, primarily in the tropics where the slowly varying lower boundary conditions are important. The atmospheric model may also generate potentially predictable variations internally from, for example, interactions with land surface processes by the atmosphere's internal dynamics.

   We assume β_y, $y = 1, \ldots, Y$, to be independently distributed Gaussian random variables with mean zero and variance σ_β^2. The Gaussian assumption would appear to be reasonable for seasonal mean quantities. On the other hand, it seems clear that both the independence and constant variance assumptions are at best approximations. The independence assumption is in question, for example, because the ocean varies on time scales much longer than a year and because the response to a volcanic event often persists for longer than a year. The constant variance assumption may be in question, for example, because climate variations during and after a volcanic event might have different amplitude than variations at other times.
2. The weather noise time series $\{\varepsilon_{yt}, t = 1, \ldots, T\}$, $y = 1, \ldots, Y$ are assumed to be independent realizations of the same Gaussian stochastic process. Again, we might question both the independence assumption and the assumption that the stochastic properties of $\{\varepsilon_{yt}, t = 1, \ldots, T\}$ do not vary from year to year.
3. The weather noise process is assumed to be independent of the potential predictable process $\{\beta_y, y = 1, \ldots, Y\}$. Again, it seems obvious that this is an approximation.
4. Each weather noise time series $\{\varepsilon_{yt}, t = 1, \ldots, T\}$ is assumed to behave as red noise. This assumption is stronger (and thus less robust) than Madden's (1976) low-frequency white noise assumption because it prescribes a specific form for the power spectrum of the weather noise at all frequencies. Madden assumes that the power spectrum of the weather noise is continuous at the origin so that the spectrum is flat at zero frequency (Zwiers 1987). Implicitly Madden also assumes that the curvature of the power spectrum is low at the origin so that the power at zero frequency can be estimated from daily variation within seasons with relatively little bias. Madden's assumptions would appear to be quite reasonable for the variables we consider. The stronger assumption is required so that the weather noise induced variance can be estimated from the standard AMIP monthly mean data products.

Clearly there are limitations to the utility of this model. Nonetheless, departures from these assumptions would appear to be mild enough that model (1) remains a useful device for partitioning variance into potentially predictable and non-predictable components. We now discuss how this model is used for this purpose.

The total inter-annual variance of the seasonal mean of $\mathbf{X}_{yt}$ is estimated by

$$\hat{\sigma}^2_{X_{y\circ}} = \frac{1}{(Y-1)} \sum_{y=1}^{Y} (\mathbf{X}_{y\circ} - \mathbf{X}_{\circ\circ})^2. \tag{2}$$

Here $\mathbf{X}_{y\circ}$ denotes the year y seasonal mean (i.e. the "circle" in the subscript indicates that an average has been taken over the index which it represents). Similarly, $\mathbf{X}_{\circ\circ}$ denotes the mean of the seasonal means. Substituting model (1) into Eq. (2) we see that

$$\begin{aligned}
\hat{\sigma}^2_{X_{y\circ}} &= \frac{1}{(Y-1)} \sum_{y=1}^{Y} (\mathbf{X}_{y\circ} - \mathbf{X}_{\circ\circ})^2 \\
&= \frac{1}{(Y-1)} \sum_{y=1}^{Y} (\beta_y + \varepsilon_{y\circ} - (\beta_\circ + \varepsilon_{\circ\circ}))^2 \\
&= \frac{1}{(Y-1)} \left(\sum_{y=1}^{Y} (\beta_y - \beta_\circ)^2 + \sum_{y=1}^{Y} (\varepsilon_{y\circ} - \varepsilon_{\circ\circ})^2 \right. \\
&\qquad \left. + 2 \sum_{y=1}^{Y} (\beta_y - \beta_\circ)(\varepsilon_{y\circ} - \varepsilon_{\circ\circ}) \right).
\end{aligned}$$

The first term in this expression is an unbiased estimator of the variance of β_y and the second term is an unbiased estimator of the variance $\sigma^2_{\varepsilon_{y\circ}}$ of the seasonal mean weather noise. Since the expectation of the cross-product term is zero as a consequence of our independence assumptions, we obtain

$$\mathscr{E}(\hat{\sigma}^2_{X_{y\circ}}) = \sigma_\beta^2 + \sigma^2_{\varepsilon_{y\circ}}. \tag{3}$$

Hence the total inter-annual variance of the seasonal mean in model (1) contains climate signal and weather-noise components.

Equation (3) indicates that we can determine whether the observed process is potentially predictable by testing the null hypothesis that the seasonal means of $\mathbf{X}$ vary only because of weather noise. That is, we need to test

$$H_0\colon \sigma_\beta^2 = 0. \tag{4}$$

To conveniently perform the test we require an estimate of the variance of the time average of the weather noise $\sigma^2_{\varepsilon_{y\circ}}$ that is statistically independent of the total variance.

When daily data are available, a method first described by Madden (1976) is often used to estimate $\sigma^2_{\varepsilon_{y\circ}}$ by assuming that the weather noise induced variance is well approximated by $f_{\varepsilon\varepsilon}(0)/T$ where $f_{\varepsilon\varepsilon}(\lambda)$ is the spectral density function of the daily weather noise process.

When only monthly means are available, a method proposed by Z96 can be used. Z96 showed that an unbiased estimate of $\sigma^2_{\varepsilon_{y\circ}}$ can be made by computing

$$\hat{\sigma}^2_{\varepsilon_{y\circ}} = SSE \, \frac{3 + 4\rho_1 + 2\rho_2}{6Y(3 - 2\rho_1 - \rho_2)} \tag{5}$$

where SSE is the sum of squared within season deviations

$$SSE = \sum_{y} \sum_{t=1}^{3} (\mathbf{X}_{yt} - \mathbf{X}_{y\circ})^2$$

and ρ_1 and ρ_2 are the lag-1 month and lag-2 month auto-correlations of the monthly mean weather noise process. SSE is statistically independent of $\hat{\sigma}^2_{X_{y\circ}}$ when assumptions 1–3 hold. Thus $\hat{\sigma}^2_{\varepsilon_{y\circ}}$ will provide the necessary estimate of the weather noise induced inter-annual variance that is required to test null hypothesis (4) provided that suitable estimates of ρ_1 and ρ_2 can be found for utilisation in (5).

There are several possibilities for estimating ρ_1 and ρ_2. The obvious approach is to estimate these correlations from the within season monthly deviations $\mathbf{X}_{yt} - \mathbf{X}_{y\circ}$. Unfortunately, this approach results in estimates with very large bias because the within season deviations are algebraically related; the three deviations within each season must sum to zero. Another approach is based on the observation that variation in the monthly anomalies $\mathbf{X}_{yt}$ is due only to weather noise when null hypothesis (4) holds. Correlations computed from the monthly anomalies are very nearly unbiased under the null hypothesis. However, Monte Carlo simulations reveal that the estimates obtained in this way are subject to considerable sampling variability. Unfortunately, this uncertainty compromises the test described later. To avoid this problem, a third approach that estimates ρ_1 and ρ_2 indirectly from daily data is used.

The lagged correlations are estimated as follows. First, the mean annual cycle is removed from the NMC analyses. Then, for each season, the seasonal mean is subtracted. Next, it is assumed that the resulting deviations behave as a Markov process (i.e. as an autoregressive process of order one). The Markov process is fitted by estimating the lag 24-h autocorrelation ρ of the deviations. The estimated ρ is subject to very little sampling variability because the number of days available in the sample is large. The auto-correlation function of the monthly mean deviations is then deduced algebraically from the fitted Markov model.

The estimate $\hat{\sigma}^2_{\varepsilon_{y\circ}}$ obtained in this way is now used to test null hypothesis (4) by computing

$$F = \hat{\sigma}^2_{X_{y\circ}} / \hat{\sigma}^2_{\varepsilon_{y\circ}} \tag{6}$$

and comparing it against critical values from the F-distribution which best approximates that of F when null hypothesis (4) is true. Z96 shows that this approximating distribution has $(Y-1)$ and $T^*(Y-1)$ degrees of freedom where

$$T^* = \frac{2(3 - 2\rho_1 - \rho_2)^2}{9 - 12\rho_1 + 8\rho_1^2 - 6\rho_2 - 4\rho_1\rho_2 + 5\rho_2^2}. \tag{7}$$

The ρ_1 and ρ_2 estimated from NMC analyses were used in Eqs. (5) and (7) for all models. This approach is justified by the fact that the estimates of $\hat{\sigma}^2_{\varepsilon_{y\circ}}$ and T^* are relatively insensitive to the lag-1 day correlation coefficient ρ, unless day-to-day dependence is very strong. In particular, $1.95 \leq T^* \leq 2$ for all $\rho < 0.9$. This is true over the most of the globe for all variables used in this study, when ρ is estimated from the NMC analysed daily data. We find estimates of $\rho > 0.9$ only in some areas at high-northern latitudes in the DJF season for ϕ_{500} and ψ_{300}. The factor that scales SSE in Eq. (5) to obtain the estimate of $\hat{\sigma}^2_{\varepsilon_{y\circ}}$ is also relatively insensitive to ρ for small to moderate day-to-day correlations.

Z96 used the same technique to make decisions about his null hypothesis H_I that there is no potential predictability from internal sources of variability in an ensemble of AMIP simulations performed with CCC GCM2. He found that very similar decisions were made when F was computed as above using lag-1 day auto-correlations computed from either the daily NMC analyses or the daily model output. Also these decisions were almost identical to those obtained directly from the daily data with Madden's (1976) method. Thus we assume that modest errors in the estimated lag-1 day correlation will not have a substantial effect on the decisions made with F.

3 Global averages of total and induced variance

The global averages of the total and weather noise induced inter-annual variances of DJF and JJA mean T_{850}, ϕ_{500} and ψ_{300} simulated by the 30 AMIP models are displayed in Fig. 1 as a proportion of the corresponding "observed" quantity. Several things can be noted about Fig. 1.

1. With the exception of one outlier, the total T_{850} variance simulated by the models in both JJA and DJF is less than that which is contained in the corresponding ECMWF analyses. A number of factors may be responsible for this. Part of the explanation is likely variability induced in the analysed temperatures by analysis system changes. Analysis error is likely also a contributor. Trenberth (1992) describes the ECMWF analyses. Other possible factors include:
   a. The simulations see only SST variability on monthly and longer time scales.
   b. Some low frequency variance may be missing depending upon how monthly SSTs are interpolated (Sheng and Zwiers 1998).
   c. There may be errors in the prescribed SSTs (e.g. the SST analysis system may have damped spatial variations).
   d. The variability in midlatitudes may be under simulated on long time scales compared to that in full coupled models (Barsugli and Battisti 1998) because feedbacks from the atmosphere to the ocean are missing.
   e. Model insensitivity to the specified boundary conditions.
2. The weather noise induced T_{850} variance clusters closely around the "observed" value in both seasons. The fact that almost all points lie above the line shows that the models tend to simulate relatively more weather noise induced variance in JJA than in DJF.
3. There is considerably less scatter for T_{850} than for ϕ_{500}. This is understandable considering the strong constraint that is exercised by the prescribed SSTs and sea-ice boundaries on 850 hPa temperature. There is even more scatter in the corresponding figure for ψ_{300}. One might conjecture that, in general, the amount of scatter increases in the vertical.
4. DJF variance of both kinds is generally a good predictor of the corresponding JJA variance. This is not all that surprising because the variance statistics considered here are global averages and because, to first order, the spatial distribution of variance in JJA is antisymmetric about the equator to that in DJF.
5. A majority of the AMIP models under simulate both total and weather noise induced ϕ_{500} variability. As with T_{850}, weather noise induced ϕ_{500} variance tends to be relatively greater in JJA than in DJF.

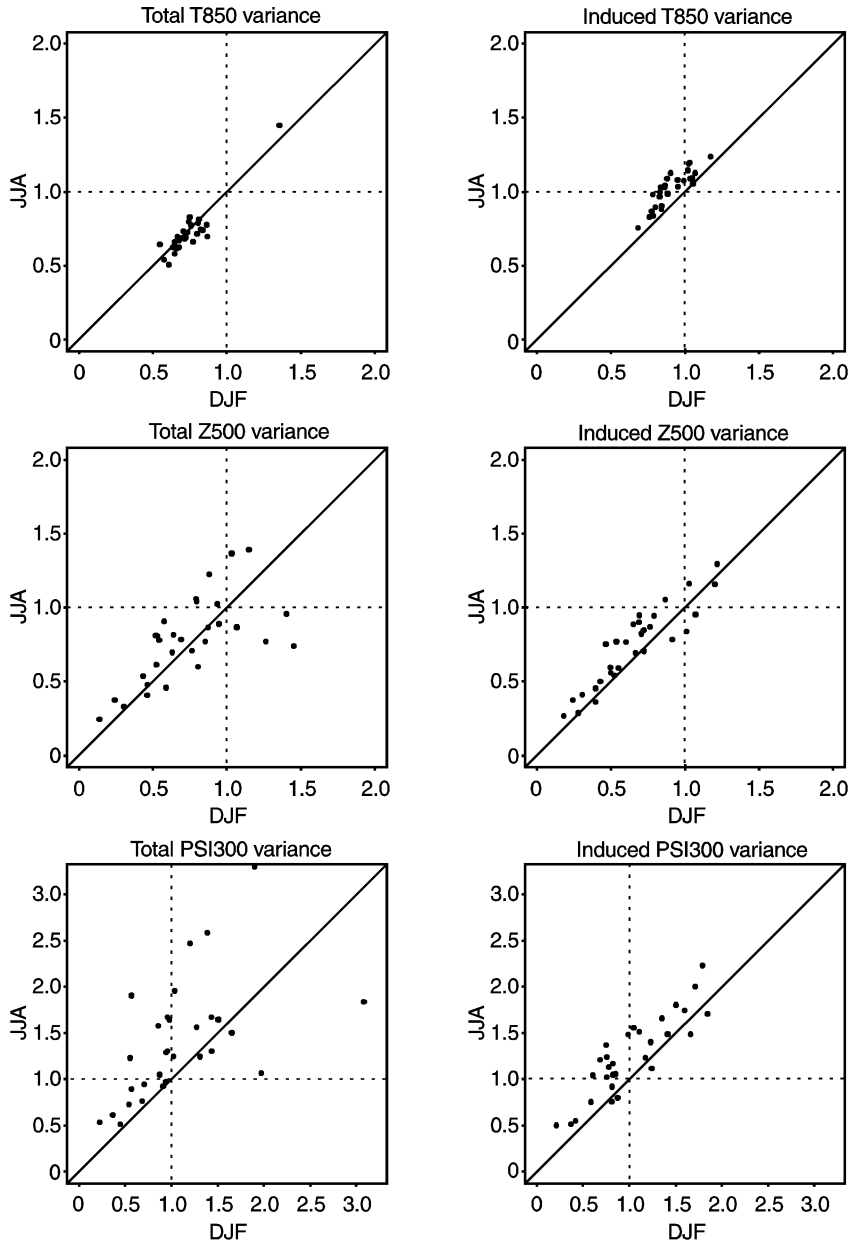

Fig. 1 The globally averaged total (*left hand column*) and induced (*right hand column*) inter-annual variance of DJF and JJA mean T_{850} (*upper row*), ϕ_{500} (*middle row*) and ψ_{300} (*lower row*) as simulated by the 30 AMIP models. The variances are expressed as a proportion of the corresponding observed quantity. There is one point outside of bounds at (2.05, 2.80) in the *upper right hand panel*

6. While the median model simulates DJF ψ_{300} total and weather noise induced variability correctly, a majority of models over simulate JJA ψ_{300} variability.

The variance indices displayed in Fig. 1 were cross-tabulated with a number of descriptors of model characteristics (e.g. horizontal and vertical resolution, whether grid point or spectral, type of radiation, cloud and convection treatment, etc.) as summarised in Phillips (1994). While this attempt to explain a gross feature of the simulated climate using a gross descriptor of the model was largely unsuccessful, we did uncover evidence suggesting that the amount of variability simulated by a model is related to its resolution. The globally averaged total variance of ϕ_{500} tends to increase with vertical and horizontal resolution in both seasons (Fig. 2) and similarly for ψ_{300} (not shown). In contrast, the globally averaged total variance of T_{850} does not appear to be related to resolution (not shown). Similarly, the weather noise induced simulated variance tends to increase with vertical and horizontal resolution in both seasons for ϕ_{500} and ψ_{300} while that of T_{850} (not shown) is virtually unrelated to resolution. Note that the resolution effect may be strongly confounded with those of other model features since the high resolution models also tend to be the more modern models in the AMIP cohort (Boer; personal communication). The apparent resolution effect may

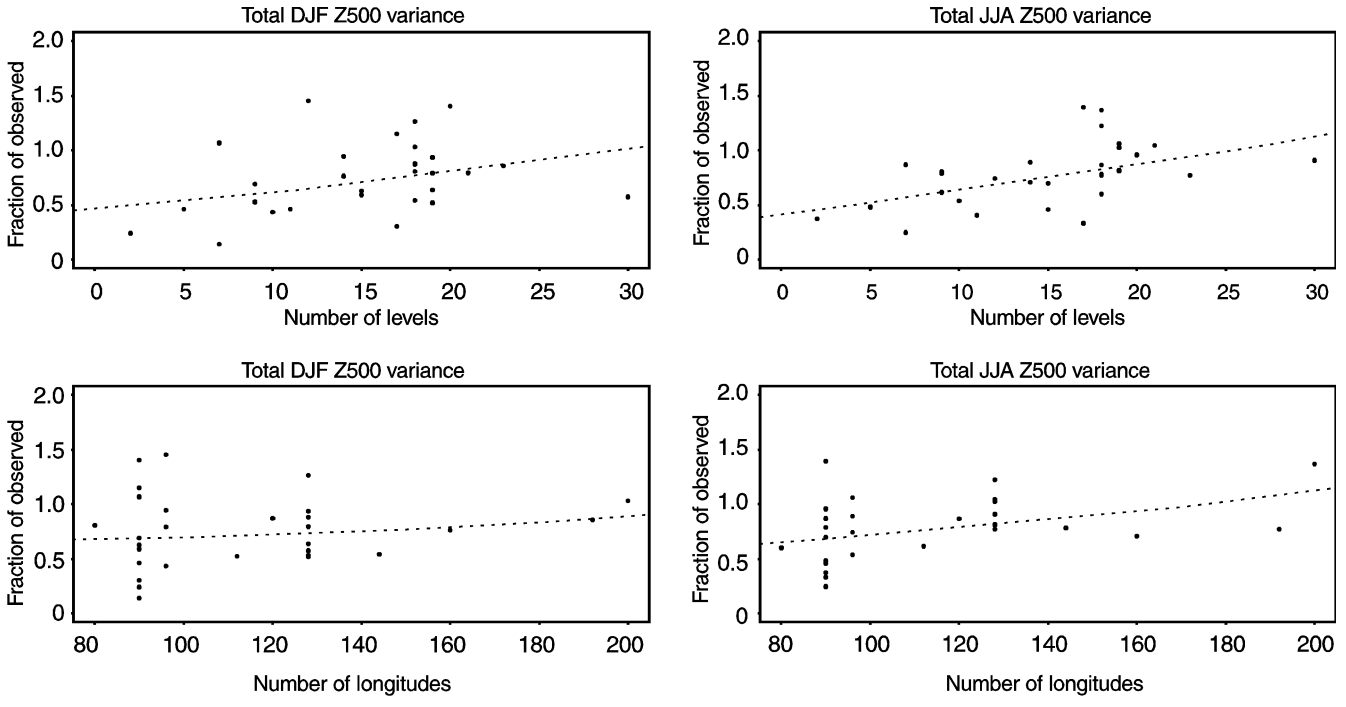

Fig. 2 Globally averaged total ϕ_{500} variance expressed as a proportion of the corresponding observed variance and plotted as a function of the vertical resolution (*upper row*) and horizontal *x*-resolution (*lower row*) of the simulating model. Dashed lines are linear least-squares fits

therefore also be due to other evolutionary changes which have been made to these models.

We also tried to classify the models according to their ability to simulate total and weather noise induced variability. For this purpose, the variability properties of each model were crudely characterised using a 12-number variance summary consisting of the global means of the total and weather noise induced interannual variance of T_{850}, ϕ_{500} and ψ_{300} in DJF and JJA. Hartigan's k-means clustering algorithm (Hartigan, 1979) was applied to the variance summaries to divide the models into two groups. We then examined the relationship between cluster membership and a number of model characteristics using cross tabulations of cluster membership with various model characteristics. We found weak evidence that the clustering procedure could discriminate between two groups of models, one of which is characterised by greater variance and higher resolution and the other by lower variance and lower resolution. However the clustering procedure did not generally succeed in classifying the models in terms of their physical parametrisations.

4 Potential predictability

The potential predictability F statistic, Eq. (6), is displayed in Fig. 3a for "observed" DJF ϕ_{500}. Null hypothesis (4) is rejected over 51% of the globe at the 10% significance level. The general structure of this map is very similar to that of the forced SST/sea-ice signal obtained from CCC GCM2 (see Z96, Fig. 14). That is, we see a wide band of large F values in the tropics which reflect variations in the direct heating of the tropical atmosphere by the tropical ocean. There is weak evidence of potential predictability over North America where F is generally greater than one, but not by a statistically significant amount. The picture is very similar in JJA (not shown) except that now evidence of possible predictability over North America is restricted to the south-western part of the continent. There is also some evidence of a potentially predictable signal over South Asia. Z96 discusses these diagrams further and compares these results with previous studies.

The finding that there is only weak evidence of potential predictability over North America in the observed climate is consistent with the result of Z96 that the SST/sea-ice signal in CCC GCM2 is relatively weak over North America. Such a signal will be difficult to detect in the observed climate using F because of the short period of record. Specifically, F has substantially fewer degrees of freedom than the F_B statistic used by Z96 to detect the SST/sea-ice signal in an ensemble of AMIP simulations. In the present setup, where there is a single 9-y record (10-y in JJA), F has just 8 and approximately 16 degrees of freedom. In contrast, the F_B statistic used in the 2-way ANOVA of Z96 has 8 and

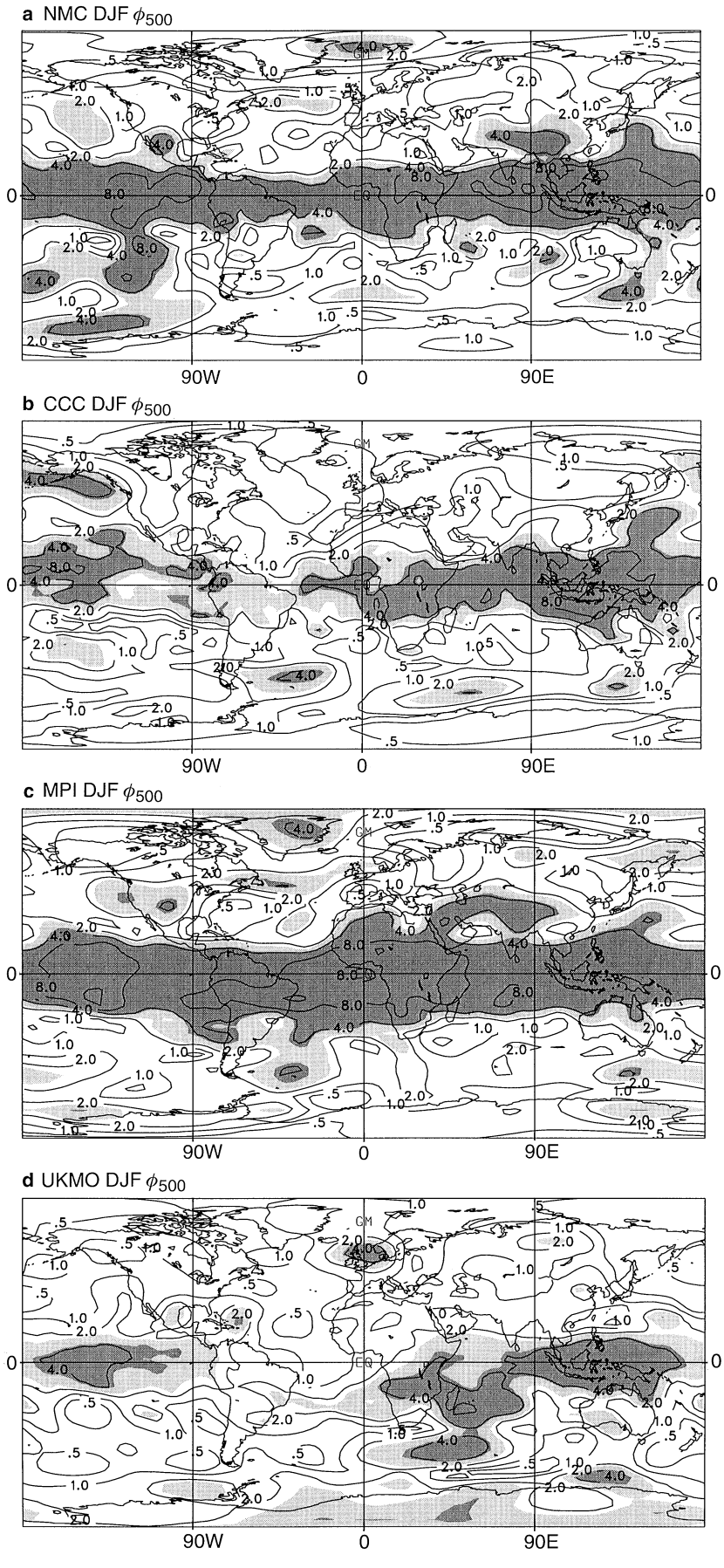

Fig. 3a–d The F ratio between the interannual variance of the seasonal mean and the corresponding weather noise induced variance for NMC analysed DJF ϕ_{500} **a** and **b** as simulated by the CCC model **c** MPI model and **d** UKMO model. Contours are 0.5, 1, 2, 4, 8 and 16. The F statistic is used to test the hypothesis that the climate is not potentially predictable. *Light* (*dark*) *shading* corresponds to F ratios that are significantly greater than one at the 5% (1%) significance level

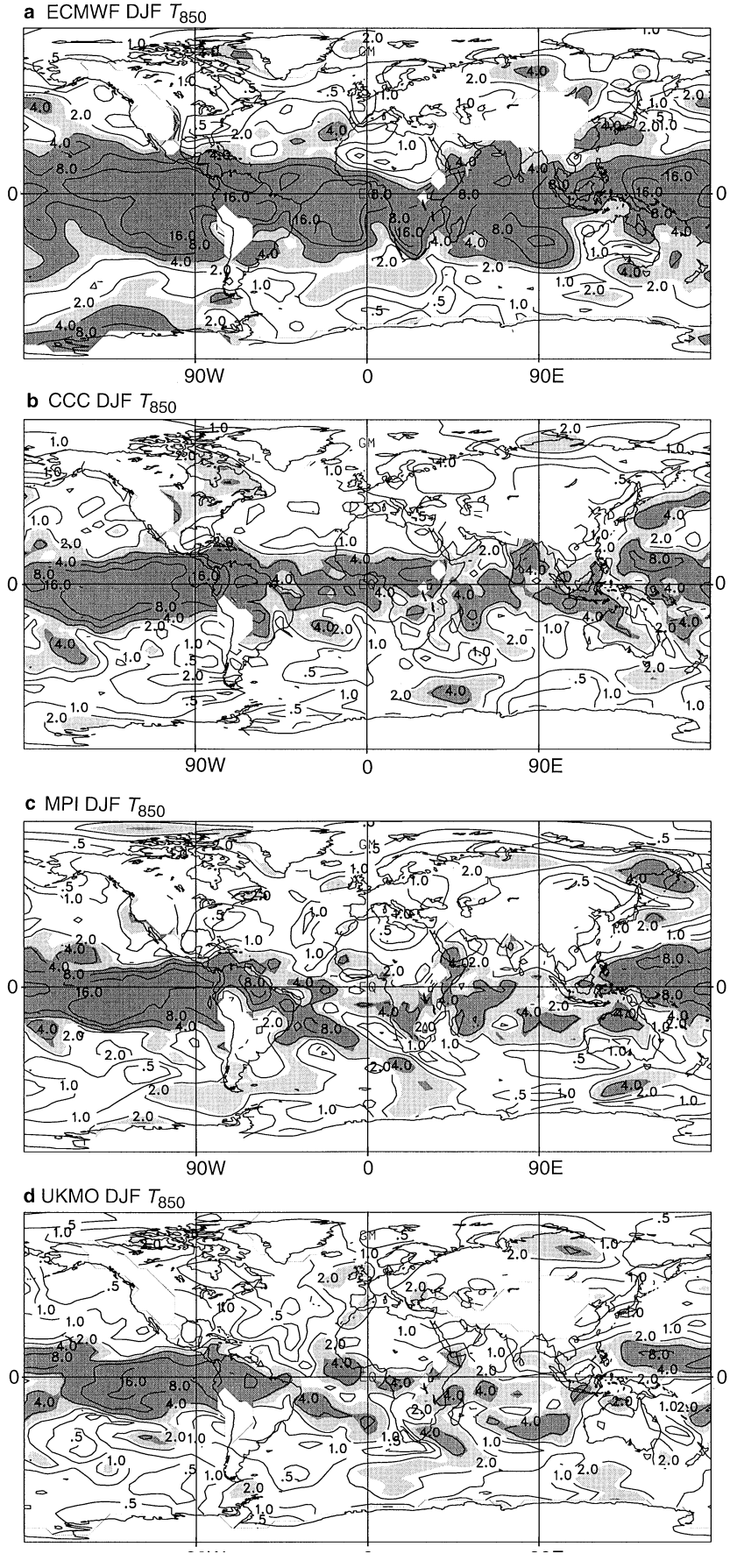

Fig. 4a–d The *F* ratio between the interannual variance of the seasonal mean and the corresponding weather noise induced variance for ECMWF analysed DJF mean **a** T_{850} and **b** as simulated by the CCC model, **c** MPI model and **d** UKMO model. Contours and *shading* are as in Fig. 3. High terrain areas have been *masked* (see text)

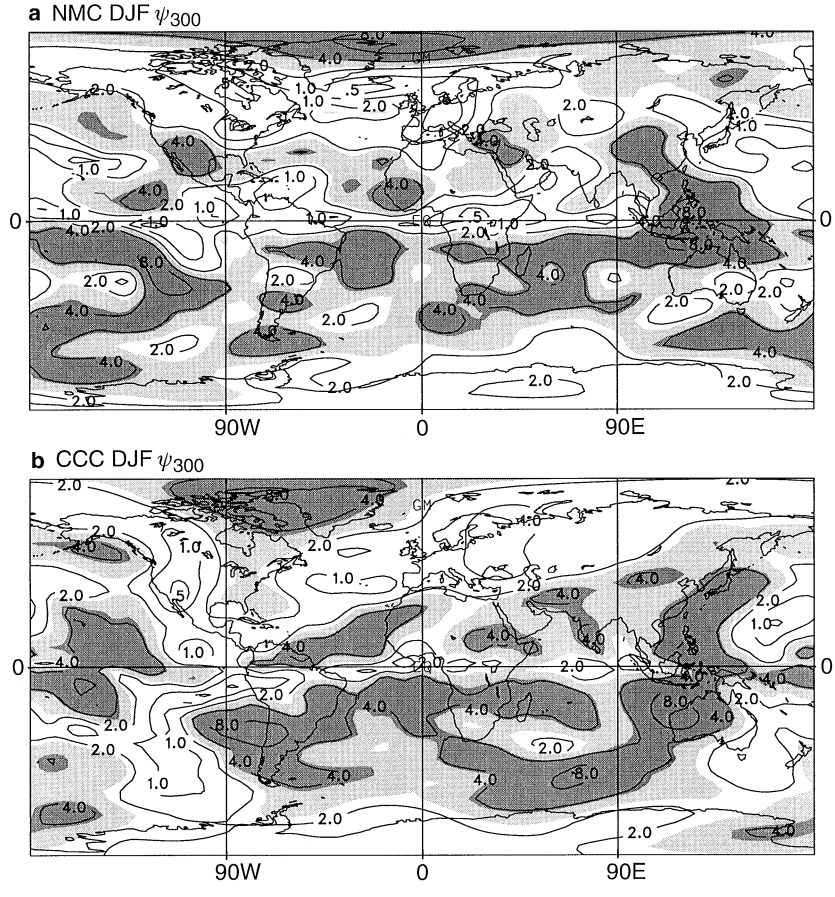

Fig. 5a, b The F ratio between the interannual variance of the seasonal mean and the corresponding weather noise induced variance for **a** NMC analysed DJF ψ_{300} and **b** as simulated by the CCC model. Contours and *shading* are as in Fig. 3

48 degrees of freedom. Power charts for the F-test displayed in Pearson and Hartley (1976; see pp 67 and 250–259) give some indication of the gain in power which occurs when the degrees of freedom in the denominator of the F statistic are increased from 16 to 48. The increase in power is particularly dramatic when the alternative to the null hypothesis is a weak signal.

Models generally simulate the observed potential predictability statistic for DJF ϕ_{500} quite well. The DJF results for the CCC, MPI and UKMO models displayed in Fig. 3b–d are typical and indicate some of the variation in results obtained between simulations. Differences between maps are due to both differences in model formulation and substantial sampling variability. An indication of the effects of sample variation can be obtained from the six member CCC GCM2 AMIP ensemble. In these simulations the rate of rejection of Eq. (4) varies between 41% and 56%. The MPI map is probably most typical of other models. In contrast to Fig. 3a–d, at least one model did not show significant DJF F statistics in the tropical band. Also, several models showed significantly large F statistics at high-southern latitudes.

Observations show moderate potential predictability in some areas in midlatitudes, in particular, in the North Pacific in DJF (Fig. 3a). This is believed to be largely due to the atmospheric response to the equatorial SST anomalies associated with ENSO. About half of the models were able to reproduce potential predictability to some extent in this area.

The models also simulated the observed potential predictability of JJA ϕ_{500} quite well. If anything, there was a slightly weaker tendency towards pathological behaviour in this season than in DJF. For example, the model cited already which simulated the DJF F statistics poorly in the tropical band, did quite well in JJA. Many models tend to have a wider area of potentially predictable variability in the tropics in the summer hemisphere than that in the winter hemisphere. Several models exhibited significantly large F statistics at high-northern latitudes.

The potential predictability statistic for observed DJF mean T_{850} is displayed in Fig. 4a. Note that high terrain has been masked out in this diagram. As with ϕ_{500}, strong evidence of potential predictability is seen in the tropics.

All models participating in AMIP simulated less T_{850} potential predictability than computed from the analyses. Typical examples for DJF, which correspond to those displayed above for ϕ_{500}, are illustrated in Fig. 4b–d. Although there is consistency between

models in terms of the proportion of the globe at which significant potential predictability is found, there is considerable variation in the spatial pattern of significant F statistics.

The ECMWF analysis shows high values of T_{850} potential predictability and standard deviation (not shown) in the equatorial eastern Pacific south of the equator. Many models do not reproduce or strongly under estimate this feature.

We have also computed the potential predictability of observed and simulated ψ_{300}. F statistics for DJF ψ_{300} computed from observations are displayed in Fig. 5a. The corresponding diagram for the CCC AMIP simulation, which is typical of many other models, is displayed in Fig. 5b. We see substantially more evidence in both diagrams for potential predictability of the extra-tropical circulation at the 300 hPa level than we do in the corresponding diagrams (Fig. 3a, b) for geopotential at the 500 hPa level. There is also evidence of potential predictability in the rotational flow on the 300 hPa surface in the tropics which appears to be the result of the atmosphere's response to variations in tropical SSTs.

There is a great deal of intermodel variation in both the total interannual variance of ψ_{300} that is simulated and also in the diagnosed potential predictability. This is perhaps not surprising considering that the 300 hPa surface is near the tropopause where we might expect to see strong effects from variation in the large-scale overturning circulations (such as the Walker and

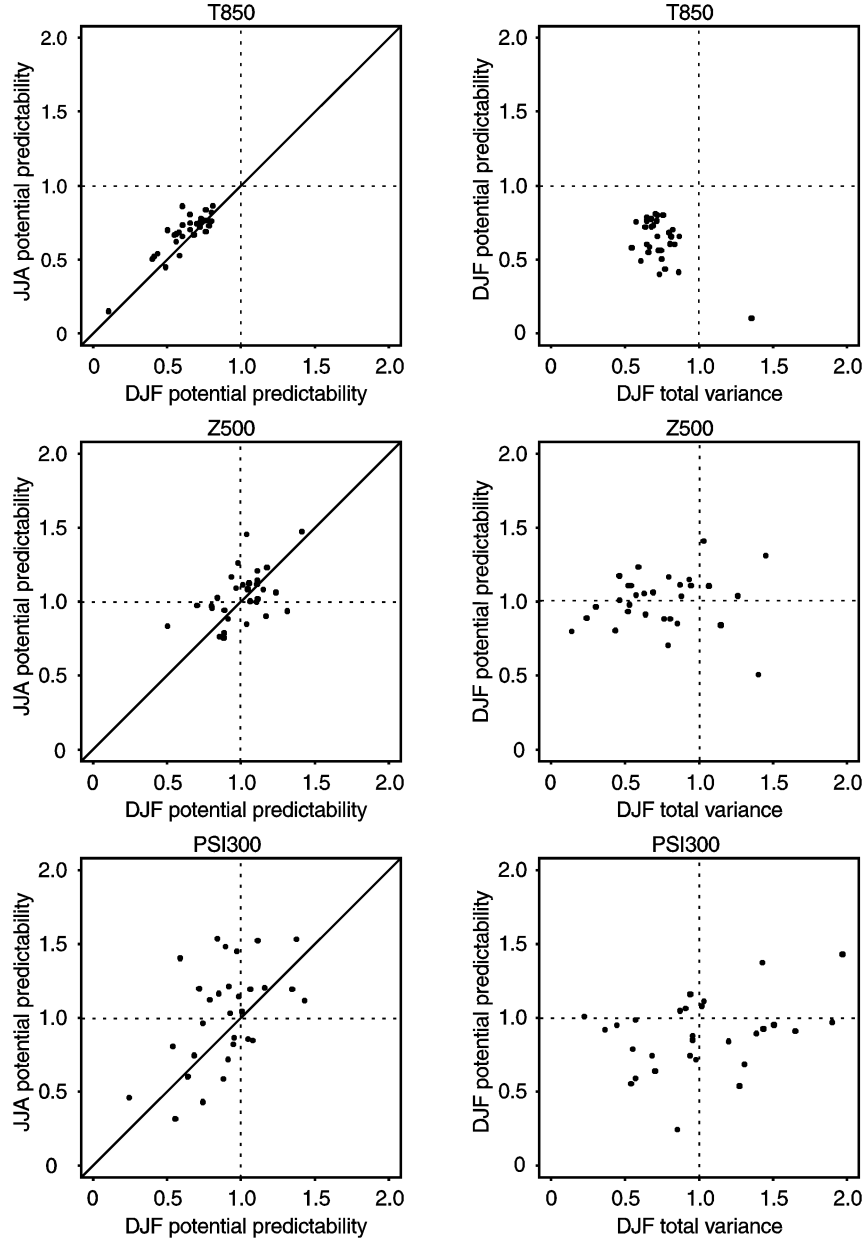

Fig. 6 *Left column*, the proportion of the globe over which the simulated climate is potentially predictable expressed as a fraction of the proportion over which the "observed" climate is potentially predictable. The JJA potential predictability ratio is displayed as a function of the DJF ratio. *Right column*, the DJF potential predictability ratio (see above) displayed as a function of the DJF global mean total variance (expressed as a fraction of that which is "observed"). The *bottom right hand panel* has one dot out of bounds at (3.09, 1.59)

Hadley cells) of the atmosphere. There are apparently large differences between models in the sensitivity of these circulations to variations in the conditions occurring at the atmosphere's lower boundary.

Figure 6 summarises the ability of the models to simulate the observed potential predictability. The left hand panels display the proportion of the globe with significant F statistics expressed as a fraction of the proportion of the globe over which the observed climate has significant F statistics. The right hand panels display this quantity for DJF as a function of globally averaged total interannual variance simulated by the model in DJF.

Several things can be noted about Fig. 6.

1. Intermodel variation is least for T_{850}, presumably because T_{850} is more strongly constrained by the prescribed SSTs and sea-ice boundaries. Also, there is more scatter for ψ_{300} than for ϕ_{500}.
2. DJF potential predictability is generally a good predictor of that in JJA.
3. As noted, the models simulate less potential predictability for T_{850} than computed from ECMWF analyses.
4. The "average" model simulates the observed ϕ_{500} potential predictability well. This is indicated by the fact that the points in the centre left panel of Fig. 6 cluster about the (1, 1) location. The same is true of ψ_{300}.
5. The right hand column in Fig. 6 shows that the ability to simulate the climate's potential predictability is not related to the ability to simulate the correct amount of interannual variability. That is, the ratio between potentially predictable and weather noise induced variance in a simulated climate is not determined by its low frequency variance.

5 Conclusions

We have examined the interannual variability and potential predictability of all of the climates simulated by the models participating in AMIP in a number of ways and now summarise by drawing a few general conclusions.

1. There is wide variation in the ability to simulate observed inter-annual variability, both total and induced. The range of abilities apparently increases in the vertical as the degree of contact between the simulated atmosphere and the prescribed SST/sea-ice conditions decreases.
2. A majority of models under simulate seasonal mean ϕ_{500} variability in both DJF and JJA. Also, a majority of models over simulates ψ_{300} variability in JJA. We could not judge the ability of models to simulate T_{850} variability because of data problems. We conjecture, however, that models simulate T_{850} variability quite well (see Z96 for related discussion).
3. With the possible exceptions of resolution and model type (grid point or spectral), there is little apparent connection between gross model characteristics and the corresponding ability to simulate observed variability. Models with higher resolution tend to simulate more total interannual and weather-noise induced variance in both seasons.
4. Most models simulate the ratio between total and weather noise induced inter-annual variance quite well for both ϕ_{500} and ψ_{300}. That is, models exhibit potential predictability similar to that which is observed even if they do not simulate observed variability very well.

Acknowledgements We thank Dave Ramsden for help with obtaining and processing the data used in this study. Also we thank David Straus, Akio Kitoh, Martin Dix and Ali Harzallah for comments on an interim version of this manuscript. Work by V. Kharin was supported by the Canadian Institute for Climate Studies via Collaborative Research Agreement (Collaborative Research Agreement Number 7 CICS-VARIABILITY Number 4-44300) with the University of Victoria. We also thank all AMIP participants for providing the data for this study.

References

Anderson JL, Stern WF (1996) Evaluating the potential predictive utility of ensemble forecasts J Clim 9:260–269

Barsugli JJ, Battisti DS (1998) The basic effect of atmosphere-ocean thermal coupling on midlatitude variability. J Atmos Sci 55: 477–493

Boer GJ (1985) Modelling the atmospheric response to the 1982/3 El Nino. In: Nihoul JCJ (ed) Coupled ocean-atmosphere models. Elsevier, Amsterdam, pp 1–17

Dix MR, Hunt BG (1995) Chaotic influences and the problem of deterministic seasonal predictions. Int J Climatol 15:729–752

Gates WL (1992) AMIP: The Atmospheric Model Intercomparison Project. Bull Am Meteorol Soc 73:1962–1970

Gill AE (1980) Some simple solutions for heat-induced tropical circulation. Q J R Meteorol Soc 106:447–462

Hartigan JA (1979) Clustering algorithms. Wiley

Kumar A, Hoerling MP (1995) Prospects and limitations of seasonal atmospheric GCM predictions. Bull Am Meteorol Soc 76:335–345

Kumar A, Hoerling MJ, Leetmaa A, Sardeshmukh P (1996) Assessing a GCM's suitability for making seasonal predictions. J Clim 9:115–129

Madden RA (1976) Estimates of the natural variability of time averaged sea level pressure. Mon Weather Rev 104:942–952

McFarlane NA, Boer GJ, Blanchet J-P, Lazare M (1992) The Canadian Climate Centre second generation general circulation model and its equilibrium climate. J Clim 5:1013–1044

Pearson ES, Hartley HO (1976) Biometrika tables for statisticians, vol 2. Biometrika Trust, University College, London, England, 385 pp

Phillips TJ (1994) A Summary Documentation of the AMIP Models. Technical Report PCMDI Rep 18, PCMDI, Lawrence Livermore National Laboratory, Livermore, CA 94550

Robock A, Schlosser CA, Vinnikov KY, Liu S (1996) Validation of humidity, moisture fluxes, and soil moisture in GCMs: report of AMIP diagnostic subproject 11 part-1 soil moisture. In: Proc First Int AMIP Sci Conf (15–19 May 1995, Monterey, CA). World Climate Research Programme, Geneva

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

Sheng J, Zwiers FW (1998) A time-interpolation scheme for bottom boundary conditions in atmospheric general circulation models. Clim Dyn (in press)

Stern W, Miyakoda K (1995) Feasibility of seasonal forecasts inferred from multiple GCM simulations. J Clim 8: 1071–1085

Trenberth KE (1992) Global analyses from ECMWF and atlas of 1000 to 10 mb circulation statistics. NCAR/TN-373 + STR, NCAR, Boulder, CO, 191 pp + 24 fiche

Webster PJ (1981) Mechanisms determining the atmospheric response to sea surface temperature anomalies. J Atmos Sci 38: 554–571

Zwiers FW (1987) A potential predictability study conducted with an atmospheric general circulation model. Mon Weather Rev 115: 2957–2974

Zwiers FW (1996) Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. Clim Dyn 12: 825–847