



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities

### Citation for published version:

Collins, M, Osborn, TJ, Tett, SFB, Briffa, KR & Schweingruber, FH 2002, 'A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities', *Journal of Climate*, vol. 15, no. 13, pp. 1497-1515. [https://doi.org/10.1175/1520-0442\(2002\)015<1497:ACOTVO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1497:ACOTVO>2.0.CO;2)

### Digital Object Identifier (DOI):

[10.1175/1520-0442\(2002\)015<1497:ACOTVO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1497:ACOTVO>2.0.CO;2)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Journal of Climate

### Publisher Rights Statement:

Published in the Journal of Climate by the American Meteorological Society (2002)

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## A Comparison of the Variability of a Climate Model with Paleotemperature Estimates from a Network of Tree-Ring Densities

MATTHEW COLLINS

*Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, Berkshire, United Kingdom, and Centre for Global Atmospheric Research, Department of Meteorology, University of Reading, Reading, Berkshire, United Kingdom*

TIMOTHY J. OSBORN

*Climate Research Unit, University of East Anglia, Norwich, Norfolk, United Kingdom*

SIMON F. B. TETT

*Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, Berkshire, United Kingdom*

KEITH R. BRIFFA

*Climate Research Unit, University of East Anglia, Norwich, Norfolk, United Kingdom*

FRITZ H. SCHWEINGRUBER

*Swiss Federal Institute for Forest and Snow Research, Birmensdorf, Switzerland*

(Manuscript received 13 November 2001, in final form 28 October 2001)

### ABSTRACT

Validation of the decadal to centennial timescale variability of coupled climate models is limited by the scarcity of long observational records. Proxy indicators of climate, such as tree rings, ice cores, etc., can be utilized for this purpose. This study presents a quantitative comparison of the variability of the third version of the Hadley Centre ocean–atmosphere coupled model with a network of temperature-sensitive tree-ring densities covering the northern high latitudes. The tree-ring density records are up to 600 years long, and temperature reconstructions based on two different methods of removing the bias due to changing tree age are used. The first is a standard method that may remove low-frequency variability on timescales of the order of the tree life span (i.e., multi-decadal to century timescales). The second (age-band decomposition) maintains low-frequency variability by only comparing similar age tree rings at each site, thus avoiding the need to remove the age effect (but at the cost of greater uncertainty in the earlier years when fewer tree cores are available). The variability of the model control simulation, which represents only the internal variability of the climate system, agrees reasonably well with the tree-ring reconstructions using the standard method at the regional level, although the model may underestimate the variance of mean Northern Hemisphere land temperature by as much as a factor of 1.8 on all timescales if one takes account of the uncertainty in the tree-ring reconstructions. Agreement with the age-band decomposition tree-ring reconstructions is less good with the model underestimating the hemispheric variance by as much as a factor of 2.1 on all timescales and by as much as a factor of 3.0 on decadal to centennial timescales. Underestimation of the natural variability of climate by the model would be serious as it may lead to false detections of climate change or erroneously low uncertainty estimates in future climate predictions. However, it is shown that some of this underestimation may be due to the lack of natural climate forcing in the model control simulation due, for example, to solar variability and volcanic eruptions. The study suggests that further quantification of the uncertainties in the proxy data, and inclusion of natural climate forcings in the model simulations, are important steps in making comparisons of climate models with the proxy record over the last 1000 years.

### 1. Introduction

Multicentury integrations of coupled climate models are now routinely produced at climate modeling centers. Such simulations are performed for various purpose,

examples of which include testing the stability of the coupled model (e.g., Gordon et al. 2000); examining low-frequency variability of the climate system (e.g., Collins et al. 2001); and defining “natural” climate variability as a basis for statistical testing in studies of the detection, attribution, and prediction of climate change (e.g., Tett et al. 1999). Thus it is important that the climate of the model is validated against observations of the real climate system. This involves not only val-

---

*Corresponding author address:* Matthew Collins, Centre for Global Atmospheric Research, Dept. of Meteorology, University of Reading, Reading, Berkshire RG6 6BB, United Kingdom.  
E-mail: matcollins@met.rdg.ac.uk

idating the mean climate of the model, but also validating the climate variability over a range of time- and space scales.

In order to validate the variability of a model on timescales of centuries one requires multicentury records of observed climate variables. These are, however, sparse. The longest instrumentally recorded series is the Central England Temperature record (Parker et al. 1992; Manley 1974), reliable back to 1659, but only representative of temperature over a relatively small geographic region. This confounds the comparison with direct climate model output that is representative of, at best, temperatures at the gridbox size (e.g., 300 km) and may be more representative of temperatures on the scales of many grid boxes [due to model formulation that restricts the accumulation of energy at the grid scale, for example, Stott and Tett (1998)]. Also, we are often interested in global- or hemispheric-scale variability, because of its relevance to the detection, attribution, and prediction of climate change, which single-site records cannot provide. Long records of climate can however be extracted from proxy indicators such as tree rings, ice cores, corals, borehole temperatures, etc. (e.g., Briffa et al. 1998a; Jones et al. 1998; Mann et al. 1998; Huang et al. 2000; Harris and Chapman 2001, and many others) that are widely spread over the globe. While these proxy records are often not direct measurements of, for example, temperature or precipitation, they do contain a wealth of information about past climate variations.

Previous efforts to compare model and proxy variability have been somewhat qualitative in their approach. Some examples include Jones et al. (1998), who compared the temperature variability of two coupled models with 17 normalized proxy series representing a variety of historical and instrumental records, tree-ring density and width, ice cores, and corals. They found that one of the models showed similar behavior to the proxy data in terms of their principal spatial patterns of variability, while the other model showed markedly different behavior. Crowley and Kim (1999) compared the spectra of global temperature variability of two coupled models with the multiproxy reconstructions of Mann et al. (1998) and found them to agree at the 90% level once they had removed an estimate of the naturally forced variability from the Mann et al. (1998) data using an energy balance model. Delworth and Mann (2000) also compared the Mann et al. (1998) reconstructions with model-generated interdecadal variability in the North Atlantic region and found similar patterns of SST variability.

There are considerable challenges that the climate community must overcome to successfully utilize proxy records in the validation of climate models. This includes developing methods to make *quantitative* comparisons of models and proxy data. We may distinguish two distinct approaches to this problem: (i) an “inverse” method, by which the proxy data are calibrated to rep-

resent large-scale climate variables that the model can simulate, such as temperature and precipitation, prior to the comparison (e.g., the multiproxy reconstructions of Mann et al. 1998), with care taken to quantify an residual variance not accounted for in the calibration procedure; (ii) a “forward” method, proposed by Reichert et al. (1999), by which the model variables are processed to represent the proxy data. This may involve scaling down the grid-scale variables of the model to a more representative local scale and then using physically or statistically based methods to generate the proxy in question, for example, glacier length (Oerlemans and Reichert 2000) or oxygen isotope ratios.

In the study described here, a multicentury integration of the third version of the Hadley Centre ocean-atmosphere coupled model (HadCM3; Gordon et al. 2000) is compared with a number of regional temperature time series, estimated from a network of tree-ring density chronologies spread over the Northern Hemisphere (Briffa et al. 2001b), primarily representative of summer (April–September) conditions and extending back 600 years. The climate system is unlikely to be greatly affected by orbital variations over this period so we use the control simulation of the model, which has orbital parameters that are fixed to present-day values. We adopt the “inverse” method of comparing the tree-ring data and the model. The tree-ring data have been previously averaged into nine regions that are more representative of the space scales that can be realistically simulated by the model (Stott and Tett 1998) and are calibrated using observed records of temperature from recent times (Briffa et al. 2001b). Correspondingly, we subsample the model in the nine tree-ring regions during the optimal (April–September) growth season of the trees (Briffa et al. 2001b), in order to make a like-with-like comparison.

In section 2 of the paper the tree-ring network and its calibration are introduced and in section 3 HadCM3 is described briefly. In section 4 the model and tree-ring reconstructions are compared visually. In section 5 their variances are compared and in section 6 their power spectra are compared. In section 7 the dominant spatial patterns of variability of the model and tree rings are contrasted and in section 8 the possible contribution to the model variability by solar and volcanic forcing is assessed. Finally, conclusions are drawn in section 9.

## 2. Tree-ring density chronologies

The climate reconstructions used here are based on a network of 387 tree-ring density chronologies located over much of the Northern Hemisphere extratropics (Fig. 1). The chronologies range in length from 100 to more than 600 years, with each consisting of, on average, data from 25 tree cores from a site close to the present timberline (i.e., at high elevation or high latitude) to maximize the potential temperature signal. Trees that grow near the timberline tend to be stressed

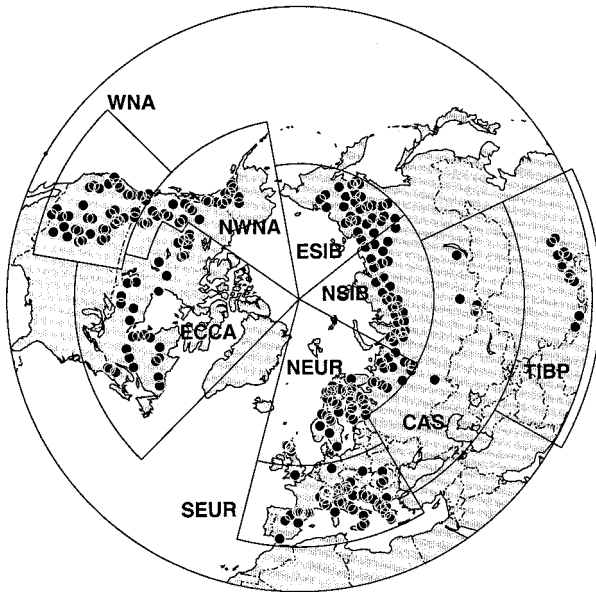


FIG. 1. Locations of individual tree-ring density chronologies (black dots; overlap areas outlined in white) and the definition of the nine regional series used in the calibration against observed temperature (after Briffa et al. 2001a).

and are hence more sensitive to variations in climate. Briffa et al. (2001b) show that, for this network, a much more reliable temperature reconstruction can be obtained from the density of the wood formed toward the end of each growing season (the maximum latewood density), than from other measures such as the width of each tree ring. The dominant climate signal in the dataset as a whole is the growing season temperature (Briffa et al. 2001b). The timing of the growing season varies with location, but the mean temperature from April to September provides the best overall correlations with tree-ring density.

Briffa et al. (2001b) combine the chronologies into nine regional averages (also shown in Fig. 1), and calibrate them against April–September temperatures for each of the regions, using simple linear regression over the period 1881–1960. The regions were defined partly arbitrarily and partly on the basis of tree genera, climate regime, and treeline type (elevational or latitudinal). There are a number of requirements that constrain the most appropriate size of the regions: (i) they need to be subhemispheric to capture spatial variability; (ii) if they are too small (e.g., individual grid boxes) then the reliability of the reconstructions is less and it is already known that the variance of the climate model is underestimated at this spatial scale (Stott and Tett 1998); and (iii) the model–data comparison should inform us about the reliability at spatial scales appropriate to the climate signal detection and attribution problem. The defined regions satisfy these requirements in general.

The correlations between the regional tree-ring density records and the temperatures are reproduced in Ta-

TABLE 1. Parameters of the calibration process of the standard tree-ring density regional chronologies in which the tree-ring series are regressed onto the observed temperature from 1881 to 1960 (see text for the description of the algorithm). Here  $R$  is the correlation coefficient,  $R^2$  is the explained variance,  $\beta$  is the regression coefficient, and rmse is the root-mean-squared error (after Briffa et al. (2001b)).

	$R$	$R^2$ (%)	$\beta$	Rmse
NEUR	0.84	71	0.58	0.37
SEUR	0.72	52	0.35	0.33
NSIB	0.76	58	0.70	0.59
ESIB	0.61	37	0.46	0.60
CAS	0.56	31	0.25	0.36
TIBP	0.36	13	0.08	0.21
WNA	0.74	55	0.38	0.34
NWNA	0.60	36	0.37	0.48
ECCA	0.75	56	0.51	0.45
NH	0.71	50	0.18	0.18

ble 1, together with the standard deviations of the regression residuals [i.e., the root-mean-squared error (rmse)]. Over the calibration period, the variance of each reconstruction is less than the variance of the observed temperature time series, with the difference in variance equal to the square of the rmse. This residual variance must be considered when comparing the variance of the reconstructions with the climate model simulations, as it is the component of temperature variability not recorded by the trees. When comparing the power spectra or the variance of decadal means, some assumption must also be made about the spectrum of the residuals. For simplicity we assume here that they are white noise. In fact, the residuals exhibit significant autocorrelation in 4 out of the 9 regions during 1881–1960, but it is unclear how representative this is of the pre-1881 period due to the presence of (warming) trends during 1881–1960. This is the subject of further study, but here we retain the white noise assumption.

Briffa et al. (2001b) also formed a weighted average of the nine regional tree-ring density records, prior to their calibration. The weighting was time dependent, and was based on a statistical measure of the strength of the climate signal in each region. This weighted average was calibrated against the observed April–September temperature averaged over all land areas north of 20°N, and is denoted by “NH.” The calibration parameters for the NH series are also given in Table 1. Briffa et al. (2001b) show that for the NH reconstruction, and for 7 out of the 9 subregions, the correlation with temperature is higher at the decadal timescale than that given in Table 1.

The calibrated temperature reconstructions for the nine subregions and the NH region are shown in Fig. 2. These are the standard tree-ring series used in the present study. Most regions show a warming during the first half of the twentieth century, followed by a cooling. Summer temperatures in these regions did cool from 1940 to 1970, but this has been followed by a rapid warming that the tree-ring density has not responded to. Briffa et al. (1998b) identified this widespread decline

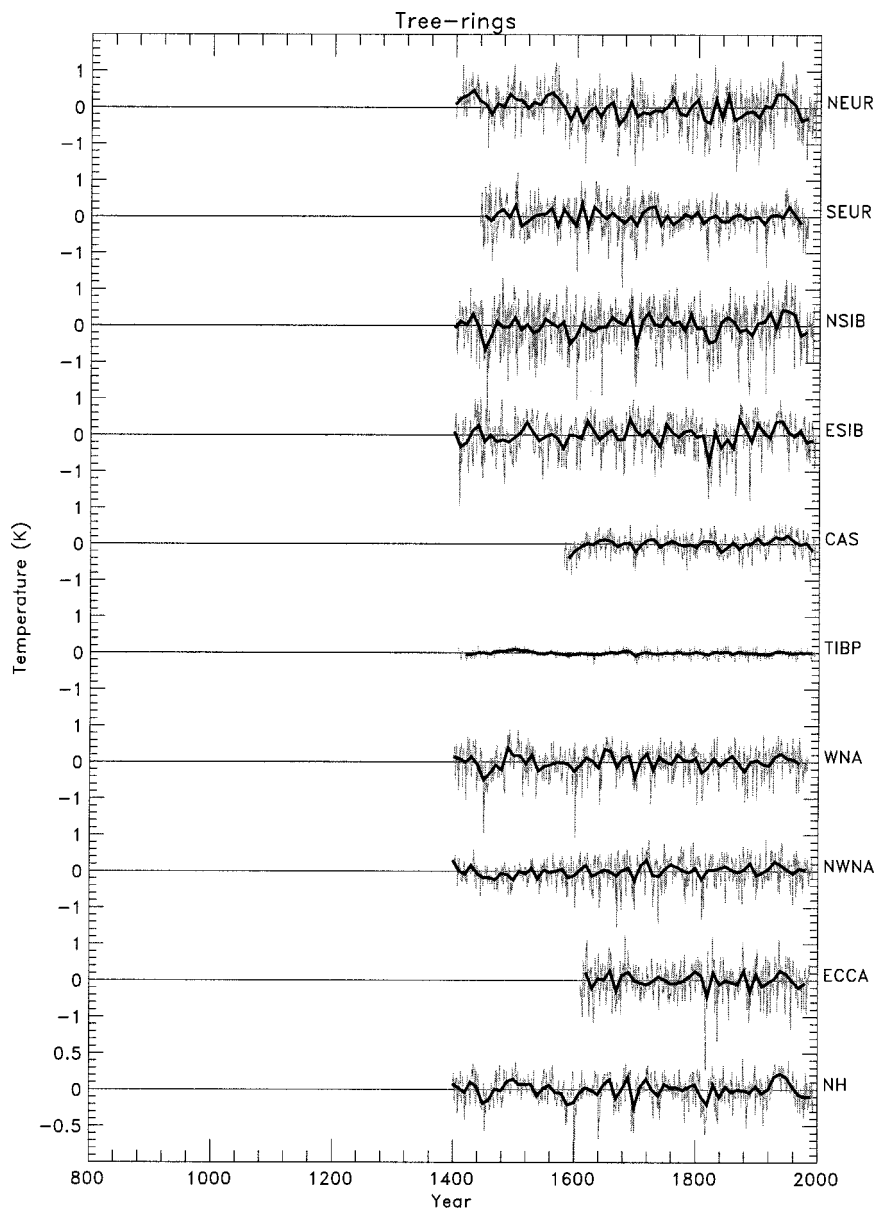


FIG. 2. The calibrated standard tree-ring series for the nine regions and for NH (see Fig. 1 for the definitions of the regions). The gray lines are the annually resolved data and the black lines are decadal averages formed from these. Note the change in vertical scale for the NH series.

in the density data (and, to a lesser extent, in the tree-ring width data not used here), and speculated that some anthropogenic cause was likely to have superimposed this non-temperature-related trend onto the tree-ring density records. In this study we make the assumption that this unknown factor is of anthropogenic origin and did not occur in the past. Hence we attribute all earlier variations to changes in growing-season temperatures.

In addition to responding to growing-season temperature, the maximum latewood density of each tree ring also depends upon the age of the ring (generally showing a downward trend with increasing tree age). Since the

age-density function is unknown (and is considered to vary with location and tree species), a generalized exponential function (a “Hugershoff” function, Bräker 1981) was fitted to each tree core and removed. This detrending technique is known as “standardization,” and results in a loss of multicentury variance, the extent of which is dependent on tree longevity (Cook et al. 1995; Briffa et al. 1996). The standard reconstructions (Fig. 2) may, therefore, be lacking in multicentury variability.

To broaden our comparisons we also use the reconstructions of Briffa et al. (2001a). These are based on

TABLE 2. Parameters of the calibration process of the ABD tree-ring density chronologies (see text for the description of the algorithm). Here  $R$  is the correlation coefficient,  $R^2$  is the explained variance,  $\beta$  is the regression coefficient, and rmse is the root-mean-squared error. There is no value of  $\beta$  for the NH reconstruction because multiple linear regression is used (after Briffa et al. (2001a)).

	$R$	$R^2$ (%)	$\beta$	Rmse
NEUR	0.76	58	0.58	0.44
SEUR	0.65	42	0.39	0.37
NSIB	0.77	59	0.63	0.58
ESIB	0.52	27	0.43	0.64
CAS	0.53	28	0.27	0.37
TIBP	0.43	18	0.12	0.20
WNA	0.72	52	0.39	0.37
NWNA	0.65	42	0.40	0.46
ECCA	0.69	48	0.49	0.49
NH	0.68	46	—	0.19

the same tree-ring density dataset, and are for the same regions as the standard reconstructions. The tree cores are separated in subgroups defined by age class (e.g., 101–150 yr from the pith), assembled into individual age-band time series, scaled to have equal mean and variance and recombined for different age classes into a single series of relative growth changes. The age dependence is thus accounted for by only combining in absolute units the density from tree rings whose age falls in a restricted range (or band); combining different sites, species, and age bands (to form the regional chronologies) is only allowed after normalization has been applied to remove differences in absolute means and variances. The difference is that the tree cores are not standardized (i.e., detrended) and, therefore, lose no low-frequency variability. Full details of the procedure are given in Briffa et al. (2001a).

These age-band decomposition (ABD) regional chronologies are calibrated in the same way as the standard reconstructions. The calibration statistics (reproduced from Briffa et al. 2001a) are given in Table 2, while the time series are shown in Fig. 3. There is no artificial loss of multicentury variability, but this is at the cost of greater uncertainty in the earlier part of the record for which there are fewer tree cores (because the ABD technique requires more tree cores). As described in Briffa et al. (2001a), one measure of the reliability is to evaluate the strength of the variability that is common to all age bands. This measure leads us to truncate some of the regional ABD series, compared to the standard series (cf. Fig. 3 and Fig. 2).

The ABD approach to tree-ring chronology construction is a useful addition to the range of processing methods employed to remove sample age-related bias in radial tree growth measurements. It has strengths and weaknesses in comparison to other such methods and should not be considered to be wholly superior on inferior to any of them. For example, other methods can better reconstruct interannual variability (Briffa et al. 2001a). We use the ABD method for this study because of its potential to express more low-frequency vari-

ability in the final chronology than other approaches. It is the low-frequency variability that is most important in studies of the detection and attribution of climate change.

These two alternative, but related, sets of reconstructions of regional and near-hemispheric growing season temperatures are used for comparison with the simulated variability. These sets are denoted by “standard” and by ABD.

### 3. HadCM3

The atmospheric component of HadCM3 is a version of the Met Office’s Unified Model (Cullen 1993). The model dynamics and physics are solved on a  $2.5^\circ \times 3.75^\circ$  latitude–longitude 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  $1.25^\circ \times 1.25^\circ$  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 that 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).

We analyze the HadCM3 control simulation in which all concentrations of greenhouse gases and aerosols, etc., are set as constants representative of the preindustrial era. Surface air temperatures (at a height of 1.5 m) were extracted from the HadCM3 control experiment in a way consistent with the calibrated tree-ring densities. Monthly mean temperatures were averaged at model land grid points for April–September. The Northern Hemisphere of the model was then subsampled according to the regions defined in Fig. 1 to form 9 regional time series of model temperatures of length 1200 yr (the first 100 yr of the experiment were ignored because of spinup effects). The model NH series was formed by averaging all the April–September land temperatures north of  $20^\circ\text{N}$  taking into account the relative areas of each grid box. This is in contrast to the reconstructed tree-ring NH series that is formed as a weighted mean of the nine regional tree-ring reconstructions (although this mean series was then calibrated against the observed mean temperature from all land north of  $20^\circ\text{N}$ ). A simple mean of the nine regional series from the model explains 72% of the variance of the true NH series at all timescales and 90% of the variance on decadal timescales. Forming a weighted mean of the nine model regional series, using linear regression to determine the weights, increases these fractions to 88% and 96%, respectively.

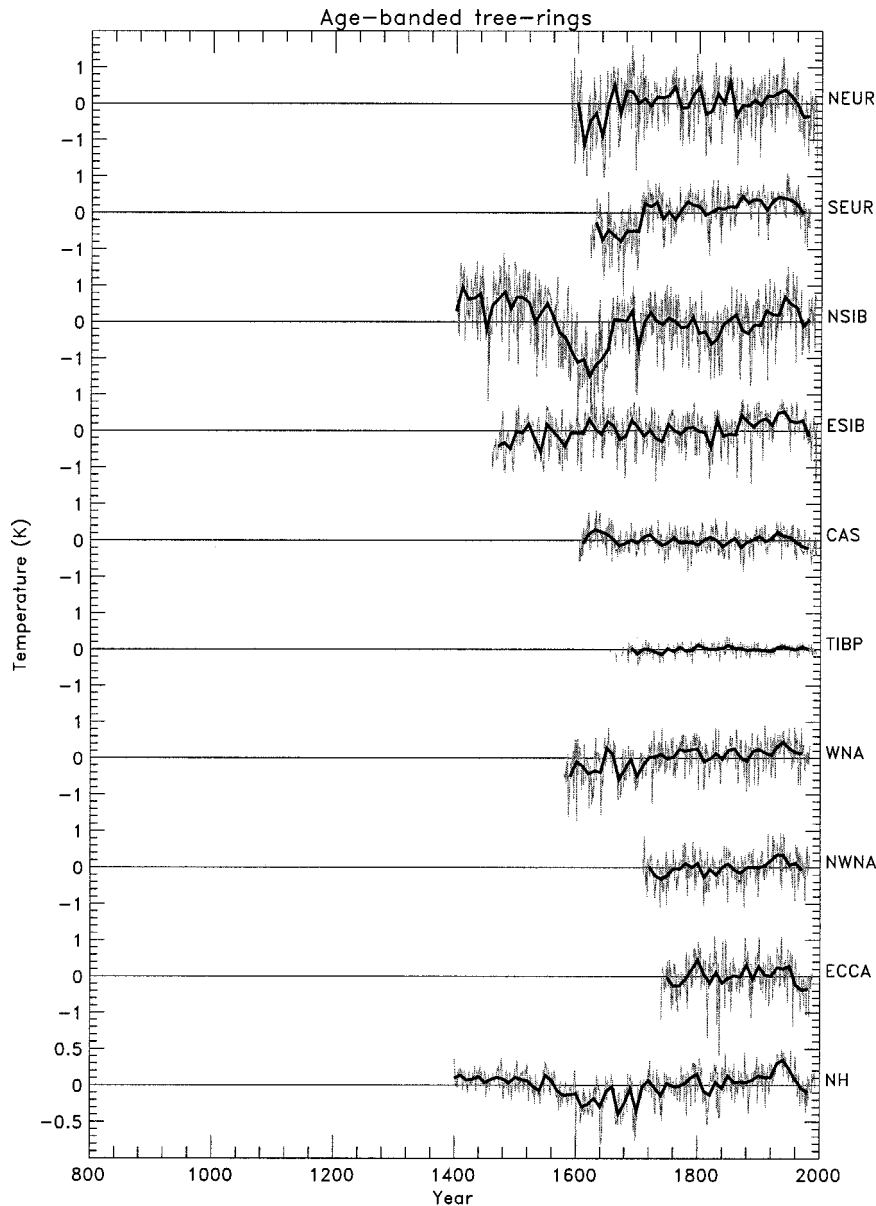


FIG. 3. The calibrated ABD tree-ring series for the nine regions and for NH (see Fig. 1 for the definitions of the regions). The gray lines are the annually resolved data and the black lines are decadal averages formed from these. Note the change in vertical scale for the NH series.

#### 4. Examination of time series

The regional and NH series of the tree-ring temperature data, with the age effect removed by using the standard technique, or removed by using the age-banding technique, and the model series are all shown in Figs. 2–4. A clear difference between the model and standard tree-ring temperature reconstructions is the existence of large negative temperature anomalies on yearly timescales or “negative spikes” in the tree-ring reconstruction. These are most likely due to volcanic eruptions, which can cause widespread cooling if the eruption is violent enough to inject sufficient aerosol

particles into the stratosphere (e.g., Sato et al. 1993). Briffa et al. (1998a) have performed a detailed examination of the signature of volcanic eruptions in the tree-ring chronologies and have related them to historical records. In the 595 years of the standard NH series there are 5 events in which the temperature anomaly drops below 3 standard deviations from the mean, a number that is significantly greater than that expected from a Gaussian process at less than the 1% level. In the 1200-yr simulated NH series there is only one negative  $3\sigma$  event that would have a 73% chance of occurring in a Gaussian time series of the same length (i.e., it is not

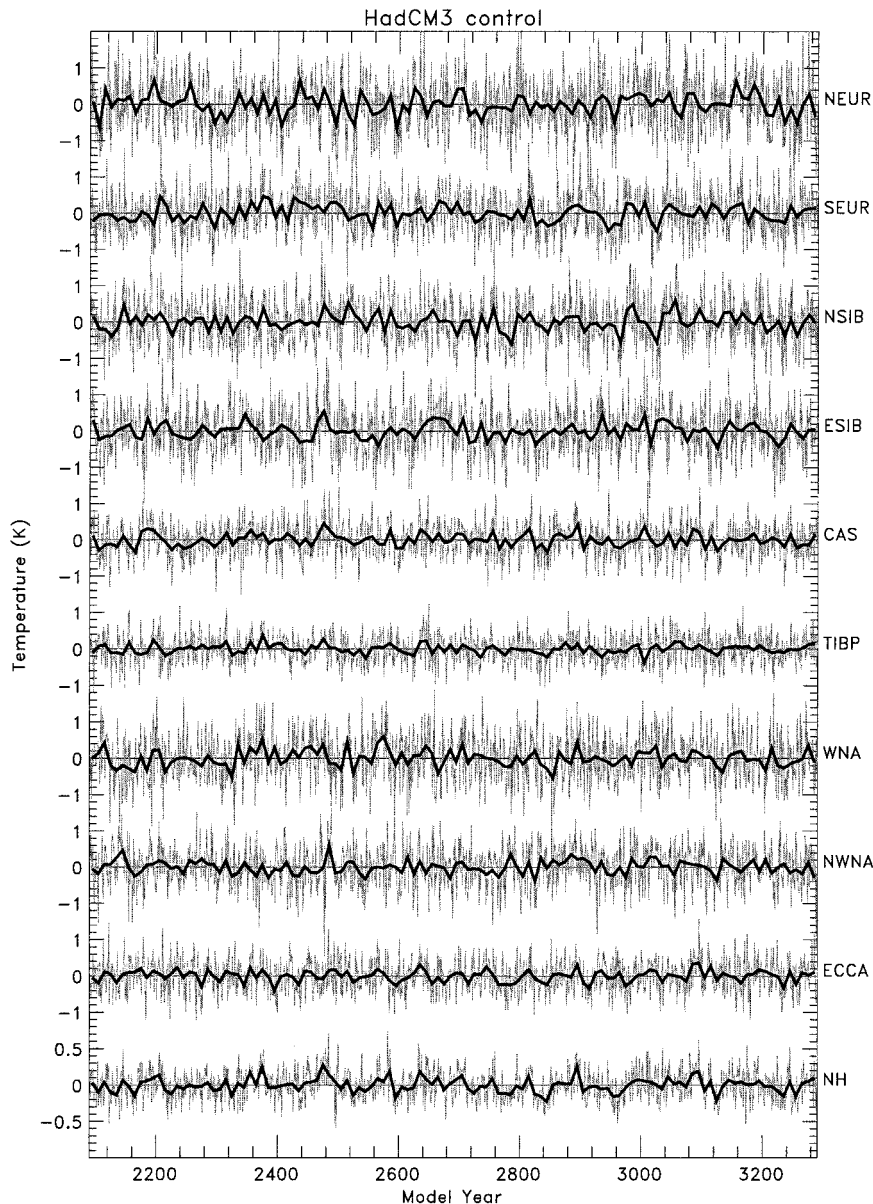


FIG. 4. HadCM3 control simulation Apr-Sep land temperatures for the nine regions and for NH (see Fig. 1 for the definitions of the regions). The gray lines are the annual (i.e., Apr-Sep) data and the black lines are decadal averages formed from these. Note the change in vertical scale for the NH series. Model years are arbitrary but are expressed in terms of a model start year at 1990.

significant). The regional tree-ring temperature reconstructions show similar large negative anomalies while the model region series do not [e.g., there are three  $-3\sigma$  summers in the eastern and central Canada (ECCA) standard tree-ring series (significant at  $<1\%$ ) but only one  $-3\sigma$  summer for the simulated ECCA series (73% chance of occurring)]. The ABD series also show these negative spikes but are not so amenable to the simple analysis of counting negative anomalies because of the complication of removing the low-frequency variability from the series (see later). Thus the model control does

not agree with the tree-ring reconstructions in that it lacks large negative temperature anomalies that are associated with historical volcanic eruptions. Of course this is not surprising as the model is not forced with the radiative effects of such volcanoes. In section 8 we include a parameterization of “naturally forced variability,” including volcanic eruptions, and assess the differences in variability between those simulations and the control.

Figures 2–4 also show the time series averaged into decades, which emphasizes lower-frequency variability



and, to a certain extent, averages out the effects of volcanic eruptions. Throughout the paper we use simple nonoverlapping 10-yr averages in order to avoid complications arising from estimating reduced degrees of freedom in statistical tests and in estimating residual variance in the calibration. The level of variability of the standard tree-ring decadal NH series seems to be in good agreement with the model NH variability. There is clearly more low-frequency variability in the ABD tree-ring series in comparison with the standard tree-ring data, which is expected because of the nature of the ABD technique. For example, there is one particularly large and extended negative anomaly in the ABD northern Siberia (NSIB) series, from 1560 to 1650, which is not evident in the standard NSIB series. Because of this low-frequency behavior, there does not appear to be such a good agreement between the ABD series and the model series as in the case of the standard tree rings. In the following sections we make a quantitative comparison of model and tree-ring temperature reconstructions.

## 5. Comparison of variance

A simple way of comparing the variability of the tree-ring reconstructions and the model is to compute the standard deviation or variance of each series (Fig. 5). As indicated in section 2, there is some uncertainty in the variance of the temperature reconstructions due to residual variance not captured by tree-ring series during the calibration period (the temperature variability not recorded by the trees). Hence we make a comparison with and without this extra variance included in the calculation.

In this section, and in all subsequent sections, we perform a quantitative comparison of the tree-ring reconstructions and the model. We eliminate the problem of the recent decline of tree-ring temperature correlation by using only pre-1960 tree-ring data. There may, however, be some anthropogenic influence in the early part of the twentieth century. Tett et al. 1999 estimate that the early twentieth-century warming is due to a combination of anthropogenic and solar effects, so the tree-ring data may represent a slight overestimate of the natural climate variability pre-1960. We consider this possible error to be small in comparison with other errors due to calibration and the removal of the age effect from the tree cores.

For the regional series, the differences between the standard tree-ring reconstructions and model standard deviations on all timescales (Fig. 5a) are generally statistically significant (because of the large sample sizes) although they are not generally large in absolute terms. The biggest difference is 0.25 K for both the ECCA and NSIB series with residual variance accounted for, leading to an underestimation of the variance by the model in these regions by factors of 2.2 and 1.8, respectively. It seems though that there is no systematic over- or

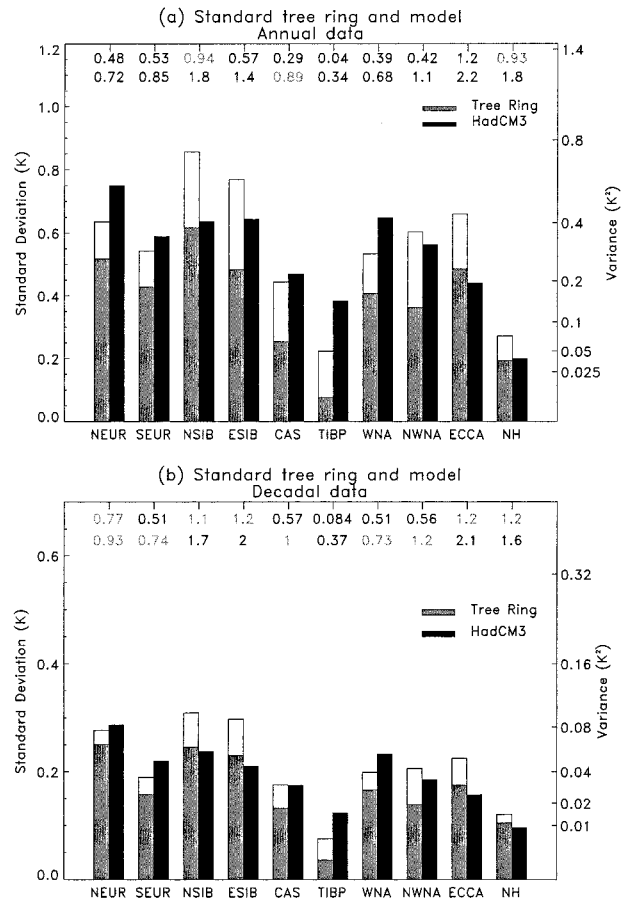


FIG. 5. Standard deviations/variances of the standard tree-ring series and the HadCM3 control series for (a) annual data and (b) decadal data. The tree-ring standard deviations are the gray bars with the residual variance from the calibration period placed on top of these in white. Model standard deviations are the black bars. The ratios of the variances are shown by the numbers above the bars. The upper numbers are the ratios without the residual variance taken into account and the lower numbers with it taken into account. Significantly different ratios (at the 95% level calculated using an  $F$  test) are shown in black and those that are not significant are shown in gray. Allowing for autocorrelation in the  $F$  test makes little difference.

underestimation of variance in the model control simulation, nor is there any obvious regional pattern. This is also true for the decadal standard deviations (Fig. 5b) although some differences are less likely to be statistically significant because of the reduced sample size. Hence there appears to be no consistent bias in the amplitude of the variability simulated by the model control, when compared with the standard tree-ring reconstructions for the regional series. For the NH series, the amplitude of the model variability is statistically indistinguishable from the standard tree-ring-based reconstructions when the residual variance from the calibration procedure is not taken into account. However, when the residual is considered, the model underestimates the NH variance by a factor of 1.8 on all timescales and a factor of 1.6 on decadal and longer timescales.

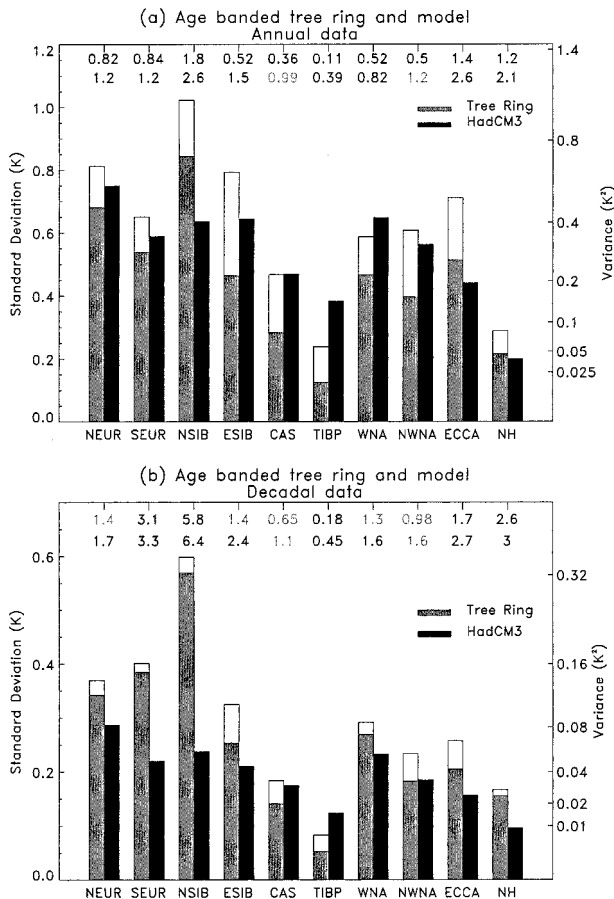


FIG. 6. Standard deviations/variances of the ABD tree-ring series and the HadCM3 control series for (a) annual data and (b) decadal data. The tree-ring standard deviations are the gray bars with the residual variance from the calibration period placed on top of these in white. Model standard deviations are the black bars. The ratios of the variances are shown by the numbers above the bars. The upper numbers are the ratios without the residual variance taken into account and the lower numbers with it taken into account. Significantly different ratios (at the 95% level calculated using an  $F$  test) are shown in black and those that are not significant are shown in gray. Allowing for autocorrelation in the  $F$  test makes little difference.

As is evident by the ABD tree-ring series in Fig. 3, the agreement between the amplitude of variability of the model and the ABD tree rings is less good. On the whole the model underestimates the standard deviation of temperature, especially for the decadal averaged data (Fig. 6b) where the tree-ring reconstructions have more variance for all the time series, except central Asia (CAS) and the Tibetan Plateau (TIBP). The maximum disagreement is with the NSIB series that has around 6 times more variance than the corresponding model series. The standard deviation of the ABD NH tree-ring series, including the residual from the calibration period, is 0.29 K (variance of 0.084 K<sup>2</sup>) compared to 0.20 K (variance of 0.040 K<sup>2</sup>) for the model NH series. This would mean that the variance of the HadCM3 summer NH land temperatures could be underestimated by as

much as a factor of 2.1 on all timescales and a possible factor of 3.0 on timescales of a decade and greater. Hence there is a consistent underestimation of hemispheric-scale variability by the model in comparison with both the ABD tree-ring reconstructions and the tree rings analyzed using the standard method.

## 6. Power spectra

A comparison of standard deviations or variances only highlights differences in the absolute amplitude of variability. The climate system varies differently on different timescales and hence it might be more instructive to compare the relative variance of the model and tree-ring temperature reconstructions at different timescales. This can be achieved by comparing the power spectra of each series. We calculate power spectra by taking the Fourier transform of the autocovariance function of the time series and then applying the Tukey–Hanning window to get a consistent and unbiased estimate (e.g., Chatfield 1984). A window width of 100 yr was used in order to get relatively smooth, bias-free spectra, while retaining the century timescale (the lowest frequency estimated by the procedure corresponds to twice the window width, 200 yr). As in the comparison of variance, account was taken of the residual variance from the calibration procedure. This was assumed to be uncorrelated in time (white) so that the same residual power was added to the tree-ring spectra at each frequency.

Figure 7 shows the power spectra of the model and standard tree-ring NH, northern Europe (NEUR), and ECCA series, and Fig. 8 shows the same for the model compared to the ABD tree-ring reconstructions. The spectra of both the model and the tree-ring series are generally red in character (i.e., there is more power at low frequencies than at high frequencies) although they are not strongly so. We can judge the redness of the spectra by fitting an autoregressive process of order 1 [AR(1) or Markov process defined as  $x_{n+1} = a_1 x_n + a_0 z$ , where the subscripts refer to discrete time intervals (years in this case) and  $z$  is random variable (white noise) with unit variance] to each series in turn (Table 3). In an AR(1) process the  $a_1$  coefficient (i.e., the lag-one autocorrelation) defines the year-to-year “memory” of the system. The ABD tree-ring data has larger  $a_1$  coefficients than the standard data (as expected because of the ABD technique) thus there is a greater level of low-frequency variability compared to high-frequency variability. The  $a_1$  coefficients for both the standard and ABD NH series are relatively larger than nearly all of the corresponding regional series indicating that there is more year-to-year memory on the hemispheric scale compared to the regional scale. In comparison with the model, both the standard and ABD tree rings have greater  $a_1$  coefficients and smaller  $a_0$  coefficients, and hence redder spectra.

There are no obvious peaks in the model or tree-ring spectra shown in Figs. 7 and 8 and testing each indi-

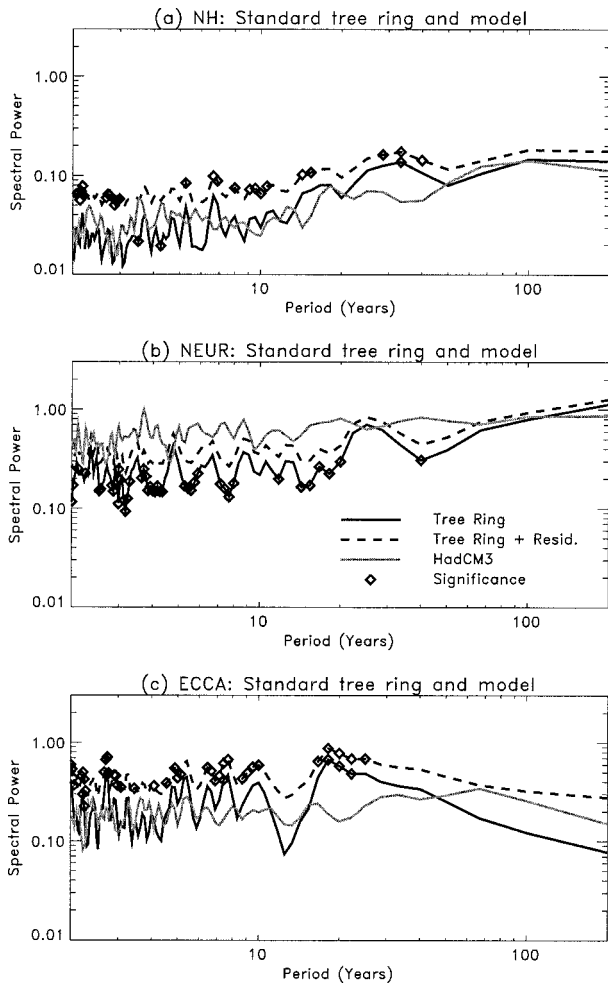


FIG. 7. Power spectra of the model and standard tree-ring series for (a) NH, (b) NEUR, and (c) ECCA. In each graph the model spectra is shown in gray, the calibrated tree-ring spectra is shown as the black solid line and the tree-ring spectra with the residual variance added is shown as the black dashed line. The residuals were assumed to be uncorrelated in time. Where the model and tree-ring spectra are different at the 95% level according to an  $F$  test, the center of the frequency bin is marked with a diamond.

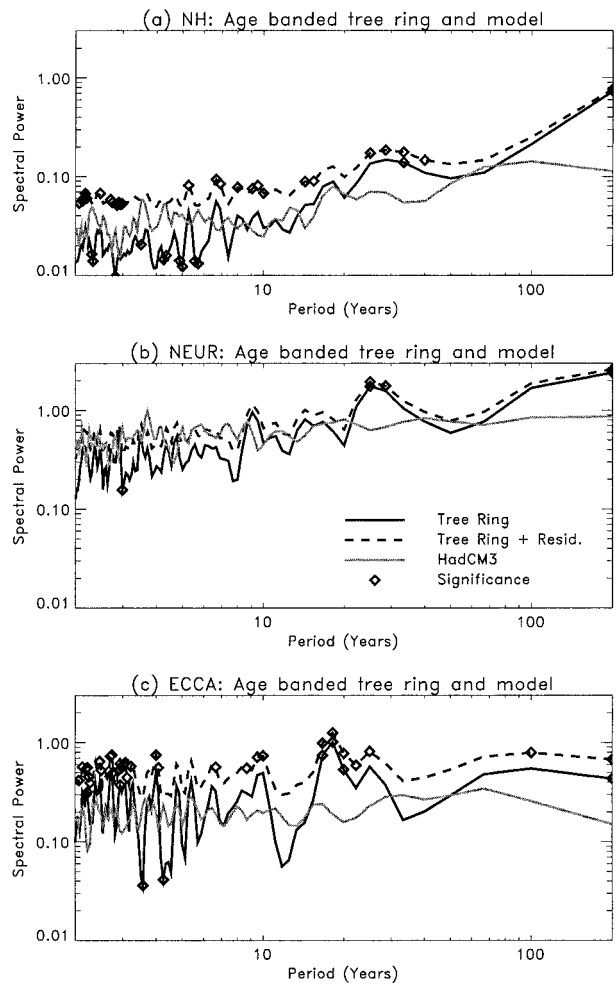


FIG. 8. Power spectra of the model and ABD tree-ring series for (a) NH, (b) NEUR, and (c) ECCA. In each graph the model spectra is shown in gray, the calibrated tree-ring spectra is shown as the black solid line and the tree-ring spectra with the residual variance added is shown as the black dashed line. The residuals were assumed to be uncorrelated in time. Where the model and tree-ring spectra are different at the 95% level according to an  $F$  test, the center of the frequency bin is marked with a diamond.

TABLE 3. Fitted coefficients of an AR(1) process,  $\chi_{n+1} = \alpha_1\chi_n + \alpha_0z$ , where  $z$  is unit variance white noise for the tree ring and HadCM3 control time series.

Region	Standard tree ring		ABD tree ring		HadCM3	
	$\alpha_1$	$\alpha_0$	$\alpha_1$	$\alpha_0$	$\alpha_1$	$\alpha_0$
NEUR	0.17	0.51	0.31	0.65	0.10	0.75
SEUR	0.15	0.42	0.48	0.47	0.058	0.59
NSIB	0.11	0.61	0.41	0.77	0.097	0.63
ESIB	0.35	0.45	0.42	0.42	0.027	0.64
CAS	0.25	0.25	0.23	0.28	0.097	0.47
TIBP	0.11	0.076	0.044	0.12	0.025	0.38
WNA	0.077	0.41	0.26	0.45	0.013	0.65
NWNA	0.0019	0.36	0.069	0.40	0.11	0.56
ECCA	0.12	0.48	0.089	0.51	0.056	0.44
NH	0.35	0.18	0.56	0.18	0.18	0.2

vidual spectrum against a null hypothesis of AR(1) noise gives no consistent peaks above, for example, the 95% confidence level. In terms of the detection, attribution and prediction of climate change, the most important timescales are decades to centuries and there are some statistically significant differences between the model and the standard tree-ring reconstructions on these timescales, and many significant differences between the ABD and model spectra. For both the tree-ring NH temperature series (Figs. 7a and 8a) there is significantly more power at the 20–40-yr timescale than the model control. Thus there may be a “mode” of variability of the Northern Hemisphere summer that has a timescale of the order of several decades that is not simulated by the model. We use the term mode loosely as the enhanced power at 20–40 yr in the tree-ring reconstructions can be explained by an AR(1) model, so may not be generated by a cyclic phenomenon. Nevertheless, there is a clear “hump” in the spectra at 20–40 yr in many of the regional tree-ring series (see also Briffa and Schweingruber 1992) and no corresponding features in the model.

For the ABD data there is also significantly more power than the model control at periods of around 100 years and longer. Again, this is unlikely to be a cyclic mode of variability, but if there is real variability in the climate system that is not simulated by the model on these timescales, this will have implications for estimating uncertainties in detection, attribution, and prediction of climate change if the model is used as a surrogate for natural climate variability. As we shall see in section 8 below, there are natural climate forcings that vary on these timescales that may be responsible for the underestimation of variability by the model control.

## 7. Spatial variability

To compare spatial patterns of variability of the model and tree-ring temperature reconstructions we use empirical orthogonal function (EOF) analysis (e.g., North 1984). First we averaged the data into decades as we are interested in patterns of low-frequency variability (again we take pre-1960 data only). The TIBP series was discarded because of its poor performance in the temperature calibration procedure (see section 2; Tables 1, 2). EOFs were computed from the covariance matrix of the eight remaining series where covariances between individual series were calculated using all the contemporaneous data for those series and each series was weighted by relative area. Principal component (PC) time series, which give the time-varying amplitude of the EOF patterns, were formed by projecting the EOFs onto the data only when data in all eight regions exists. This is true for years 1615–1985 of the standard tree-ring data, but only for years 1745–1985 for the ABD data due to the reduced length of the series that results from the age banding of individual trees (see section 2

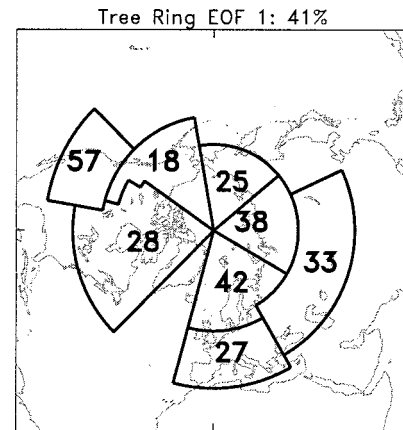


FIG. 9. Leading EOF of the decadal-averaged standard tree-ring series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. The percentage at the top is the percent total variance explained by the leading EOF.

and Briffa et al. 2001a). In comparing the model and tree-ring variance (Fig. 5) and power spectra (Fig. 7) account was taken of the residual variance in the calibration procedure. In computing the EOFs we assume that these residuals are uncorrelated in space and hence do not affect the shape of the EOF (although they will affect the fraction of the variance captured by the leading EOF). All spatial patterns shown are orthonormal so that the dot product of the pattern with itself is unity.

The leading EOF of the standard tree-ring reconstructions (Fig. 9) explains 41% of the total variance and has positive loadings for all regions. Thus there is a mode of variability that involves coherent temperature anomalies of the same sign over the whole Northern Hemisphere land area. Again, we use the term mode loosely as there is no evidence for any cyclic behavior as the time coefficient of the leading EOF is statistically indistinguishable from an AR(1) process. The second EOF explains 16% of the variance and the third EOF explains 13%. Typical decadal anomalies associated with a 1 standard deviation anomaly in the amplitude of EOF1 are, for example, 0.13 K in the western North America (WNA) region and 0.04 K in the northwestern North America (NNA) region. There is no obvious physically based spatial pattern to the EOF (other than being a hemispheric pattern), and the pattern is stable to truncating each series by up to a half and repeating the analysis. The PC for EOF1 has a correlation with the NH series of 0.98 indicating further the existence of a hemispheric-wide mode of variability.

The leading EOF of the ABD tree-ring reconstructions (Fig. 10) explains 54% of the total variance and has positive loadings in all of the regions. The second EOF explains 17% of the variance and the third EOF explains 11%. A measure of the similarity of EOFs is the dot product of the two vectors that, as the EOFs are constrained to be orthonormal, is 1 for vectors which are identical and 0 for vectors that are orthogonal (i.e.,

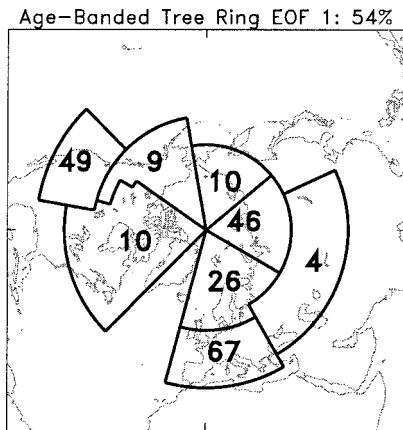


FIG. 10. Leading EOF of the decadal-averaged ABD tree-ring series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. The percentage at the top is as in Fig. 9.

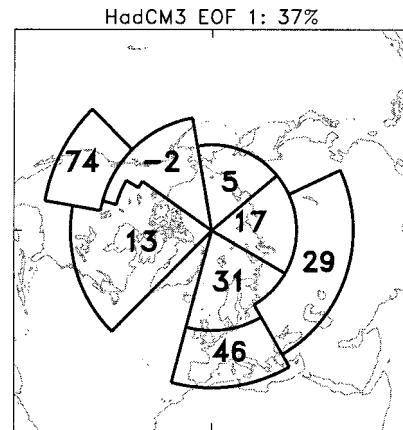


FIG. 11. Leading EOF of the decadal-averaged HadCM3 series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. The percentage at the top is as in Fig. 9.

equivalent to a pattern correlation). The dot product of the standard tree-ring and ABD leading EOFs is 0.81 indicating a high degree of similarity between the leading modes of variability. The PC for the ABD EOF1 has a correlation with the ABD NH series of 0.95.

The leading EOF of the model data (Fig. 11) explains 37% of the total variance and has positive loadings for each time series, with the exception of NWNA, which has a weak negative loading. The second EOF explains 20% of the variance and the third EOF explains 13%, so there is less variance separation than in the case of the tree rings. The PC for the model EOF1 has a correlation of 0.76 with the model NH series, somewhat less than is the case for the tree rings again indicating that the leading EOF is not so dominant a mode for the model. This discrepancy is more apparent when one considers that the tree-ring reconstructions are not completely accurate and that the regions are not fully sampled by the tree-ring network, both of which would tend to reduce the spatial coherence of the tree-ring reconstructions. The dot product of the model EOF with the tree-ring EOFs is 0.89 for the standard tree rings and 0.87 for the ABD tree rings indicating though that the leading modes of variability are similar for the model and the tree-ring reconstructions.

Thus we conclude that for both the tree-ring reconstructions, and for the model control, there is a leading mode of variability that is of one sign over the entire Northern Hemisphere. Moreover there is a large degree of similarity between both tree-ring-leading EOFs and the model-leading EOF.

Taking the EOF analysis of tree-ring and model series and comparing them as above gives some qualitative idea of the similarity or difference in the spatial patterns of variability. The algorithm described by Venzke et al. (1999) and outlined in the appendix, however, provides a more quantitative approach to identifying differences between the model and the tree-ring data. The algorithm essentially maximizes the signal-to-noise ratio (S/N) of

the tree-ring and model data in the spatial domain giving the pattern of variability that is "most different" between the two datasets [Eq. (A6)]. The S/N ratio is a measure of the difference in the spatial patterns of variability between the datasets.

The S/N pattern of the standard tree-ring reconstructions and the model is shown in Fig. 12. It has an eigenvalue (signal-to-noise ratio) of 1.27 and explains 37% of the tree-ring variance and 24% of the model variance. An eigenvalue of 1.27 is not particularly large and is not statistically significant at any high probability based on the significance test outlined in the appendix. The pattern is similar to the leading EOFs of both the model and tree-ring data (dot products are 0.74 and 0.92, respectively). Thus the most different mode of variability between the standard tree-ring data and the model points in the direction of the leading tree-ring EOF, has

Tree Ring and HadCM3 S/N PATTERN 1: 1.27  
TRUNCATION=6, 37% (OBS), 24% (MODEL)

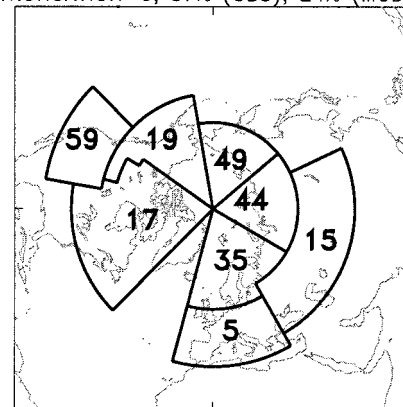


FIG. 12. Leading signal-to-noise maximizing pattern of the standard tree ring vs model series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. Eigenvalue (S/N) and percentage variance explained of the tree-ring and the model are give on at the top.

Age-Banded Tree Ring and HadCM3 S/N PATTERN 1: 6.27  
TRUNCATION=6, 11% (OBS), 11% (MODEL)

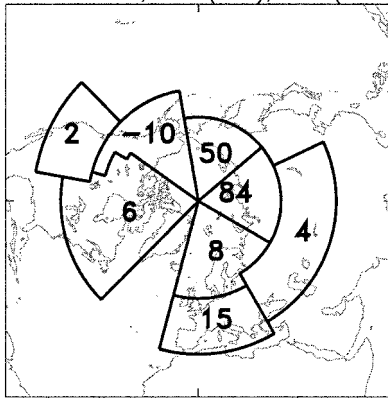


FIG. 13. Leading signal-to-noise maximizing pattern of the ABD tree ring vs model series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. Values at the top are as in Fig. 12.

only small signal-to-noise ratio and is not significantly different to that which might be expected by chance. The time coefficient of the S/N pattern has relatively more power at the 20–40-yr timescales in the standard tree-ring reconstructions than in the model (not shown), a feature that is evident in the power spectra of the NH series (Fig. 7), although this result is somewhat marginal due to the short period of time over which the time coefficients can be estimated (1615–1960).

The S/N pattern of the most different mode between the model and the ABD data (Fig. 13) does however have a significant eigenvalue of 6.27 (i.e., there is more than 6 times as much variance in the ABD tree-ring reconstructions in this pattern in comparison with the model) and is somewhat different from the leading EOFs of the model and ABD tree rings (dot products of 0.31 and 0.57, respectively). The maximum loadings occur over the Siberian region (NSIB and ESIB), which corresponds to a region with a large prolonged negative anomaly during the seventeenth century the like of which is not seen in the model NSIB series (Fig. 3). The spatial pattern and the S/N ratio greater than 6 agrees well with the simple variance comparison (Fig. 6) that found around 6 times more variance in the ABD NSIB series on decadal timescales in comparison with the model.

Thus, as was also indicated by the analysis of the standard deviations and power spectra, there is little difference in the leading spatial patterns of variability of the model and the standard tree-ring reconstructions, apart from an indication of some variability at 20–40 yr that the model does not simulate (although the statistical significance of this result is questionable). There is, however, a significant pattern of variability of the ABD tree-ring reconstructions that the model does not simulate. This pattern is dominated by the Siberian region where the ABD tree rings show a period of negative anomalies that lasted for several decades.

## 8. The contribution of “natural” climate forcing

We have seen that the model control simulation underestimates the variability of summer NH temperatures in comparison with both the standard and the ABD tree-ring reconstructions. Part of this underestimation may be due to what we term naturally forced variability (e.g., changes in total solar irradiance, volcanic eruptions, natural variations in greenhouse gases, land-use changes, etc.) that are not represented in the control simulation. It would be desirable to run HadCM3 with these forcings included over the time period of interest here, the last 600 years, to see if the model could simulate features such as the negative spikes and more prolonged large regional anomalies. There is, however, considerable uncertainty in such forcings (e.g., Crowley and Kim 1999) suggesting that a number of simulations utilizing different forcing data should be considered.

We have been unable to perform such an ensemble because of constraints on computer resources, but we have performed an ensemble of four simulations from 1860 to 1997 that include estimates of solar (Lean et al. 1995) and volcanic (Sato et al. 1993) forcing for use in the detection and attribution of recent climate change. This natural ensemble provides some information about how much extra variance the natural forcings may add to the control estimates of climate variability. Although it should be remembered that the simulations use forcing estimates only for the period 1860–1997, which may exhibit a different level of variability than over the full period from 1400 to the present, which the tree-ring data represent.

The regional and NH series from the natural ensemble are shown in Fig. 14. There are six negative spikes in the four simulation as defined by  $-3\sigma$  anomalies in the natural ensemble corresponding to volcanic eruptions, compared to only one such event in the control that has no representation of volcanoes and hence occurs by chance. The  $-3\sigma$  definition of a volcano is clearly arbitrary as we might expect the number of volcanoes showing up in the natural ensemble to be a multiple of 4. This clearly highlights the problematic nature of identifying volcanic eruptions purely from their temperature imprint. Careful cross validation with historical and other paleovariables, as in Briffa et al. (1998a), is also desirable.

The NH time series for the four ensemble members show some common variability and the variance of the NH series is significantly greater than the control NH variance (Fig. 15), by a factor of 1.6 for all timescales and a factor of 1.9 for decadal and greater timescales. The regional series also show more variance than the control, although some of the differences are not statistically significant.

We estimate the power spectrum of the natural ensemble by computing the spectrum of each individual member and averaging (Fig. 16). For the 20–40-yr timescale, where the model was found to be lacking in var-

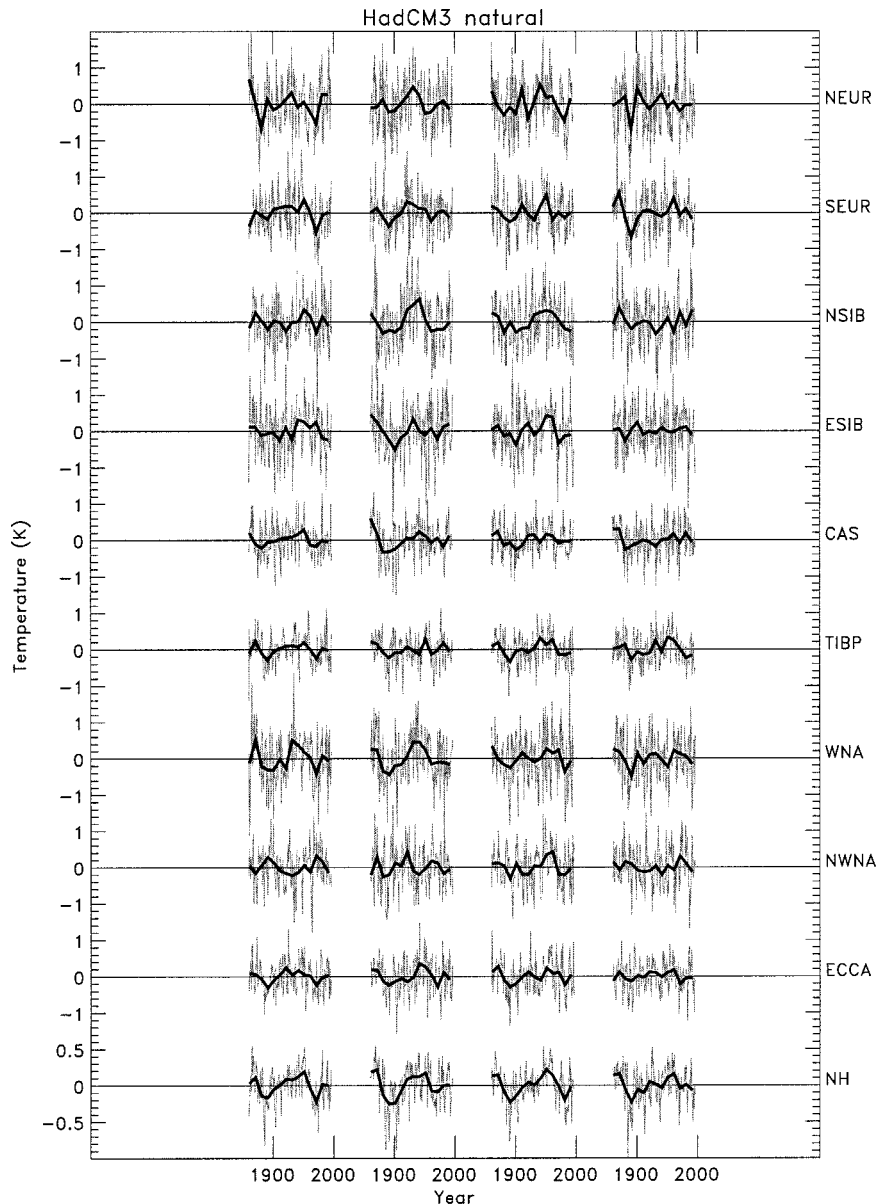


FIG. 14. Regional and NH series from an ensemble of four HadCM3 experiments with solar and volcanic forcing. The gray lines are the annual data and the black lines are decadal averages formed from these. Note the change in scale for the NH series.

iance (Fig. 7), the power spectrum of the NH series of the natural simulations shows some enhancement in comparison with the control although there is little sign of the hump that is apparent in both tree-ring spectra. At low frequencies, there is significant enhancement in the natural ensemble in comparison with the control, indicating that the underestimation of variability by the model on timescales of 100 yr and greater is possibly due to the lack of natural forcings. Again it is difficult to be certain because of the relatively short time period of the natural runs.

Comparing the average AR(1) coefficients of the nat-

ural ensemble (Table 4) with the AR(1) coefficients of the control (Table 3) we see that the natural forcings produce little change in the  $a_0$ , or noise, coefficient. For  $a_1$  coefficients, which give the year-to-year memory or autocorrelation, there are small increases in the regional coefficients in the natural simulations in comparison to the control, and a more than doubling of the coefficient for the NH series. Hence the naturally forced runs have more than twice the hemispheric year-to-year memory of the control simulation for summer land temperatures and thus have a redder NH spectrum (Fig. 16).

There is also a significant enhancement of spectral

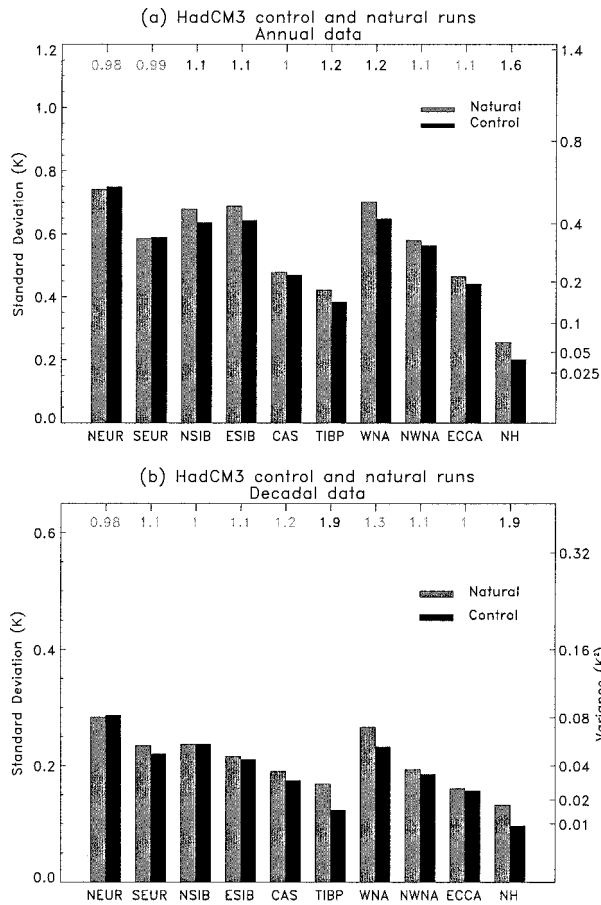


FIG. 15. Standard deviations of the HadCM3 natural ensemble and the HadCM3 control series for (a) annual data and (b) decadal data. The natural ensemble standard deviations are the gray bars and the control standard deviations are the black bars. The ratios of the variances are shown by the numbers above the bars. Significantly different ratios (at the 95% level calculated using an  $F$  test) are shown in black and those that are not significant are shown in gray.

power in the natural NH series around the 10-yr timescale in comparison with the control. While it is impossible to assess the individual contribution of the solar and volcanic forcing without running a further 2 ensembles with each forcing separately, a spectral analysis of the forcing series shows that, around the 10-yr timescale, much of the forcing comes from the volcanic term not from the 11-yr solar cycle.

The EOF analysis of the natural ensemble (Fig. 17) produces a leading mode of variability that is very similar to the leading EOF of the control simulation (dot product of 0.98), but explains more of the total variance (55%). The most different or S/N pattern analysis (Fig. 18) shows no significant mode of variability that is different from the control (the pattern has a signal-to-noise ratio of only 1.52 and a dot product of 0.97 with the control EOF1). Thus the solar and volcanic forcing project onto, and enhance, the natural hemisphere-wide mode of model climate variability, but do not force a

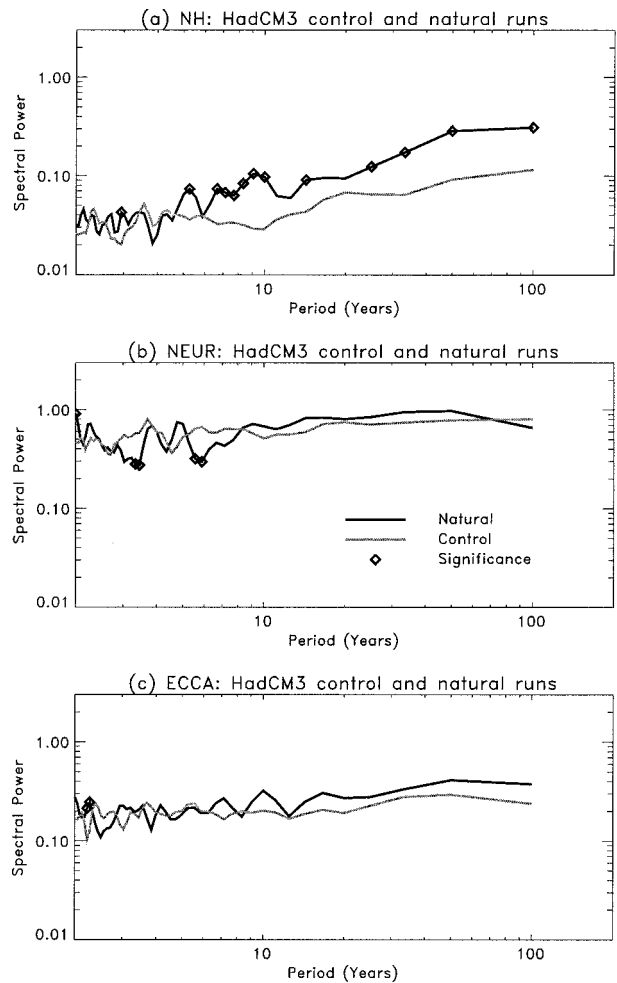


FIG. 16. Power spectra of the natural ensemble and control for (a) NH, (b) NEUR, and (c) ECCA. In each graph the control spectra is shown in gray and the natural spectra in black. Where the control and natural spectra are different at the 95% level according to an  $F$  test, the center of the frequency bin is marked with a diamond.

mode of variability that is significantly different from the internal variability of the model. Thus we may tentatively conclude that, although the variance is enhanced at low frequencies in the natural ensemble, the pattern of variability that is different between the ABD tree-ring reconstructions and the model (Fig. 13) is a real pattern of variability of the climate system that is not represented by the model and thus the model is inadequate in its simulation of variability. This conclusion is tentative because of uncertainties in the tree-ring data and because the natural ensemble is only valid for the 1860–1997 period, not over the full period of tree-ring data, 1400–1994.

## 9. Conclusions

We have compared the variability of a control run of a coupled atmosphere–ocean model with the variability



TABLE 4. Average fitted coefficients of an AR(1) process,  $\chi_{t+1} = \alpha_1 \chi_t + \alpha_0 z$ , where  $z$  is unit variance white noise for the HadCM3 natural ensemble.

Region	HadCM3 natural ensemble	
	$a_1$	$a_0$
NEUR	0.1	0.74
SEUR	0.071	0.58
NSIB	0.11	0.67
ESIB	0.05	0.69
CAS	0.1	0.47
TIBP	0.14	0.42
WNA	0.068	0.7
NWNA	0.21	0.57
ECCA	0.13	0.46
NH	0.4	0.23

of the real climate system inferred from a network of temperature-sensitive tree-ring densities. Two methods of combining the records from individual tree cores were used. The first (standard method) removes any age-dependent signal by fitting and removing a curve for each record, but potentially loses climate fluctuations on timescales of the order of the lifetime of the tree (Briffa et al. 2001b). The second (age-band decomposition) attempts to retain this low-frequency behavior, but at the cost of lower reliability when there are fewer tree cores (Briffa et al. 2001a). The model control is unforced by changes in greenhouse gases, solar variability etc., so thus represents only the “internal” variability of the climate system.

At the regional level, the variability of the model compares reasonably well with the standard tree-ring reconstructions in terms of the absolute magnitude of the variance, the frequency distribution of the variance and the spatial pattern of the variability. However, the model underestimates the mean Northern Hemisphere land variance by a factor of 1.8 on all timescales and a factor of 1.6 on decadal and greater timescales if we take into account residual temperature variance not captured during the calibration of the tree-ring data. Also, the model lacks large single-year negative anomalies associated with volcanic eruptions that are not represented in the control simulation. There is an indication that the tree-ring reconstructions have some variability on timescales of 20–40 yr that the model lacks, although this is unlikely to be a cyclic mode of variability of the climate system as it could easily be explained by an AR(1) process.

The variability of the model does not compare as well to the ABD tree-ring reconstructions. The model underestimates the mean Northern Hemisphere land variance by a factor of 2.1 on all timescales and a factor of 3.0 on decadal and greater timescales taking into account residual variance from the calibration procedure. As well as the negative anomalies associated with volcanoes and the 20–40-yr variability differences seen in the comparison with the standard tree-ring data, there is significantly more variance in the ABD tree-ring re-

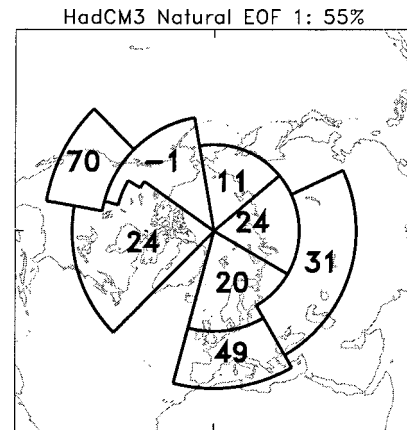


FIG. 17. Leading EOF of the decadal averaged HadCM3 natural ensemble. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. The percentage at the top is as in Fig. 9.

constructions on timescales of 100 yr—much greater than exhibited by the model control. These long timescales are important for the detection, attribution, and prediction of climate change and there exists the possibility of false claims of climate change detection or unrealistically low uncertainty estimates for future climate predictions if the tree-ring estimates are correct and the model underestimates the magnitude of “natural” climate variability.

It is important to consider uncertainties in the tree-ring reconstructions, when considering the large underestimation of climate variability by the model, and the differences in its spatial pattern in comparison with the ABD tree-ring data. Examination of Figs. 3 and 6 both highlight the large variability of the ABD NSIB series. This is the longest of the regional series and hence has a large influence on the construction of the ABD NH series. It also has the largest variance of all the regional

HadCM3 Natural and HadCM3 S/N PATTERN 1: 1.52  
TRUNCATION=6, 54% (OBS), 35% (MODEL)

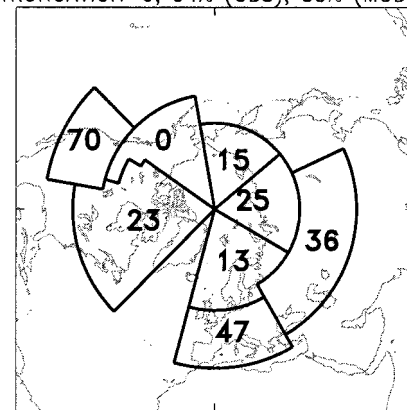


FIG. 18. Leading signal-to-noise maximizing pattern of the model natural ensemble vs model control series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated. Values at the top are as in Fig. 12.

ABD series and shows some considerable low-frequency variations. We have recomputed the ABD NH series substituting the NSIB series from the standard calibration to highlight uncertainties in the tree-ring data. A comparison of this NH series with the model gives a reduction in the ratios of the tree ring to model variance from 2.1 to 1.9 on all timescales and from 3.0 to 2.1 on decadal timescales (taking into account the residual variance of the calibration procedure). The power spectrum of this NH series is reduced at centennial timescales in comparison with the ABD NH power spectrum, but still has significantly more power than the model NH series. The pattern of spatial variability that is most different between the model and the ABD tree rings (Fig. 13) is relatively unchanged on substituting the standard NSIB series for the ABD NSIB series, although the signal-to-noise ratio is reduced from 6.3 to 3.4. Hence, while possible errors in the NSIB series do affect the details of the model–tree ring comparison, they do not significantly alter the conclusion that the model may be underestimating natural climate variability on the hemispheric scale.

The model control experiment only attempts to simulate the internal variability of the climate system. There are variations in “natural forcings” over the time period of the tree-ring data that may contribute to the climate variability. The existence of single-year negative “spikes” in the tree-ring reconstructions show the effect of large volcanic eruptions on climate that are not simulated in the model. We have not been able to produce a simulation of the naturally forced climate of the last 600 years, partly due to restrictions on computer time and partly due to the lack of accurate histories of, for example, variations in solar output, with which to force the model from 1400 to the present day. However, we have analyzed an ensemble of four simulations with solar and volcanic forcing over the period 1860–1997. Inclusion of these forcings produces volcanic negative “spikes,” enhances the NH variance in comparison with the control by a factor of 1.6 on all timescales and a factor of 1.9 on decadal timescales. In terms of the power spectra, there is enhancement of NH variability at 20–40 yr and at the centennial timescale. Hence the lack of natural forcings in the model could be responsible for the underestimation of variability in comparison with the tree-ring data. This conclusion is tentative because the model was not run with natural forcing over the full tree-ring period.

The results of this paper are in good agreement with those of Crowley (2000) who used an energy balance model to estimate the (naturally) forced component of Northern Hemisphere temperature variability and then removed this from the Mann et al. (1998) and Crowley and Lowery (2000) reconstructions. Crowley (2000) found that 41%–64% of the total variance of the reconstructions could be attributed to forcing, with the remainder due to natural internal variability. This would imply that the total preindustrial variance in NH tem-

perature would be 1.7–2.8 greater than natural internal variability alone, very similar to the 1.8–3.0 range estimated here. Crowley (2000) also compares the variability of NH temperatures from HadCM3 and other coupled ocean–atmosphere GCMs with the Mann et al. (1998) and Crowley and Lowery (2000) reconstructions and find that HadCM3 is at the high end of the distribution of combined paleo- and GCM estimates of natural internal variability. This tends to support our hypothesis in this study that it is the absence of natural climate forcings (solar variability and volcanoes) that is responsible for the apparent underestimation of variance by the model in comparison with the tree-ring reconstructions, rather than any fundamental error in the model internally generated variability.

The study raises several recommendations for the future if the proxy record is going to be useful in the validation of climate models.

First, it is important to quantify the uncertainties in the proxy record in order to make quantitative comparisons. In this study we have taken into account the residuals of the calibration in the comparison with the model. In situations where proxy records are calibrated to represent climate variables (e.g., Mann et al. 1998) knowledge of the residuals is important as they account for climate variability not recorded by the proxy. For this study it is the uncertainty in the low-frequency variability (as illustrated by the two alternative approaches used to combine all the tree density records together) that is of most importance; the calibration residuals provide only limited information about uncertainty on these timescales. We recommend further work in quantifying errors in such proxy climate records.

Second, it appears that, at least for the climate variable considered here (summer land temperatures), that there is significant variability that is forced by natural factors such as variations in solar output and volcanic eruptions, over and above the internal variability of the climate system that arises from nonlinear interactions. It is important, therefore, to represent these factors in climate models in order to make a like-with-like comparison. This, in turn, requires estimates (with known error characteristics) of these natural forcings, a procedure that requires analysis of various proxy indicators (e.g., Crowley and Kim 1999).

In the absence of instrumental records of climate, the validation of the variability of climate models on timescales of centuries and longer can only be achieved by recourse to the proxy record. Given the importance of this variability in the detection, attribution, and prediction of future climate, this validation is crucial if the models are to be used as estimates of climate variability. However, there is still considerable progress to be made by both the proxy data community and the modeling community before the wealth of past climate information can be utilized to the fullest.

*Acknowledgments.* Thanks go to Tim Johns who ran the HadCM3 control integration and to Gareth Jones who ran the ensemble of simulations with natural forcings. This work was supported by the U.K. Department of the Environment, Transport and the Regions (PECD/7/12/37) and the Public Meteorological Service Research and Development Programme—MC and SFBT; the U.K. Natural Environmental Research Council (GR3/12107) and the European Union (ENV4-CT95-0127)—TJO, KRB, and FHS.

## APPENDIX

### The Signal-to-Noise EOF Technique of Venzke et al.

The technique developed by Venzke et al. (1999) was used by them to find the common atmospheric response of an ensemble of atmosphere-only model experiments forced by observed SSTs. We reproduce the details of the method here because of our slightly different application, which is to find the patterns of variability that are “most different” between two different datasets, in this case between the model and climate proxy data.

Let  $\mathbf{M}$  be a space–time matrix of model data (e.g., a series of spatial patterns of surface temperatures) and let  $\mathbf{O}$  be a similar space–time matrix computed from the observations (climate proxy). The number of space points must be the same for  $\mathbf{M}$  and  $\mathbf{O}$  but the number of time points can be different.

First perform a singular value decomposition (EOF analysis) of  $\mathbf{M}$ :

$$\mathbf{M} = \mathbf{E}_M \mathbf{\Lambda}_M \mathbf{P}_M^T. \quad (\text{A1})$$

Next define the prewhitening operator,  $\mathbf{F}$ , in the truncated EOF space:

$$\mathbf{F}^T = [(\mathbf{\Lambda}_M)^{(\kappa)}]^{-1} \mathbf{E}_M^{(\kappa)}, \quad (\text{A2})$$

where  $\kappa$  is the number of EOFs retained in the truncation. Then prewhiten  $\mathbf{O}$  with  $\mathbf{F}$ ,

$$\mathbf{O}' = \mathbf{F}^T \mathbf{O}, \quad (\text{A3})$$

and perform an SVD of  $\mathbf{O}'$

$$\mathbf{O}' = \mathbf{E}' \mathbf{\Lambda}' \mathbf{P}'^T, \quad (\text{A4})$$

where  $\mathbf{\Lambda}'$  are the signal-to-noise ratios. Projecting the EOFs back into “real” space,

$$\tilde{\mathbf{E}} = \mathbf{F} \mathbf{E}', \quad (\text{A5})$$

gives the signal-to-noise maximizing filters. The leading spatial pattern that is most different between the model and the observations is then given by

$$\hat{e}_1 = \frac{\mathbf{O} p_1}{\lambda_1'}, \quad (\text{A6})$$

where  $p_1$  is the time series obtained by projecting  $\mathbf{O}$  onto the leading spatial filter  $\tilde{e}_1$ , that is,

$$p_1 = \frac{\mathbf{O}^T \tilde{e}_1}{\lambda_1'}. \quad (\text{A7})$$

See Venzke et al. (1999) for more details.

A subtle, but nevertheless crucial, component of the algorithm is the level of truncation,  $\kappa$ , of the EOF space in the prewhitening operator. Venzke et al. (1999) suggest examining the cumulative ratio of the variance of the model and observations, as a function of  $\kappa$ , for plateaus that suggest the stability of the algorithm. Because of the difference between their application of the algorithm and ours, this is not possible as there are no obvious plateaus. Hence we simply take the pragmatic approach of examining the S/N EOFs at all truncations and picking the maximum  $\kappa$  for which there appears to be a region of stability in the patterns.

Another difference between the Venzke et al. (1999) application of the algorithm and ours is in the magnitude of the signal-to-noise ratios. In Venzke et al. (1999) the maximum S/N ratios were reasonably large (e.g., of order 10). For our application S/N ratios are more likely to be of order unity (unless there is some large difference between the model and the observations that would be easily seen by looking at, e.g., variances) and hence there is a question of statistical significance. To judge statistical significance we substitute the matrix of observed values  $\mathbf{O}$  with randomly selected portions of the model matrix  $\mathbf{M}$  and repeat the analysis. This is then repeated many times (e.g., 1000) to give a population of S/N ratios that can be thought of as to have occurred by chance. We then perform a  $t$  test on the S/N ratio from analysis to get the significance level.

## REFERENCES

- Bräker, O., 1981: Der alterstrend bei jahringdichten und jahringbreiten von Nadelholzern und sein Ausgleich. *Mitt. Forstl. Bundesversuchsanst. Wien*, **142**, 75–102.
- Briffa, K. R., and F. H. Schweingruber, 1992: Dendroclimatic evidence: Northern and Central Europe. *Climate Since A.D. 1500*, R. S. Bradley, and P. D. Jones, Eds., Routledge, 266–392.
- , P. D. Jones, F. H. Schweingruber, W. Karlen, and S. G. Shiyatov, 1996: Tree-ring variables as proxy-climate indicators: Problems with low-frequency signals. *Climate Variations and Forcing Mechanisms of the Last 2000 Years*, P. D. Jones, R. S. Bradley, and J. Jouzel, Eds., Springer, 9–41.
- , —, —, and T. J. Osborn, 1998a: Influence of volcanic eruptions on Northern Hemisphere summer temperatures over the past 600 years. *Nature*, **393**, 450–454.
- , F. H. Schweingruber, P. D. Jones, T. J. Osborn, S. G. Shiyatov, and E. A. Vaganov, 1998b: Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature*, **391**, 678–682.
- , T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov, 2001a: Low-frequency temperature variations from a northern tree-ring density network. *J. Geophys. Res.*, **106**, 2929–2941.
- , —, —, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov, 2001b: Tree-ring width and density data around the Northern Hemisphere: Part 1. Local and regional climate signals. *Holocene*, in press.
- Chatfield, C., 1984: *The Analysis of Time Series*. 3d ed. Chapman and Hall, 241 pp.

- Collins, M., S. F. B. Tett, and C. Cooper, 2001: The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **17**, 61–81.
- Cook, E. R., K. R. Briffa, D. M. Meko, D. A. Graybill, and G. Funkhouser, 1995: The “segment-length curse” in long tree-ring chronology development for palaeoclimatic studies. *Holocene*, **5**, 229–237.
- Cox, M. D., 1984: A primitive equation, three dimensional model of the ocean. GFDL Ocean Group Tech. Rep. 1, 141 pp.
- Crowley, T. J., 2000: Causes of climate change over the past 1000 years. *Science*, **289**, 270–277.
- , and K.-Y. Kim, 1999: Modeling the temperature response to forced climate change over the last six centuries. *Geophys. Res. Lett.*, **26**, 1901–1904.
- , and T. S. Lowery, 2000: How warm was the medieval warm period? *Ambio*, **29**, 51–54.
- Cullen, M. J. P., 1993: The unified forecast/climate model. *Meteor. Mag.*, **122**, 81–94.
- Delworth, T. L., and M. E. Mann, 2000: Observed and simulated multidecadal variability in the Northern Hemisphere. *Climate Dyn.*, **16**, 661–676.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 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 Dyn.*, **16**, 147–168.
- Harris, R. N., and D. S. Chapman, 2001: Mid-latitude (30°–60°N) climatic warming inferred by combining borehole temperatures with surface air temperatures. *Geophys. Res. Lett.*, **28**, 747–750.
- Huang, S., H. N. Pollack, and P. Y. Shen, 2000: Temperature trends over the past five centuries reconstructed from borehole temperatures. *Nature*, **403**, 756–758.
- Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean–atmosphere GCM: Model description, spinup and validation. *Climate Dyn.*, **13**, 103–134.
- Jones, P. D., K. R. Briffa, T. P. Barnett, and S. F. B. Tett, 1998: High-resolution paleoclimatic records for the last millennium: Interpretation, integration and comparison with general circulation model control-run temperatures. *Holocene*, **8**, 455–471.
- Lean, J., J. Beer, and R. Bradley, 1995: Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.*, **22**, 3195–3198.
- Manley, G., 1974: Central England temperatures, monthly means 1659 to 1973. *Quart. J. Roy. Meteor. Soc.*, **100**, 389–405.
- Mann, M. E., R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392**, 779–787.
- North, G. R., 1984: Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, **41**, 879–887.
- Oerlemans, J., and B. K. Reichert, 2000: Relating glacier mass balance to meteorological data by using a seasonal sensitivity characteristic. *J. Glaciol.*, **46**, 1–6.
- Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central England temperature series. *Int. J. Climatol.*, **12**, 317–342.
- Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. *Climate Dyn.*, **16**, 123–146.
- Reichert, B. K., L. Bengtsson, and O. Akesson, 1999: A statistical modeling approach for the simulation of local paleoclimatic proxy records using general circulation model output. *J. Geophys. Res.*, **104**, 19 071–19 083.
- Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths (1850–1990). *J. Geophys. Res.*, **98** (D12), 22 987–22 994.
- Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 3282–3294.
- Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth century temperature change near the earth’s surface. *Nature*, **399**, 569–572.
- Venzke, S., M. R. Allen, R. T. Sutton, and D. P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562–2584.