Hadley Centre
for Climate Predictions and Research

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

by:

Matthew Collins
Timothy J.Osborn ,Simon F.B. Tett,
Keith R. Briffa and Fritz H. Schweingruber

HCTN 16

Hadley Centre Technical Note

March 2000

The Met.Office

A comparison of the variability of a climate model with palaeo-temperature estimates from a network of tree-ring densities.

Matthew Collins[1],

Timothy J. Osborn[2], Simon F. B. Tett[1],

Keith R. Briffa[2] and Fritz H. Schweingruber[3].

March 31, 2000

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

[2]Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, UK.

[3]Swiss Federal Institute for Forest and Snow Research, Zurcherstrasse 111, CH-8903 Birmensdorff, Switzerland.

Corresponding Author: Matthew Collins

Hadley Centre for Climate Prediction and Research,

The Met. Office, London Road, Bracknell, RG12 2SZ, UK.

Tel: (44) 01344 856370

Fax: (44) 01344 854898

matcollins@meto.gov.uk

Abstract

Validation of the decadal to centennial time scale 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 coupled ocean-atmosphere 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 we use temperature reconstructions based on two different methods of removing the bias due to changing tree age. The first is a standard method that potentially also removes the low-frequency variability on time scales of the order of the tree life-span. The second (age-banding) maintaines 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, although the model may underestimate northern hemisphere land temperature total variance by as much as a factor of 1.8. Agreement with the age-banded tree-ring reconstructions is less good with the model underestimating the hemispheric variance by as much as a factor of 2.1 on all time scales and by as much as a factor of 2.7 on decadal to centennial time scales. Underestimation of the natural variability of climate by the model is serious as it may lead to false claims of climate change detection or erroneously low uncertainty estimates in future climate predictions. 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

Multi-century integrations of coupled climate models are now routinely produced at climate modelling centres. Such simulations are performed for various purposes, 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 (2000)); 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 validating the mean climate of the model, but also validating the climate variability over a range of space and time scales.

In order to validate the variability of a model on time scales of centuries one requires multi-century 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 which is representative of, at best, temperatures at the grid box size (e.g. 300km) and may be more representative of temperatures on the scales of many grid boxes (due to model formulation which restricts the accumulation of energy at the grid scale e.g. 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) and many others) which are widely spread over the globe. While these proxy records are 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 normalised 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 behaviour to the proxy data in terms of their principal spatial patterns of variability, while the other model showed markedly different behaviour. Crowley and Kim (1999)

compared the spectra of global temperature variability of two coupled models with the multi-proxy reconstructions of Mann et al (1998) and found them to agree at the 90% level once they had removed the naturally-forced variability from the Mann et al (1998) data using an energy balance model. Delworth and Mann (1999) 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 represent large-scale climate variables that the model can simulate, such as temperature and precipitation, prior to the comparison (e.g. the multi-proxy reconstructions of Mann et al (1998)), with care taken to account for any 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 (e.g. glacier length (Oerlemans and Reichert (1999)).

In the study described here, a multi-century integration of the third version of the Hadley Centre coupled model (HadCM3 - Gordon et al (2000)) is compared with a number of regional temperature timeseries, estimated from a network of tree-ring density chronologies spread over over the northern hemisphere (Briffa et al (2000b)), primarily representative of summer (half year) 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. The validation of HadCM3 during past epochs of very different climate, such as the last glacial maximum, will be presented elsewhere. 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 which are more representative of the space scales which 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 (2000b)). Correspondingly, we sub-sample the model in the nine tree-ring regions during the optimal (April-September) growth season of the trees (Briffa et al (2000b)), 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 maximise the potential temperature signal. Briffa et al (2000b) show that, for this network, a much more reliable temperature reconstruction can be obtained from the density of the wood formed towards 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 data set as a whole is the growing season temperature (Briffa et al (2000b)). 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 (2000b) combine the chronologies into nine regional averages (also shown in fig. 1), and calibrate them against April to September temperatures for each of the regions, using simple linear regression over the period 1881-1960. The correlations between the regional tree-ring density records and the temperatures are reproduced in table 1, together with the standard deviations of the residual temperature variations not captured by the tree-ring density (RMSE). Over the calibration period, the variance of each reconstruction is less than the variance of the temperature time series, with the difference between the variances being 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. The residuals exhibit significant autocorrelation for four out of the nine regions, but the strength of the autocorrelation is uncertain - particularly with regard to how representative the calibration period is of the pre-1881 period - so we make the assumption that

all the residuals are white noise (this simplifies the inclusion of the residuals in the comparison of inter-decadal variance and power spectra).

Briffa et al (2000b) 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 (which depends upon the number of chronologies with data for a particular year, and on the mean correlation between pairs of chronologies within the region). This record was calibrated against the observed April to 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 (2000b) show that for the NH reconstruction, and for seven out of the nine sub-regions, the correlation with temperature is higher at the decadal time scale than that given in table 1.

The calibrated temperature reconstructions for the nine sub-regions 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 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 generalised exponential function (a "Hugershoff" function, Bräker (1981)) was fitted to each tree core and removed. This detrending technique is known as "standardisation", and results in a loss of multi-century 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 multi-century variability. To broaden our comparisons we also use the reconstructions of Briffa et al (2000a). These are based on the same tree-ring density data set, and are for the same regions as the standard reconstructions. The difference is that the tree

cores are not standardised (i.e., detrended) and, therefore, lose no low-frequency variability. The age-dependence is 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 normalisation has been applied to remove differences in absolute means and variances. Full details of the procedure are given in Briffa et al (2000a).

These age-banded regional chronologies are calibrated in the same way as the standard reconstructions. The only exception is that the age-banded NH reconstruction is obtained by a principal components regression of the nine regional series. The calibration statistics (reproduced from Briffa et al (2000a)) are given in table 2, while the time series are shown in fig. 3. There is no artificial loss of multi-century 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 age-banded technique requires more tree cores). As described in Briffa et al (2000a), 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 age-banded series, compared to the standard series (fig. 3 cf. fig. 2).

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 “age-banded”.

3 HadCM3

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

description of the climate variability simulated by the model is given by Collins et al (2000).

We analyse the HadCM3 control simulation in which all concentrations of greenhouse gases and aerosols etc. are set as constants representative of the pre-industrial era. Surface air temperatures (at a height of 1.5m) 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 sub-sampled according to the regions defined in fig. 1 to form nine regional time series of model temperatures of length 1200 years (the first 100 years of the experiment were ignored because of spin-up effects). The model "NH" series was formed by averaging all the April-September land temperatures north of 20°N. This is in contrast to the reconstructed tree-ring "NH" series which 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°N). A simple mean of the nine regional series from the model explains 72% of the variance of the true "NH" series at all time scales and 90% of the variance on decadal time scales. 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.

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 time scales or "negative spikes" in the tree-ring reconstruction. These are most likely due to volcanic eruptions, which can cause wide-spread cooling over the northern hemisphere land 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 three standard deviations from the mean, a number which is significantly greater than that which would be expected from a Gaussian process at less than the 1% level. In the 1200 year model NH series there is only

one negative 3σ event which would have a 73% chance of occurring in a Gaussian time series of the same length. The regional tree-ring temperature reconstructions show similar large negative anomalies while the model region series do not (e.g. there are three -3σ summers in the ECCA standard tree-ring series (significant at $<1\%$) but only one -3σ summer for the model ECCA series (73% change of occurring)). The age-banded 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 which 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 parametrisation 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. 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 age-banded tree-ring series in comparison with the standard tree-ring data, which is expected because of the nature of the age-banding technique. For example, there is one particularly large and extended negative anomaly in the age-banded NSIB series, from 1560 to 1650, which is not evident in the standard NSIB series. Because of this low-frequency behaviour, there does not appear to be such a good agreement between the age-banded 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. Hence we make a comparison with and without this extra variance included in the calculation.

For the regional series, the differences between the standard tree-ring reconstructions and model standard deviations on all time scales (fig. 5(a)) are generally statistically significant (because of the large sample sizes) although they are not generally large in absolute terms. The biggest difference is 0.25K 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 under- estimation of variance in the model control simulation, nor is there any obvious regional pattern. This is also true for the decadal standard deviations (fig. 5(b)) 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, in comparison with the standard tree-ring reconstructions for the regional series. For the NH series, the amplitude of the model variability is statistically indistinguisable 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 time scales and a factor of 1.5 on decadal and longer time scales.

As is evident by the age-banded tree-ring series in fig. 3, the agreement between the amplitude of variability of the model and the age-banded tree-rings is less good. On the whole the model underestimates the standard deviation of temperature, especially for the decadally averaged data (fig. 6(b)) where the tree-ring reconstructions have more variance for all the time series, except CAS and TIBP. The maximum disagreement is with the NSIB series which has over 5 times more variance than the corresponding model series. The standard deviation of the age-banded NH tree ring series, including the residual from the calibration period, is 0.29K (variance of $0.084K^2$) compared to 0.20K (variance of $0.040K^2$) 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 time scales and a possible factor of 2.7 on time scales of a decade and greater. Hence there is a consistent underestimation of hemispheric scale variability by the model in comparison with both the age-banded tree-ring reconstructions and the tree-rings analysed 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 time scales and hence it might be more instructive to compare the relative variance of the model and tree-ring temperature reconstructions at different time scales. 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 years was used in order to get relatively smooth, bias-free spectra, while retaining the century time scale (the lowest frequency estimated by the procedure corresponds to twice the window width, 200 years). 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, NEUR and ECCA series, and fig. 8 shows the same for the model compared to the age-banded 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 Auto-regressive 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 age-banded tree-ring data has larger a_1 coefficients than the standard data (as expected because of the age-banding 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 age-banded 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 age-banded tree-rings have greater a_1 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 individual 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 time scales are decades to centuries and there are some statistically significant differences between the model and the standard tree-ring reconstructions on these time scales, and many significant differences between the age-banded and model spectra. For both the tree-ring NH temperature series (fig. 7(a) and 8(a)) there is significantly more power at the 20-40 year time scale than the model control. Thus there may be a "mode" of variability of the northern hemisphere summer which has a time scale of the order of several decades which is not simulated by the model. We use the term "mode" loosely as the enhanced power at 20-40 years 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 years in many of the regional tree-ring series (see also Briffa and Schweingruber (1992)) and no corresponding features in the model.

For the age-banded 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 which is not simulated by the model on these time scales, this will have 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 which vary on these time scales 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)). Firstly the data were averaged into decades as we are interested in patterns of low frequency variability. The TIBP series was discarded because of its poor performance in the temperature calibration procedure (see section 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 age-banded data, due to the reduced length of the series that results from the age-banding of individual trees (see section 2 and Briffa et al (2000a)). 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 39% of the total variance and has positive loadings for all regions. Thus there is a "mode" of variability which 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 behaviour. The second EOF explains 19% of the variance and the third EOF explains 16%. Typical decadal anomalies associated with a one standard deviation anomaly in the amplitude of EOF 1 are, for example, 0.13K in the WNA region and 0.04K in the NWNA 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 EOF 1 has a correlation with the NH series of 0.96 indicating further the existence of a hemispheric-wide mode of variability.

The leading EOF of the age-banded tree-ring reconstructions (fig. 10) explains 53% of the total variance and has positive loadings in all but one of the regions, CAS, which has a weak negative loading. The second EOF explains 16% of the variance and the third EOF explains 11%. A measure of the similarity of EOFs is the dot product of the two vectors which, as the EOFs are constrained to be orthonormal, is 1 for vectors which are identical and 0 for vectors which are orthogonal (i.e. equivalent to a pattern correlation). The dot product of the standard tree-ring and age-banded leading EOFs is 0.81 indicating a high degree of similarity between the leading "modes" of variability. The PC for the age-banded EOF 1 has a correlation with the age-banded 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 EOF 1 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 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 0.90 for the standard tree-rings and 0.86 for the age-banded 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 recnstructions, and for the model control, there is a leading mode of variability which 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 maximises the signal-to-noise ratio (S/N) of the tree-ring and model data in the spatial domain giving the pattern of variability which is "most different" between the two data sets (equation 6). The S/N ratio is a measure of the difference in the spatial patterns of variability between the data sets.

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 1.29 and explains 34% of the tree-ring variance and 24% of the model variance. An eigenvalue of 1.29 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 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 year time scales 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 (1615-1985).

The S/N pattern of the "most different" mode between the model and the age-banded data (fig. 13) does however have a significant eigenvalue of 5.64 (i.e., there is more than 5 times as much variance in the age-banded tree-ring reconstructions in this pattern in comparison with the model) and is somewhat different from the leading EOFs of the model and age-banded tree-rings (dot products of 0.45 and 0.64 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 17th century, the like of which is not seen in the model NSIB series (fig. 3).

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 years which the model does not simulate (although the statistical significance of this result is questionable). There is, however, a significant pattern of variability of the age-banded tree-ring reconstructions that the model does not represent at all. This pattern is dominated by the Siberian region where the age-banded tree-rings show a period of negative anomalies which 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 age-banded 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.) which 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 1860-1997 which include estimates of solar (Lean et al (1995)) and volcanic (Sato et al (1993)) forcing for use in the detec-

tion 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 present that the tree-ring data represent.

The regional and NH series from the natural ensemble are shown in fig. 14. There are six "negative spikes" or -3σ events in the natural ensemble corresponding to volcanic eruptions, compared to only one such event in the control which has no representation of volcanoes and hence occurs by chance. 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 time scales and a factor of 1.9 for decadal and greater time scales. 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 year time scale, where the model was found to be lacking in variance (fig. 7), the power spectrum of the NH series of the natural simulations shows some enhancement in comparison with the control, indicating that this 20-40 variability could be forced. However, there is little sign of the "hump" which 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 time scales of 100 years and greater is possibly due 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 natural 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 power in the natural NH series around the

10 year time scale in comparison with the control. While it is impossible to assess the individual contribution of the solar and volcanic forcing without running a further two ensembles with each forcing separately, a spectral analysis of the forcing series shows that, around the 10 year time scale, much of the forcing comes from the volcanic term not from the 11 year solar cycle.

The EOF analysis of the natural ensemble produces a leading mode of variability which 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 shows no significant mode of variability which is different from the control (the pattern has a signal-to-noise ration of only 1.52 and a dot product of 0.97 with the control EOF 1). Thus the solar and volcanic forcing project onto and enhance the natural hemisphere wide mode of model climate variability, but do not force a mode of variability which 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 which is different between the age-banded tree-ring reconstructions and the model (fig. 13) is a real pattern of variability of the climate system which 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 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 time scales of the order of the life time of the tree (Briffa et al (2000b)). The second ("age-banding") attempts to retain this low-frequency behaviour, but at the cost of lower reliability when there are fewer tree cores (Briffa et al (2000a)). The model control is unforced by changes in greenhouse gases, solar variability etc. so thus represents only the "internal" variability of the climate system.

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 northern hemisphere land variance by a factor of 1.8 on all time scales and a factor of 1.5 on decadal and greater time scales 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 which are not represented in the control simulation. There is an indication that the tree-ring reconstructions have some variability on time scales of 20-40 which the model lacks, although this is unlikely to be a cyclic mode of variability of the climate system and it could easily be explained by an AR(1) process.

The variability of the model does not compare as well to the age-banded tree-ring reconstructions. The model underestimates the northern hemisphere land variance by a factor of 2.1 on all time scales and a factor of 2.7 on decadal and greater time scales taking into account residual variance from the calibration procedure. As well as the negative anomalies associated with volcanoes and the 20-40 year variability differences seen in the comparison with the standard tree-ring data, there is significantly more variance in the age-banded tree-ring reconstructions on time scales of 100 years - much greater than exhibited by the model control. These long time scales 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 age-banded tree-ring data. Examination of figs. 3, 6 and 8 all highlight the large variability of the age-banded NSIB series. This is the longest of the regional series and hence has a large influence on the construction of the age-banded NH series. It also has the largest variance of all the regional age-banded series and shows some considerable low-frequency variations. We have recomputed the age-banded 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 time scales and from 2.7 to 2.1 on decadal time scales (taking into account the residual variance

of the calibration procedure). The power spectrum of this NH series is reduced at centennial time scales in comparison with the age-banded 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 age-banded tree rings (fig. 13) is relatively unchanged on substituting the standard NSIB series for the age-banded NSIB series, although the signal-to-noise ratio is reduced from 5.6 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 which 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 which are not simulated in the model. We have not been able to produce a simulation of naturally forced climate, partly due to restrictions on computer time and partly due to the lack of accurate histories of e.g. variations in solar output, with which to force the model from 1400 to the present day. However, we have analysed an ensemble of four simulations with solar and volcanic forcing over the period 1860-1997. Inclusion of these forcings enhances the NH variance in comparison with the control by a factor of 1.6 on all time scales and a factor of 1.9 on decadal time scales. In terms of the power spectra, there is enhancement of NH variability at 20-40 years and at the centennial time scale. Hence the lack of natural forcings in the model could be responsible for the underestimation of variability in comparison with the tree-rind data. This conclusion is tentative because the model was not run with natural forcing over the full tree-ring period.

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

Firstly, 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. 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 time scales. We recommend

further work in quantifying errors in such proxy climate records.

Secondly, it appears that, at least for the climate variable considered here (summer land temperatures), that there is significant variability which is forced by natural factors such as variations in solar output and volcanic eruptions, over and above the internal variability of the climate system which arises from non-linear 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 which 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 time scales 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 predicting 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 modelling community before the wealth of past climate information can be utilized to the full.

10 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 data sets, 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 an Singular Value Decomposition (EOF analysis) of $\mathbf{M}$.

$$\mathbf{M} = \mathbf{E}_M \mathbf{\Lambda}_M \mathbf{P}_M^T \tag{1}$$

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

$$\mathbf{F}^T = ((\mathbf{\Lambda}_M)^{(\kappa)})^{-1} \mathbf{E}_M^{(\kappa)} \tag{2}$$

where κ is the number of EOFs retained in the truncation. Then pre-whiten $\mathbf{O}$ with $\mathbf{F}$,

$$\mathbf{O}' = \mathbf{F}^T \mathbf{O}, \tag{3}$$

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

$$\mathbf{O}' = \mathbf{E}' \mathbf{\Lambda}' \mathbf{P}'^T \tag{4}$$

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

$$\tilde{\mathbf{E}} = \mathbf{F}\mathbf{E}', \tag{5}$$

gives the signal to noise maximising filters. The leading spatial pattern which is "most different" between the model and the observations is then given by

$$\hat{e}_1 = \frac{\mathbf{O}p_1}{\lambda_1'}, \tag{6}$$

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

$$p_1 = \frac{\mathbf{O}^T \tilde{e}_1}{\lambda_1'}. \tag{7}$$

See Venzke et al (1999) for more details.

A subtle, but nevertheless crucial, component of the algorithm is the level of truncation, κ, of the EOF space in the pre-whitening operator. Venzke et al (1999) suggest examining the cumulative ratio of the variance of the model and observations, as a function of κ, for plateaus which 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 κ 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 which would be easily seen by looking at, for example, 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.

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 UK Department of the Environment, Transport and the Regions (PECP/7/12/37) and the Public Meteorological Service Research and Development Programme - MC and SFBT; the UK Natural Environmental Research Council (GR3/12107) and the European Union (ENV4-CT95-0127) - TJO, KRB, and FHS.

References

Bräker O (1981) Der alterstrend bei jahrringdichten und jahrringbreiten von Nadelholzern und sein Ausgleich. Mitt forstl Bundesvers Anst Wien 142:75–102.

Briffa KR, Schweingruber FH (1992) Dendroclimatic evidence: Northern and Central Europe. In Bradley RS, and Jones PD, editors, Climate Since A.D. 1500 266–392. Routledge, London.

Briffa KR, Jones PD, Schweingruber FH, Karlen W, Shiyatov SG (1996) Tree-ring variables as proxy-climate indicators: Problems with low-frequency signals. In Jones PD, Bradley RS, and Jouzel J, editors, Climate variations and forcing mechanisms of the last 2000 years 9–41. Springer, Berlin.

Briffa KR, Jones PD, Schweingruber FH, Osborn TJ (1998a) Influence of volcanic eruptions on Northern Hemisphere summer teperatures over the past 600 years. Nature 393:450–454.

Briffa KR, Schweingruber FH, Jones PD, Osborn TJ, Shiyatov SG, Vaganov EA (1998b) Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. Nature 391:678–682.

Briffa KR, Osborn TJ, Schweingruber FH, Harris IC, Jones PD (2000a) Extracting low-frequency temperature variations from a northern tree-ring density network. J Geophys Res In preparation.

Briffa KR, Osborn TJ, Schweingruber FH, Jones PD, Shiyatov SG, Vaganov EA (2000b) Tree-ring width and density chronologies around the northern hemisphere: Part 1, local and regional climate signals. Holocene In preparation.

Chatfield C (1984) The analysis of time series. Chapman and Hall, London, third edition.

Collins M, Tett SFB, Cooper C (2000) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics Accepted for publication.

Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G (1995) The 'segment-length curse' in long tree-ring chronology development for palaeoclimatic studies. Holocene 5:229–237.

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

Crowley TJ, Kim KY (1999) Modeling the temperature response to forced climate change over the last six centuries. Geophys Res Letts 26(13):1901–1904.

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

Delworth TL, Mann ME (1999) Observed and simulated multidecadal variability in the northern hemisphere. Climate Dynamics Submitted.

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

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

Jones PD, Briffa KR, Barnett TP, Tett SFB (1998) High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control-run temperatures. The Holocene 8(4):455–471.

Lean J, Beer J, Bradley R (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 R Met Soc 79:242–261.

Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392:779–787.

North GR (1984) Empirical orthogonal functions and normal modes. J Atmos Sci 41:879–887.

Oerlemans J, Reichert BK (1999) Relating glacier mass balance to meteorological data by using a seasonal sensitivity characteristic (SSC). J Glaciology In Press.

Parker DE, Legg TP, Folland CK (1992) A new daily Central England temperature series. Int J Climatol 12:317–342.

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

Reichert BK, Bengtsson L, Akesson O (1999) A statistical modeling approach for the simulation of local paleoclimatic proxy records using general circulation model output. J Geophys Res 104:19071–19083.

Sato M, Hansen JE, McCormick MP, Pollack JB (1993) Stratospheric aerosol optical depths (1850–1990). J Geophys Res 98(D12):22987–22994.

Stott PA, Tett SFB (1998) Scale-dependent detection of climate change. J Climate 11:3282– 3294.

Tett SFB, Stott PA, Allen MR, Ingram WJ, Mitchell JFB (1999) Causes of twentieth century temperature change. Nature 399:569–572.

Venzke S, Allen MR, Sutton RT, Rowell DP (1999) The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. J Climate 12(8):2562–2584.

	R	R^2 (%)	β	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

Table 1: Parameters of the calibration process of the standard tree-ring density regional chronologies (see text for the description of the algorithm). R is the correlation coefficient, R^2 is the explained variance, β is the regression coefficient and RMSE is the residual mean squared error (after Briffa et al (2000b)).

	R	R^2 (%)	β	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

Table 2: Parameters of the calibration process of the age-banded tree-ring density chronologies (see text for the description of the algorithm). R is the correlation coefficient, R^2 is the explained variance, β is the regression coefficient and RMSE is the residual mean squared error. There is no value of β for the NE reconstruction because multiple linear regression is used (after Briffa et al (2000b)).

REGION	standard tree-ring		Age-banded tree-ring		HadCM3	
	a_1	a_0	a_1	a_0	a_1	a_0
NEUR	0.17	0.51	0.31	0.64	0.10	0.75
SEUR	0.14	0.42	0.46	0.47	0.058	0.59
NSIB	0.1	0.61	0.4	0.76	0.097	0.63
ESIB	0.34	0.45	0.4	0.42	0.027	0.64
CAS	0.25	0.25	0.25	0.27	0.097	0.47
TIBP	0.083	0.076	0.034	0.12	0.025	0.38
WNA	0.07	0.41	0.24	0.45	0.013	0.65
NWNA	0.0065	0.36	0.057	0.39	0.11	0.56
ECCA	0.11	0.48	0.13	0.5	0.056	0.44
NH	0.33	0.18	0.54	0.18	0.18	0.2

Table 3: Fitted coefficients of an AR(1) process, $x_{n+1} = a_1 x_n + a_0 z$, where z is unit variance white noise for the tree-ring and HadCM3 control time series.

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

Table 4: Average fitted coefficients of an AR(1) process, $x_{n+1} = a_1 x_n + a_0 z$, where z is unit variance white noise for the HadCM3 natural ensemble.

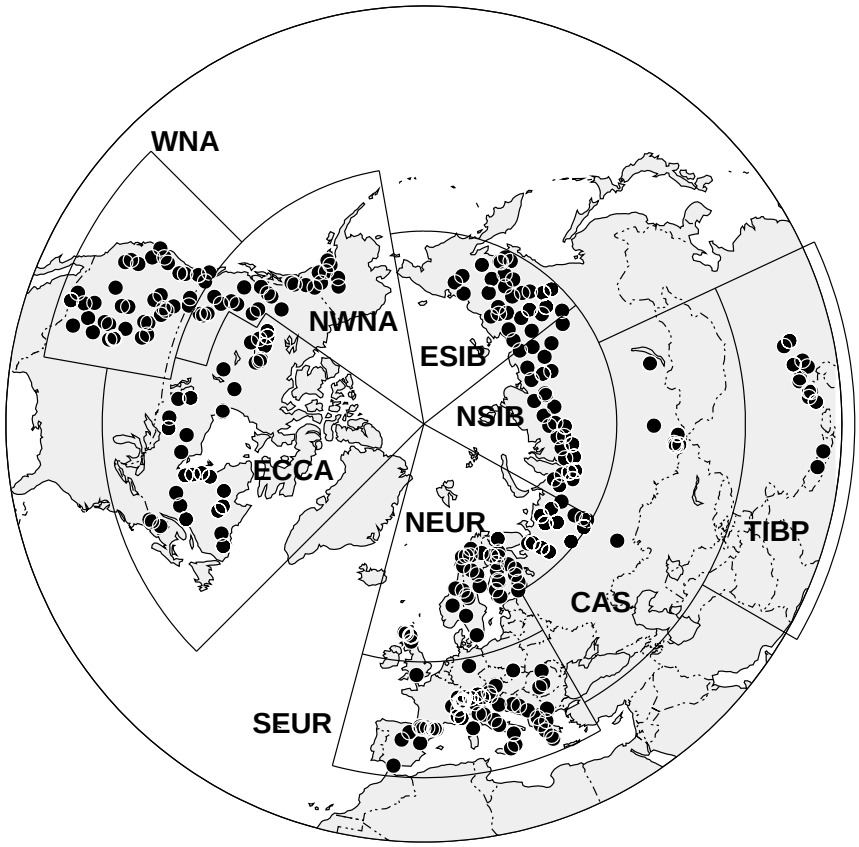

Figure 1: Locations of individual tree-ring density chronologies (dots) and the definition of the nine regional series used in the calibration against observed temperature (after Briffa et al (2000b)).

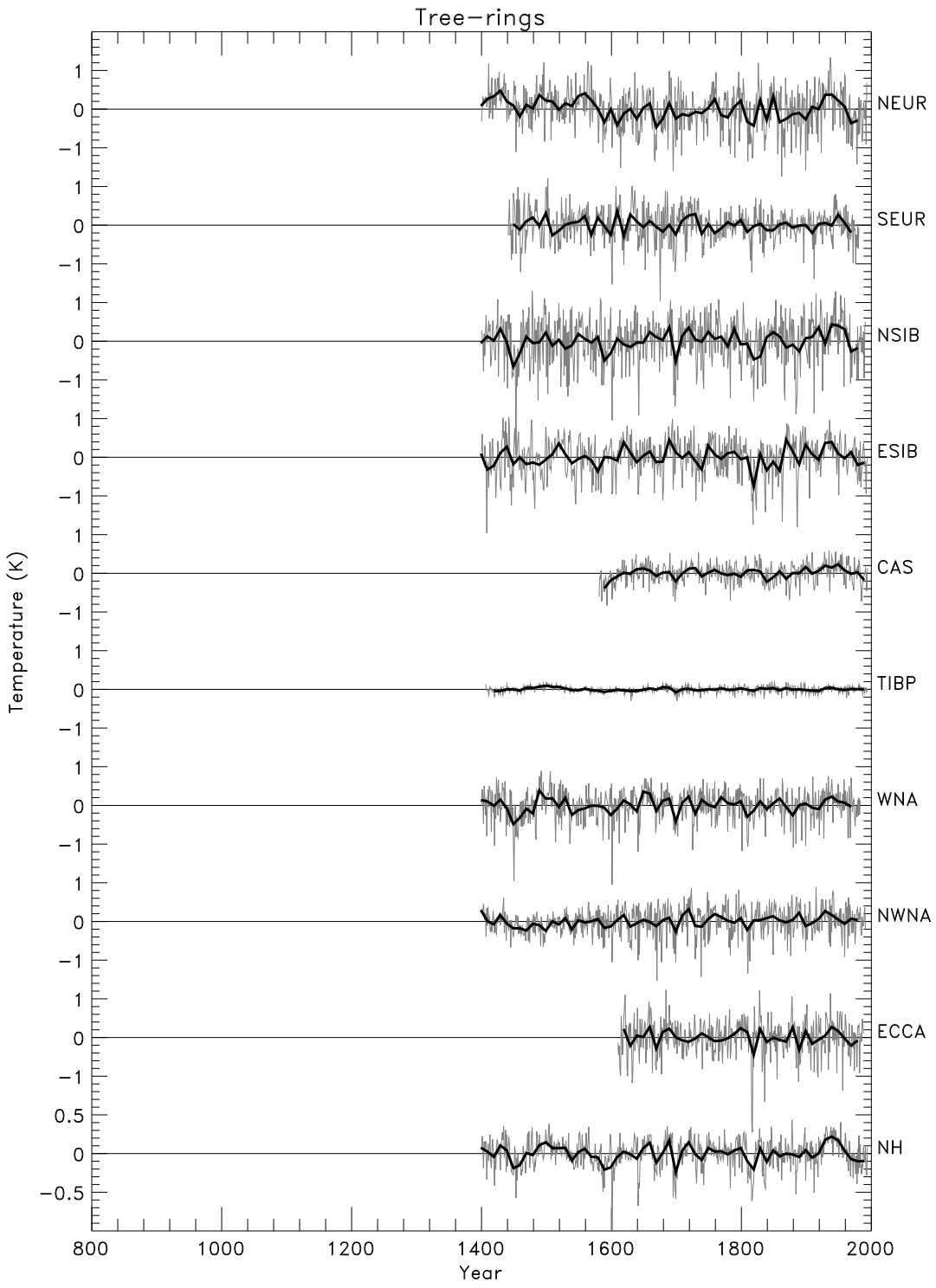

Figure 2: The calibrated standard tree-ring series for the nine regions and for NH (see fig. 1 for the definitions of the regions). The grey lines are the annually resolved data and the black lines are decadal averages formed from these. Note the change in vertical scale for the NH series.

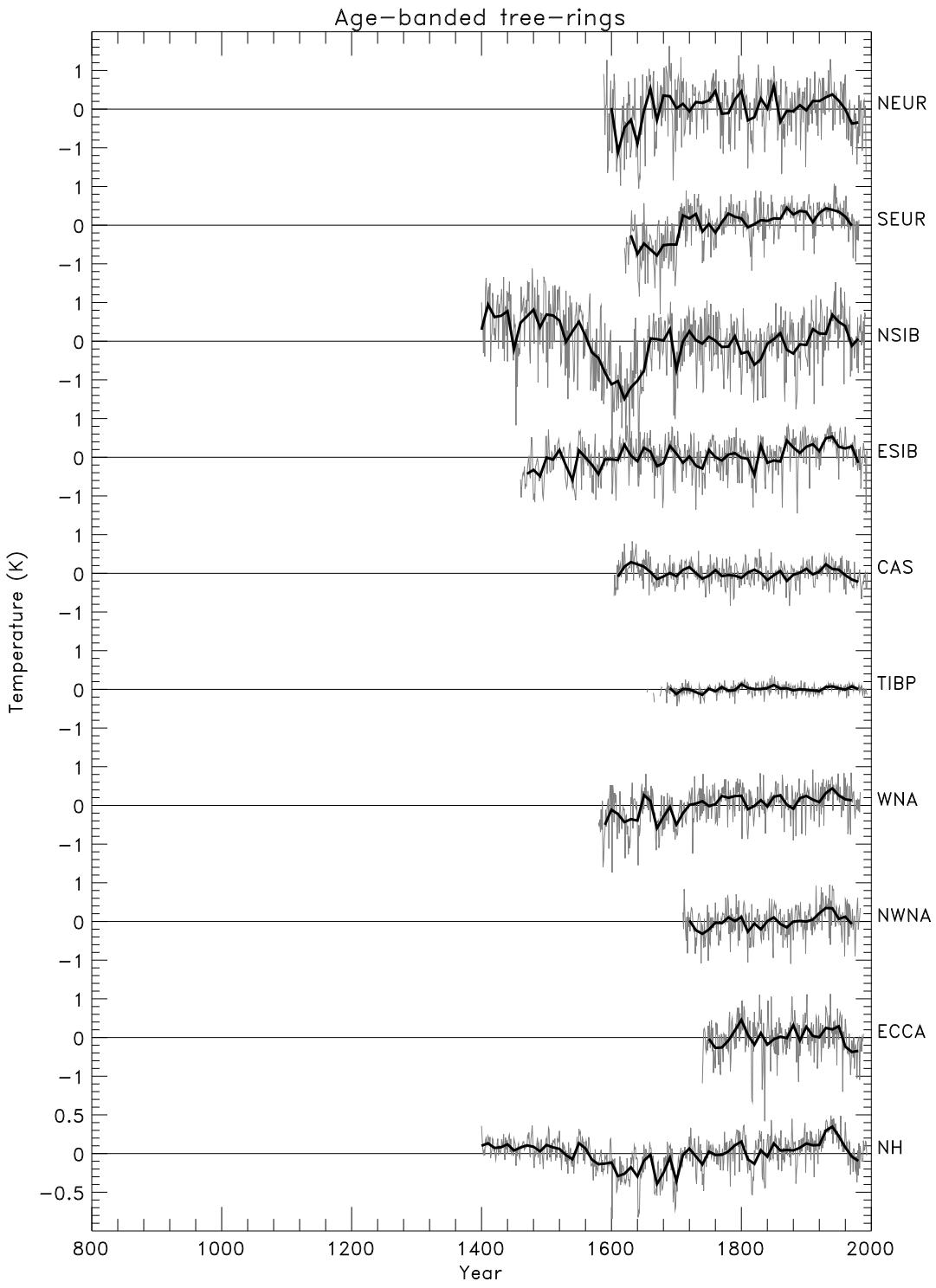

Figure 3: The calibrated age-banded tree-ring series for the nine regions and for NH (see fig. 1 for the definitions of the regions). The grey lines are the annually resolved data and the black lines are decadal averages formed from these. Note the change in vertical scale for the NH series.

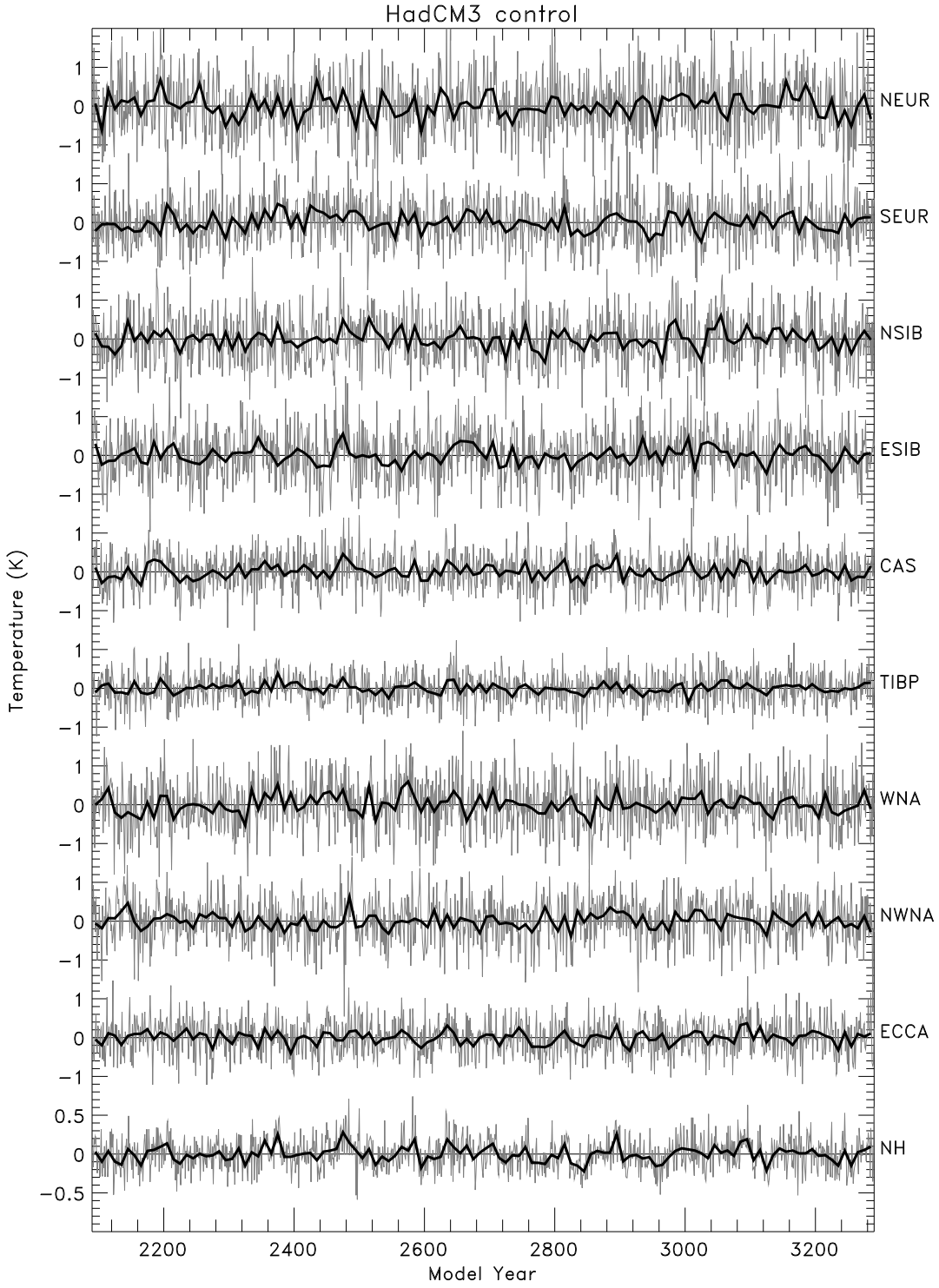

Figure 4: HadCM3 control simulation April-September land temperatures for the nine regions and for NH (see fig. 1 for the definitions of the regions). The grey lines are the annual (i.e. April-September) data and the black lines are decadal averages formed from these. Note the change in vertical scale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

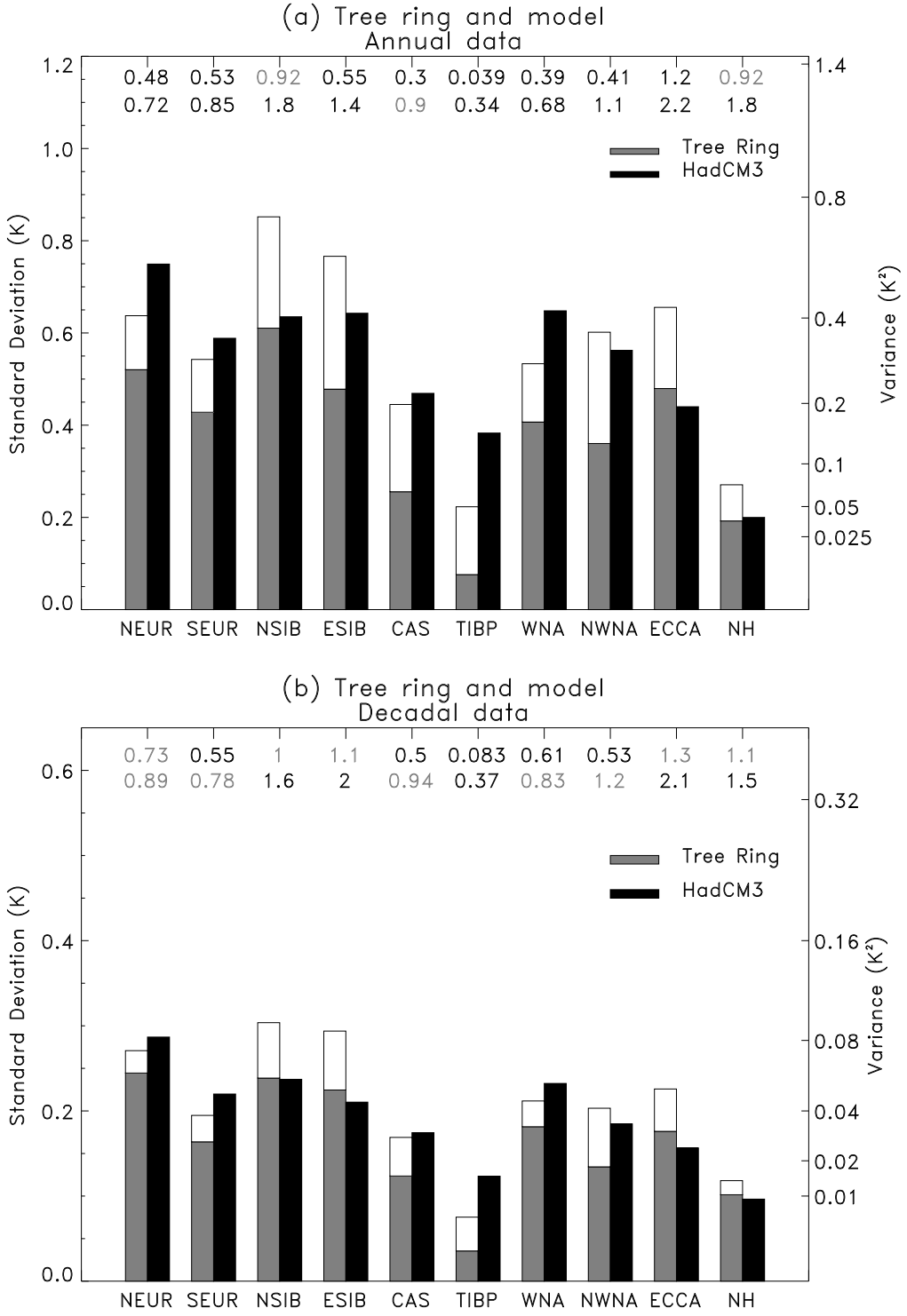

Figure 5: Standard deviations/variances of the standard tree-ring series and the HadCM3 control series for (a) annual data and (b) decadal data. The tree-ring standard deviations are the grey bars with the residual variance from the calibration period placed on top of these in white. Model standard deviations are the black bars. The ratios of the variances are shown buy the numbers above the bars. The upper numbers are the ratios without the residual variance taken into account at 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 which are not significant are shown in grey.

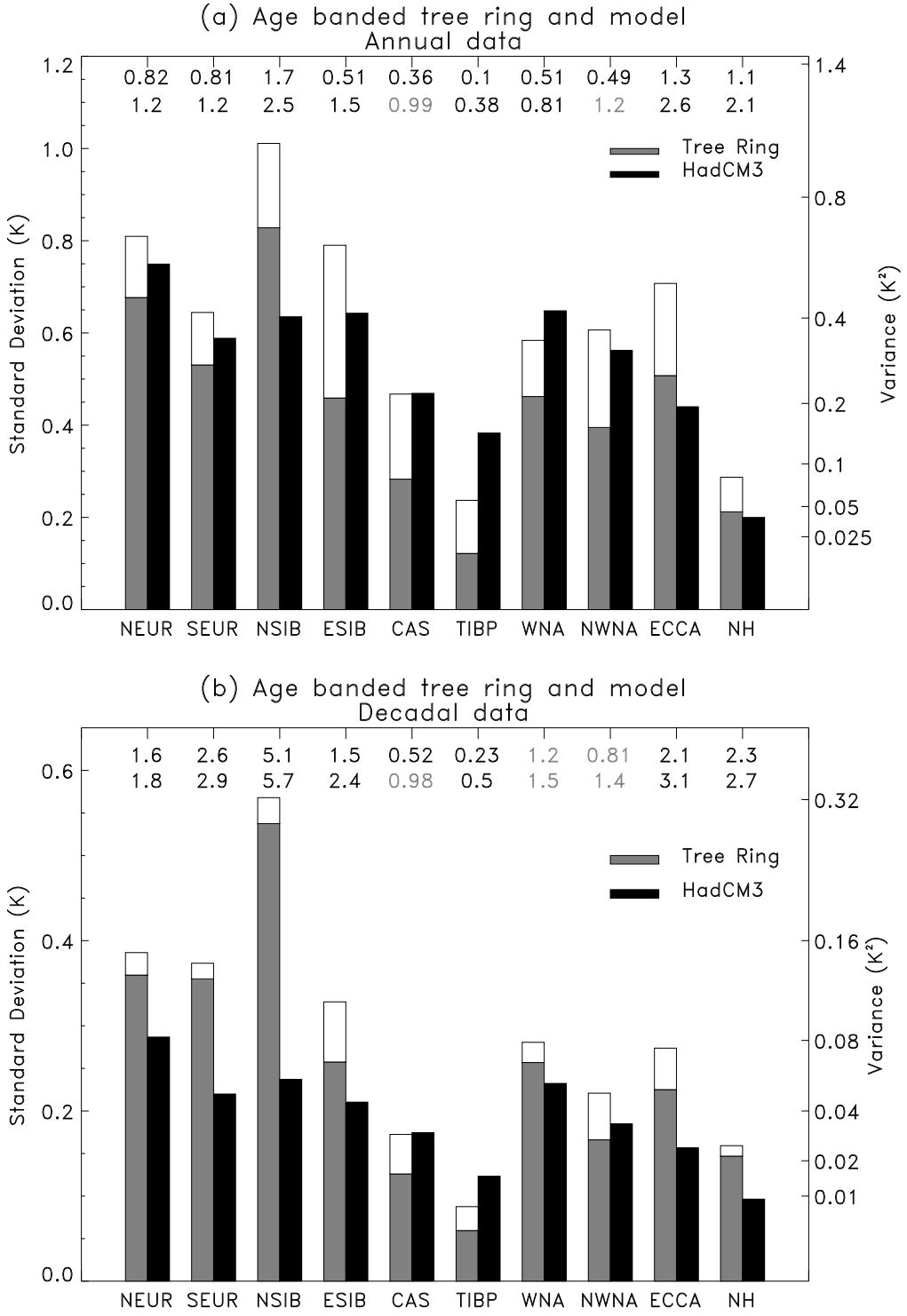

Figure 6: Standard deviations/variances of the age-banded tree-ring series and the HadCM3 control series for (a) annual data and (b) decadal data. The tree-ring standard deviations are the grey bars with the residual variance from the calibration period placed on top of these in white. Model standard deviations are the black bars. The ratios of the variances are shown buy the numbers above the bars. The upper numbers are the ratios without the residual variance taken into account at 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 which are not significant are shown in grey.

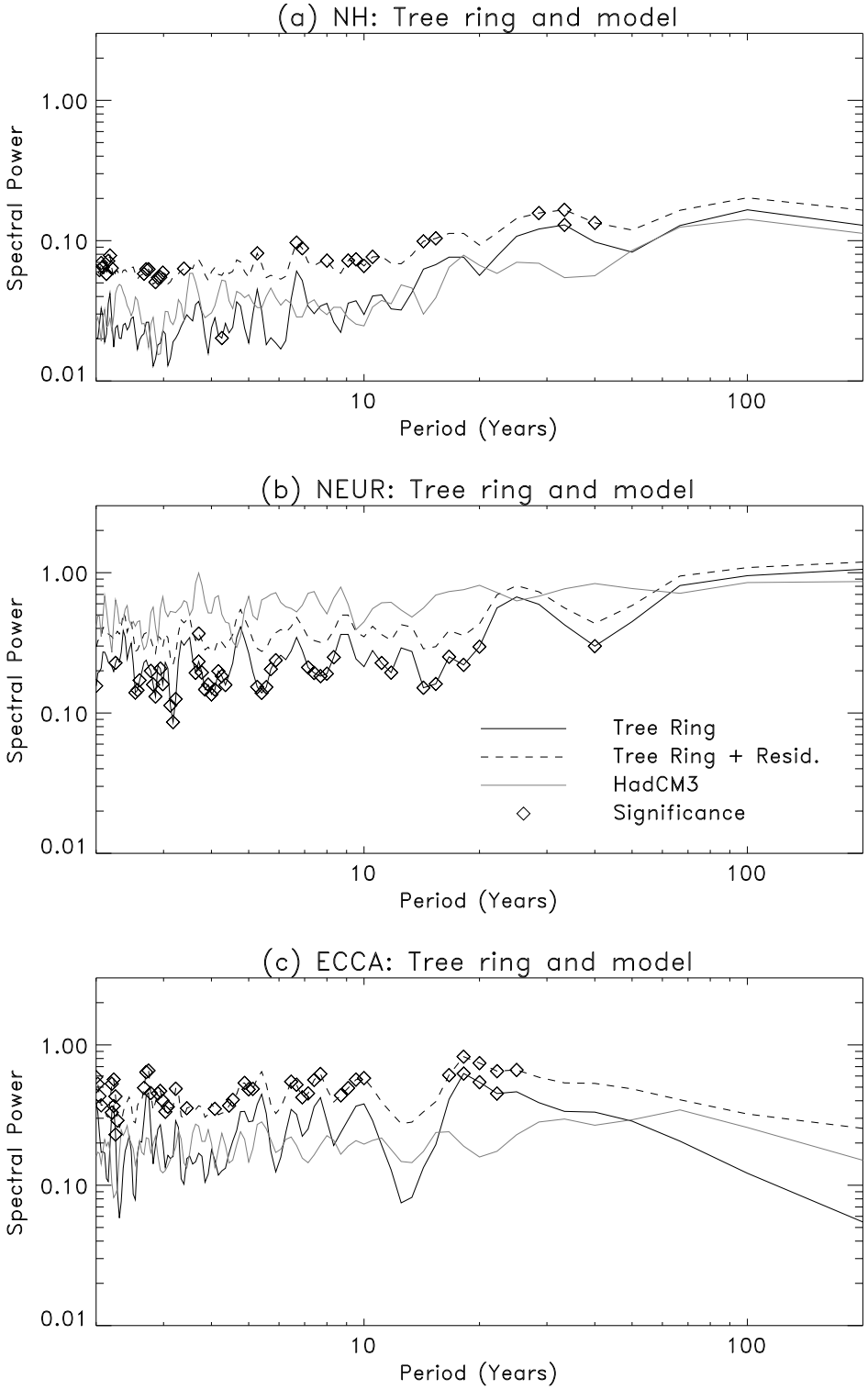

Figure 7: Power spectra of the model and standard tree-ring series for (a) NH, (b) NEUR and (c) ECCA. In each graph the model spectra is shown in grey, the calibrated tree-ring spectra is shown as the black solid line and the tree-ring spectra with the residual variance added is shown as the black dashed 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.

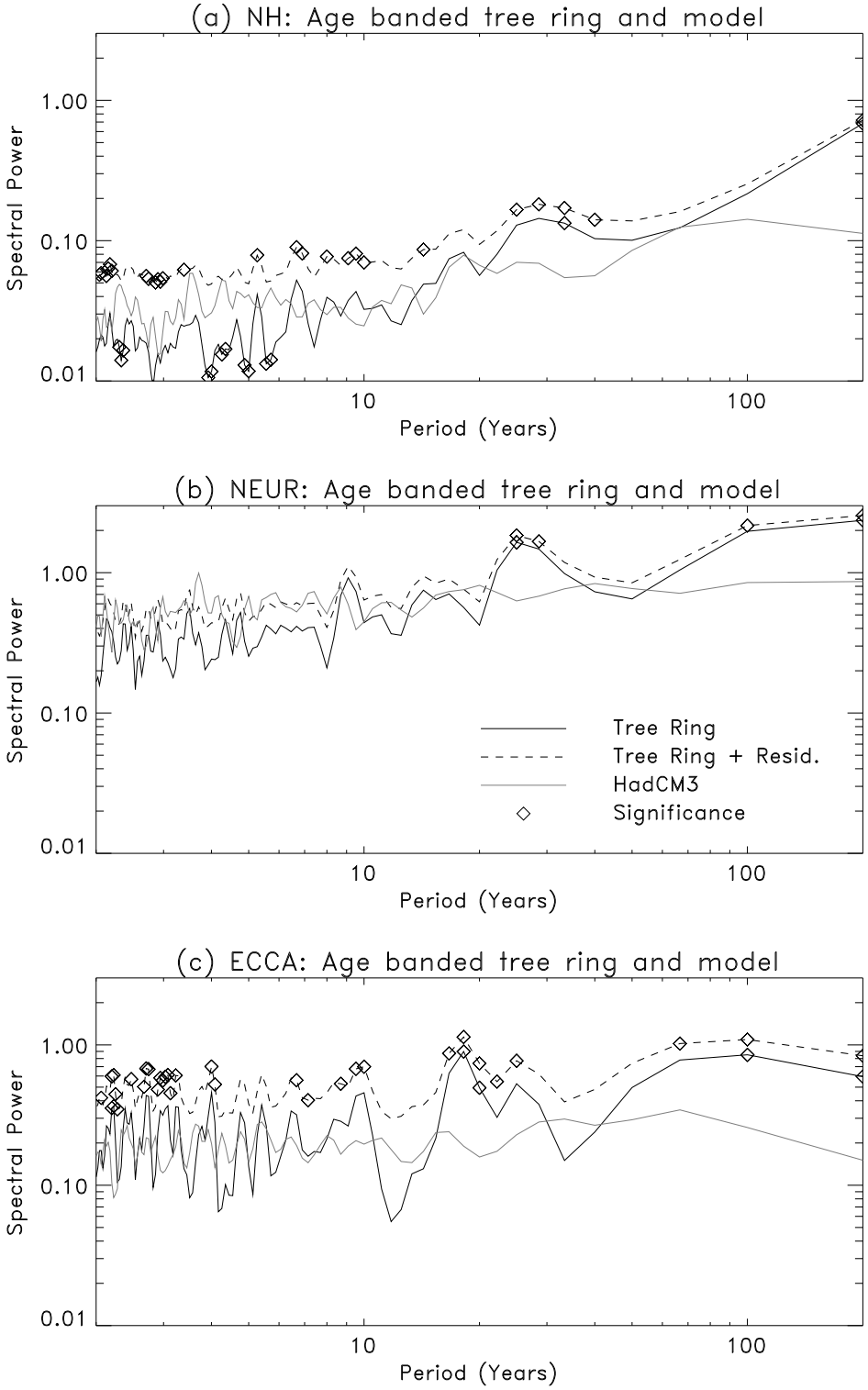

Figure 8: Power spectra of the model and age-banded tree-ring series for (a) NH, (b) NEUR and (c) ECCA. In each graph the model spectra is shown in grey, the calibrated tree-ring spectra is shown as the black solid line and the tree-ring spectra with the residual variance added is shown as the black 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.

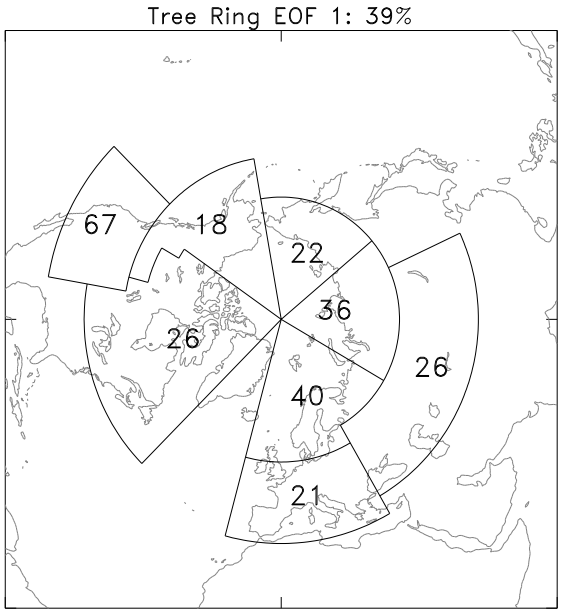

Figure 9: Leading EOF of the decadal averaged standard tree-ring series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

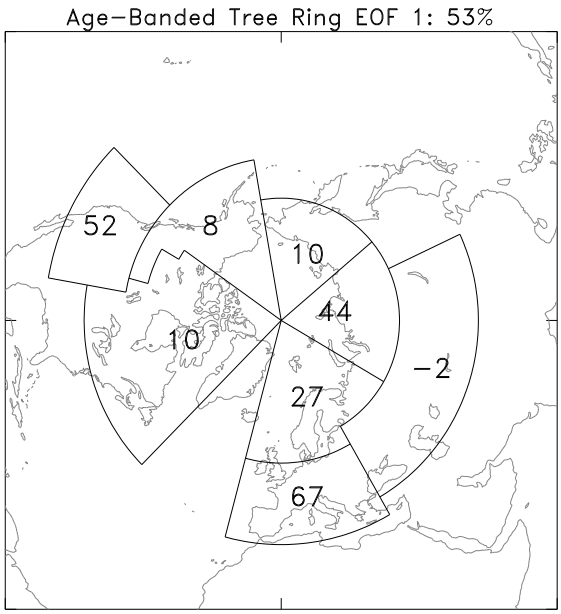

Figure 10: Leading EOF of the decadal averaged age-banded tree-ring series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

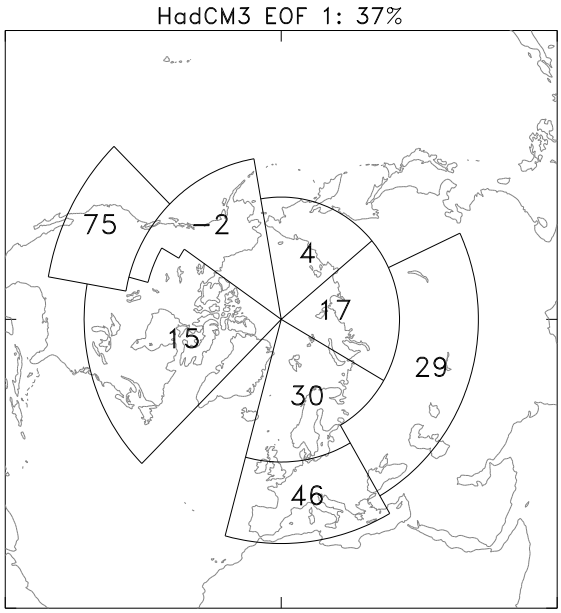

Figure 11: Leading EOF of the decadal averaged HadCM3 series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

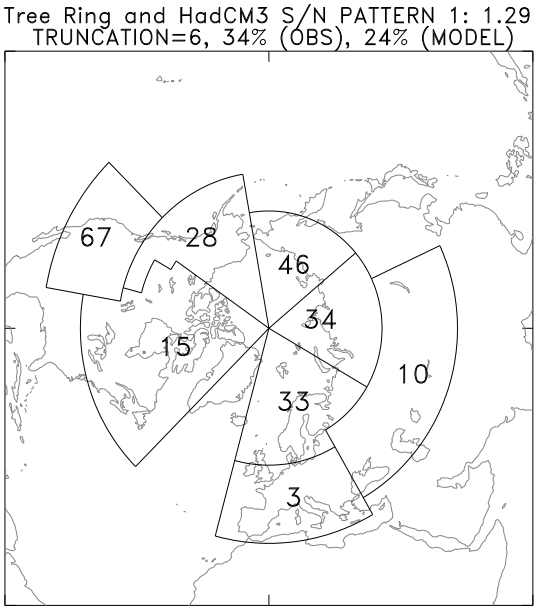

Figure 12: Leading signal-to-noise maximising pattern of the standard tree-ring vs. model series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

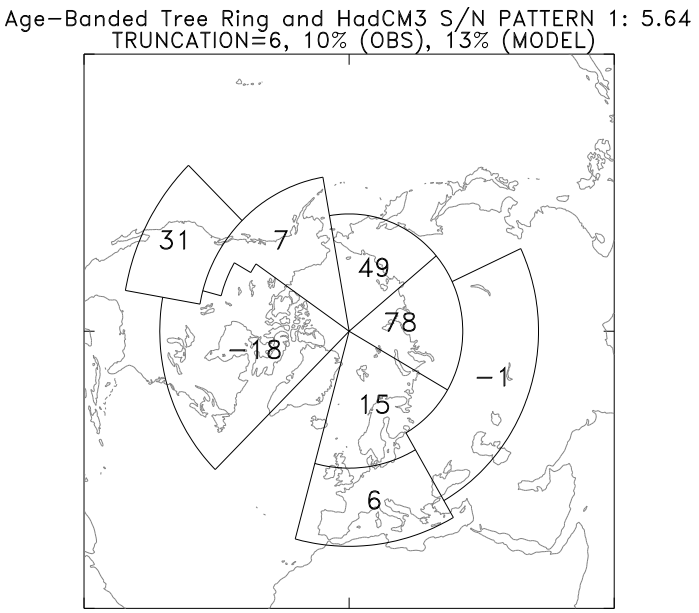

Figure 13: Leading signal-to-noise maximising pattern of the age-banded tree-ring vs. model series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

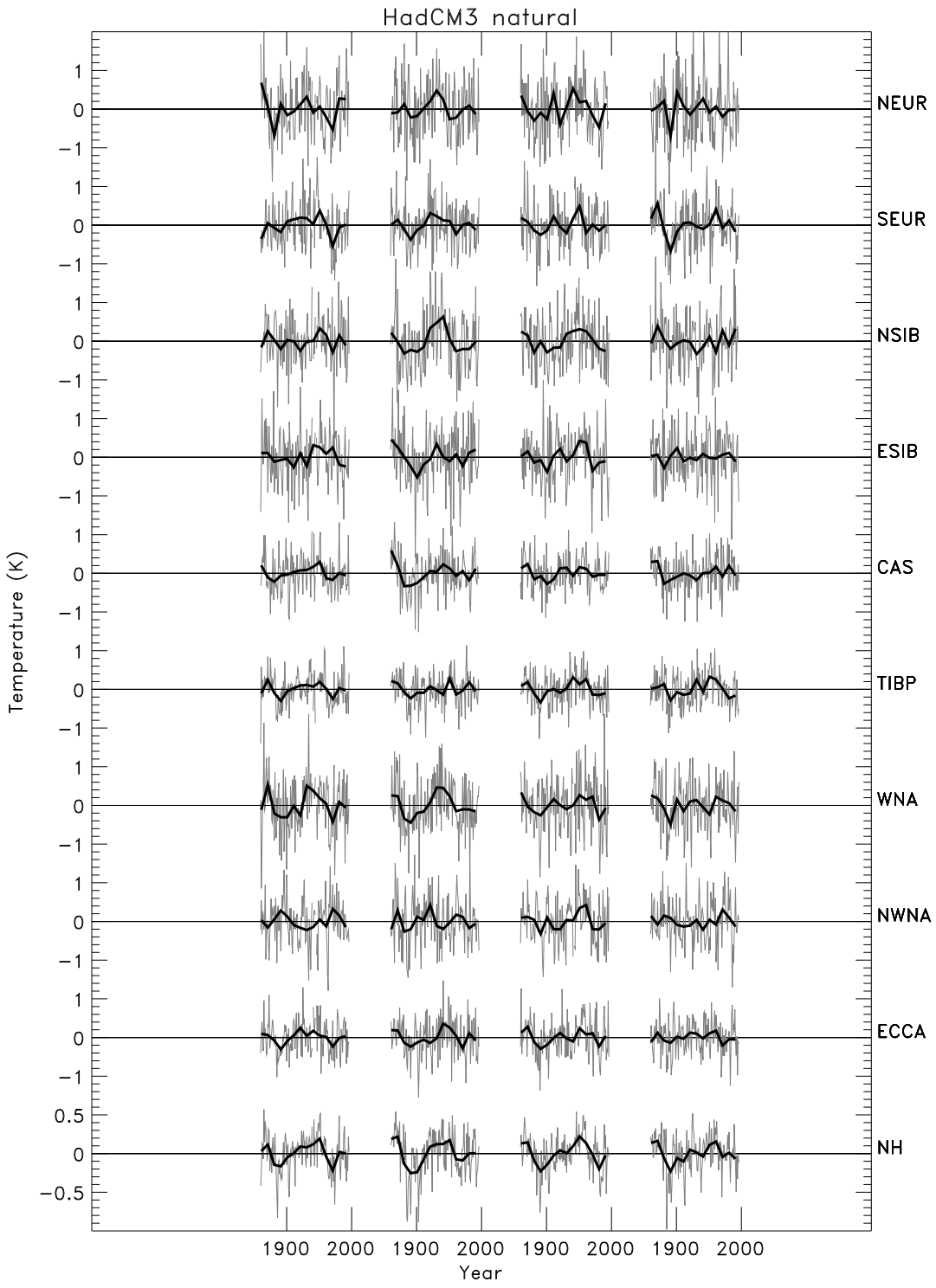

Figure 14: Regional and NH series from an ensemble of 4 HadCM3 experiments with solar and volcanic forcing. The grey lines are the annual data and the black lines are decadal averages formed from these. Note the change in scale for the NH series.

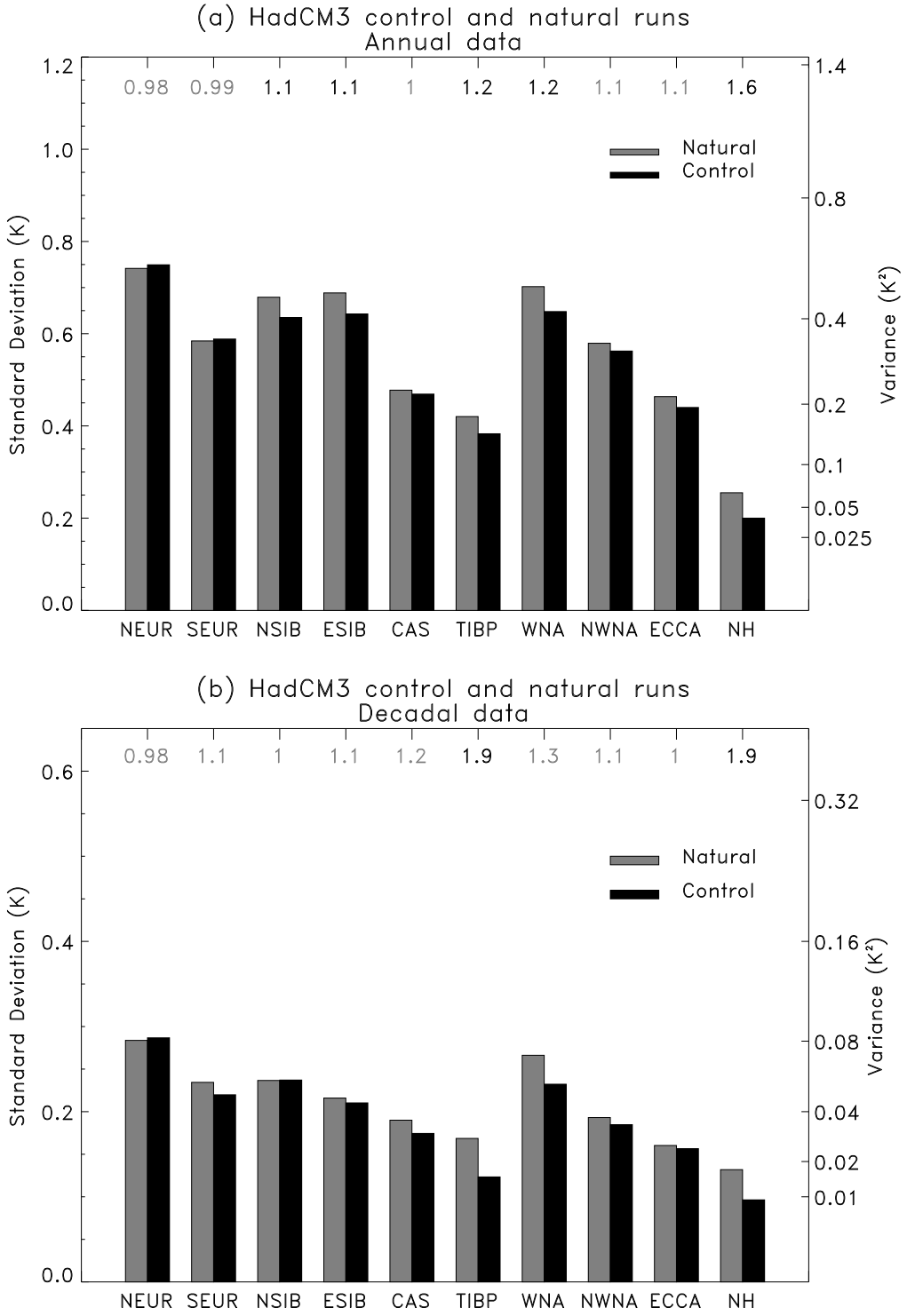

Figure 15: Standard deviations of the HadCM3 natural ensemble and the HadCM3 control series for (a) annual data and (b) decadal data. The natural ensemble standard deviations are the grey bars and the control standard deviations are the black bars. The ratios of the variances are shown buy the numbers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

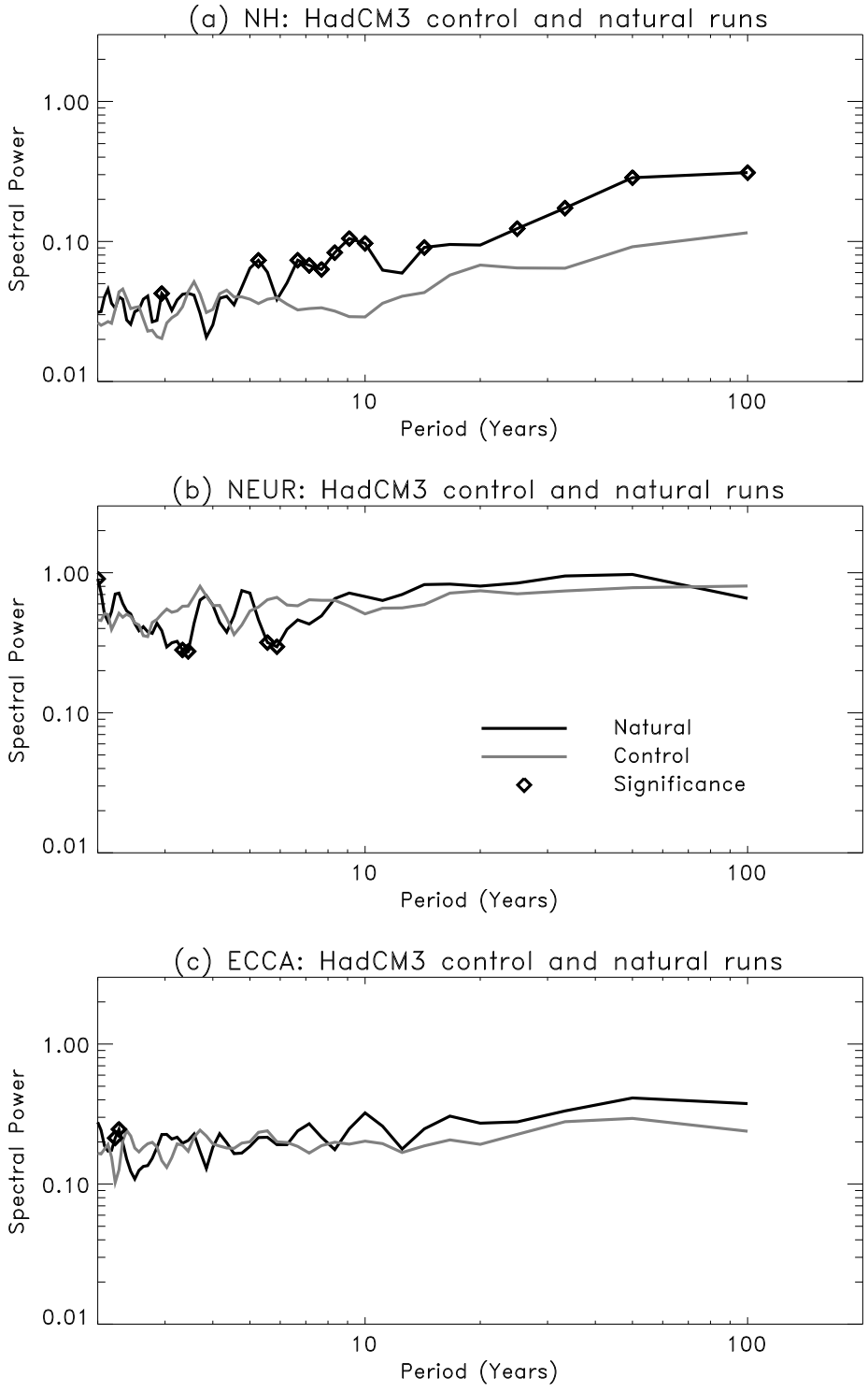

Figure 16: Power spectra of the natural ensemble and control for (a) NH, (b) NEUR and (c) ECCA. In each graph the control spectra is shown in grey and the natural spectra in black. Where the control and natural spectra are different at the 95% level according to an F-test, the center of the frequency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

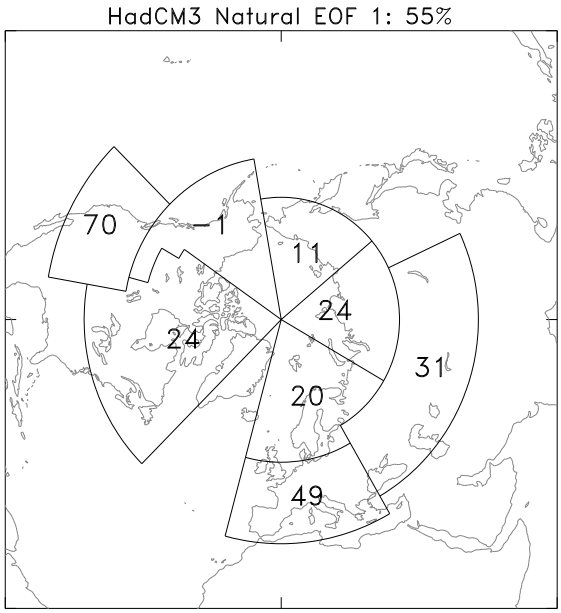

Figure 17: Leading EOF of the decadal averaged HadCM3 natural ensemble. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.

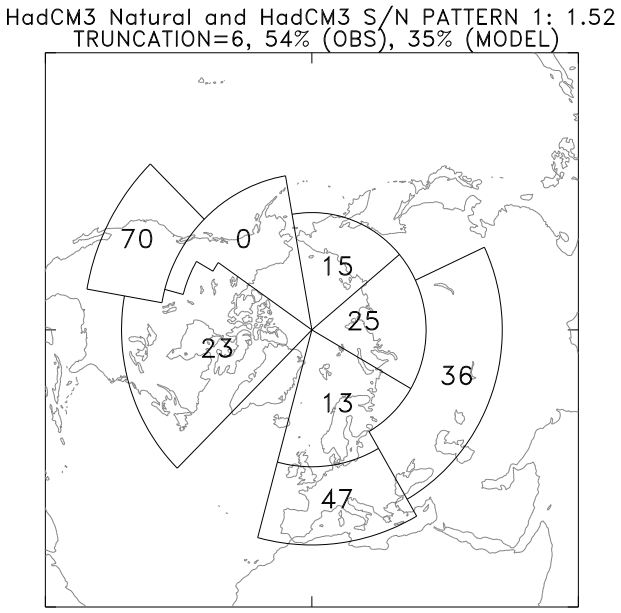

Figure 18: Leading signal-to-noise maximising pattern of the model natural ensemble vs. model control series. The figures correspond to the EOF loadings multiplied by 100 for the regions indicated.