Testing the Mann et al. (1998) Approach to Paleoclimate Reconstructions in the Context of a 1000-Yr Control Simulation with the ECHO-G Coupled Climate Model

EDUARDO ZORITA AND FIDEL GONZÁLEZ-ROUCO

Institute for Coastal Research, GKSS-Research Center, Geesthacht, Germany

STEPHANIE LEGUTKE

Max-Planck-Institut für Meteorologie, Hamburg, Germany

(Manuscript received 20 August 2001, in final form 12 August 2002)

ABSTRACT

Statistical reconstructions of past climate variability based on climate indicators face several uncertainties: for instance, to what extent is the network of available proxy indicators dense enough for a meaningful estimation of past global temperatures?; can statistical models, calibrated with data at interannual timescales be used to estimate the low-frequency variability of the past climate?; and what is the influence of the limited spatial coverage of the instrumental records used to calibrate the statistical models? Possible answers to these questions are searched by applying the statistical method of Mann et al. to a long control climate simulation as a climate surrogate. The role of the proxy indicators is played by the temperature simulated by the model at selected grid points.

It is found that generally a set of a few tens of climate indicators is enough to provide a meaningful estimation (resolved variance of about 30%) of the simulated global annual temperature at annual timescales. The reconstructions based on around 10 indicators are barely able to resolve 10% of the temperature variance. The skill of the regression model increases at lower frequencies, so that at timescales longer than 20 yr the explained variance may reach 65%. However, the reconstructions tend to underestimate some periods of global cooling that are associated with temperatures anomalies off the Antarctic coast and south of Greenland lasting for about 20 yr. Also, it is found that in one 100-yr period, the low-frequency behavior of the global temperature evolution is not well reproduced, the error being probably related to tropical dynamics.

This analysis could be influenced by the lack of a realistic variability of external forcing in the simulation and also by the quality of simulated key variability modes, such as ENSO. Both factors can affect the large-scale coherence of the temperature field and, therefore, the skill of the statistical models.

1. Introduction

The observed increase of global mean air temperatures in the last 100 yr has stimulated a great deal of efforts in the reconstruction of climate conditions in the past centuries, with the objective of putting the observed global warming into the perspective of the natural climate variability. These climate reconstructions may be based on simulations with climate models of different degrees of complexity (Cubasch et al. 1997; Crowley 2000), or on the analysis of climate proxy data, for instance, tree-ring chronologies (Cook et al. 1998; Jones et al. 1998; Briffa et al. 1998, 2001), ice-core snow deposition rates (Appenzeller et al. 1998), documentary evidence (Pfister et al. 1999), or a combination of multiple proxy types (Mann et al. 1998, hereafter MBH98).

Two outstanding examples so far are the estimation of past temperature (MBH98) and sea level pressure evolutions (Luterbacher et al. 1999) by the use of statistical models linking a variety of climatic proxy data and the climatic variable of interest. These statistical models are set up and validated in the instrumental period where both types of datasets overlap. The estimations of the evolution of the climate parameters are then performed by feeding these statistical models with the much longer time series of proxy data available. In particular, the global annual temperature reconstructions of MBH98 are based essentially on a network of tree-ring chronologies, corals, ice cores from Greenland and other glaciers, and very long instrumental time series. These authors employ a regression model that links the time series of the leading principal components of the global annual temperature field with the time series of the proxy data. The statistical model is validated using independent data from the instrumental period reserved for this purpose.

Although the methodology for this reconstruction is probably the best strategy that can be used in view of

Corresponding author address: Eduardo Zorita, GKSS-Forschungszentrum, Max-Planck-Str.1, 21502 Geesthacht, Germany.
E-mail: zorita@gkss.de

the data available, several uncertainties may remain, which are not easily resolved with the sole use of observational or proxy datasets, and that may require the analysis of long simulations with coupled climate models. One of these uncertainties is related to the mismatch between the timescales of the model setup and the timescale of the reconstructions. The statistical regression model can only be calibrated and validated in the instrumental period, which optimistically can be considered as spanning about the last 150 yr. However, the climate processes operating on longer timescales may not necessarily be statistically captured by the the values of the model parameters, which are determined by the high-frequency (interannual) climate variability in the instrumental period. The question then arises as to whether the regression equations are still valid at these longer (decadal and centennial) timescales.

A second uncertainty relates to the amount of climate proxy data available. In most of the reconstructions based on statistical methods the densities of proxy data available decreases backward in time. For instance, in MBH98, the number of proxy records available for the temperature reconstruction decreases from 112 in the period 1820 A.D. onward, down to 22 in the period 1400–50 A.D. In the North Atlantic sea level pressure reconstruction by Luterbacher et al. (1999), the number of proxies delivering information from 1650 A.D. is 20% of the number of proxies available for 1900 A.D.

Two questions can be posed in this context. One involves the minimum number of climate proxy records required for a meaningful reconstruction of the global temperature field. The other concerns the dependence of temperature reconstructions on the subset of proxy indicators that is still available for the longest reconstructions. The area covered by the complete set of indicators is mostly situated in inhabited areas, lower and midlatitudes, the obvious exception to this rule being ice-core records. The available remaining subsets that provide information for more remote periods are probably a random group of the original set of proxies, their locations being determined by reasons of opportunity (research funds, archeological campaigns, etc.). One could reasonably argue that, when the size of this subset decreases substantially, the reconstructed temperature may depend too strongly on the (quasi random) composition of this proxy subset.

In this paper we try to give an estimation of the uncertainties associated with these questions. Our objective is not to test the temperature reconstructions with a simulation using a global climate model driven by the estimated past variations of external forcing, such as solar variability, volcanic aerosols, etc. Our aims are rather directed to methodological aspects of the reconstructions: to test whether the statistical methodology for the estimation of global annual temperature remains valid at long timescales in a complex system, such as a coupled global climate model; to obtain an estimation of the minimum number of proxies necessary for a reasonable temperature reconstruction; and finally to investigate if the errors in the temperature reconstructions are equally distributed over the globe or if regions exist where the temperature variations cannot be captured by the network of proxy data.

For our purpose, a 1000-yr long global climate simulation with a state-of-the-art climate model will be analyzed. The output of this simulation will play the role of a pseudoclimate, the temperature of which will be reconstructed by a statistical regression model and a set of pseudoproxy indicators. Contrary to the real world, in a climate simulation no proxy data exist and, therefore, in this study the simulated temperature at selected grid points will play the role of pseudoclimate indicators.

In this approach several factors will contribute to a higher skill of the statistical reconstructions in the pseudoworld of the climate model, compared to the skill of the statistical model in the real climate. One factor is the relationship between the climate proxy data (e.g., tree-ring chronologies) and the local temperature at the position of the proxy time series. These relationships are known as the transfer functions, which in most cases are derived empirically. Being approximate, they describe only part of the local temperature variability. Our approach completely neglects this source of uncertainty and therefore this choice could be denominated as ''perfect-signal pseudoproxies,'' as opposed to ''noisy pseudoproxies,'' where the gridpoint temperatures would be artificially contaminated with statistical noise (Mann and Rutherford 2002). A second difference to the real world is the lack of external sources of climate variability in the climate simulation, such as changes in solar output, which may affect the validity of the statistical model at long timescales. This aspect will be taken into account in a twin climate simulation with external forcing currently under way. A third factor is the number of physical degrees of freedom present in a climate model (Jones et al. 1998). Because of the known limitation of computing power this number is clearly just a small fraction of the number of degrees of freedom of the real climate and, therefore, a statistical model linking climate variables within the climate simulation will probably tend to perform better than its counterpart in the real world.

Recently, the question of the characteristic spatial scales of the information contained in the real proxy indicators has been discussed in the literature (Barnett et al. 1999). These authors argue that the coherence between local temperature conditions and the proxy records is often poor, that is, the local signal-to-noise ratio is low. Even if this is the case, Bradley et al. (2000) argue that the information provided by local proxy indicators may still contain large-scale information, for instance, if an indicator is sensitive to salinity changes related to large-scale phenomena such as the Southern Oscillation. To test these additional aspects in a climate model is not a straightforward task. Many nontemper-

ature indicators used so far are sensitive to precipitation and it is known that unfortunately the skill of climate models in reproducing even the mean annual cycle of precipitation is quite poor. Also, the relationship between modeled precipitation and modeled large-scale phenomena is often not the same as in the real climate. Though certainly worthwhile in itself, the estimation of this source of uncertainty is beyond the scope of the present study. Therefore, it should be kept in mind that the performance of the reconstructions in the surrogate climate could become better if this effect could be reasonably incorporated in the analysis.

For these reasons the present study can be viewed as a test for the methodology of statistical climate reconstruction in clearly favorable conditions. The obvious advantage of analyzing a climate simulation with this purpose is that the state of the climate is completely known at all times, and the errors in the temperature reconstructed by the statistical model can be easily compared with the true (model simulated) temperature, and the source of errors can be much more easily investigated. This line of research follows the previous works of, among others, Madden et al. (1993) and Bradley (1994). Madden et al. (1993) focused on the error in the estimation of the global mean temperature stemming from limited spatial sampling and used a perpetual January simulation with the National Center for Atmospheric Research (NCAR) community atmospheric model for their analysis. Bradley (1994) tried to identify optimal locations for climate indicators for paleotemperature reconstructions. He also analyzed results from a long climate simulation with the Geophysical Fluid Dynamics Laboratory (GFDL) coupled climate model.

The paper is structured as follows: a short description of the model and the climate simulation is given in section 2. Section 3 is a brief description of the statistical regression model and the data used to set up the model. In section 4, the influence of the limited spatial coverage of the observational dataset used in the calibration of the statistical models is investigated. Section 5 focuses on the performance of the model as the number of proxy indicators is varied and on the spatial differences in the temperature reconstructions. In section 6, we perform the temperature reconstructions by calibrating the statistical model in a different, somewhat special, period of the simulation. The spectral characteristics of the temperature reconstructions are considered in section 7. The paper closes with some concluding remarks.

2. Model simulation

The global climate model ECHO-G (Legutke and Voss 1999) consists of the spectral atmospheric model ECHAM4 and the ocean model HOPE-G, both developed at the Max-Planck-Institut für Meteorology in Hamburg, Germany. The model ECHAM4 has, in this simulation, a horizontal resolution of T30 (approximately $3.75° \times 3.75°$) and 19 vertical levels. The horizontal resolution of the ocean model HOPE-G is about $2.8° \times 2.8°$ with a grid refinement in the tropical regions, where the meridional gridpoint separation decreases progressively to the equator, reaching a value of 0.5°. The ocean model has 20 vertical levels.

To avoid climate drift in such a long simulation, additional fluxes of heat and freshwater, diagnosed from SST and sea surface salinity (SSS) restoring boundary conditions in a coupled spinup integration are applied to the ocean. This flux adjustment is constant in time and its global integral vanishes.

The coupled model is driven by constant external forcing (solar constant and atmospheric concentrations of anthropogenic greenhouse gases), so that the climate variability is only due to the internal dynamics of the model. From this integration only the 2-m temperature field has been used in this study. The globally averaged annual temperature shows a stable behavior with a range of variability of about 0.15 K. Regionally, the temperature variability can be larger and, for instance in the European region, periods several decades long exist, where winter temperature deviations reach 0.5 K, comparable to the ones estimated for the Little Ice Age (Pfister et al. 1999).

3. Statistical method for temperature reconstruction

In this study we have used the same method as described in MBH98 for temperature reconstructions since 1400 A.D. This method is based on the calibration of regression equations between climate proxy time series and the principal components (PCs) of the annual temperature field over the globe. Once the statistical regression model has been calibrated using data from the instrumental period, it can be used to estimate the amplitudes of the temperature PCs in the period with proxy data available. These amplitudes can then be used to reconstruct the global temperature field.

Since the number of locations where proxy data are available is limited, one can only expect to reconstruct the large-scale structures of temperature variability. In earlier periods, when only a few proxy data can be used, only the leading PC is expected to be reasonably well estimated (Mann et al. 1999).

Since the climate model obviously does not simulate climate proxy records, the air temperatures simulated at the grid points situated nearest to the locations of proxy records are taken as ''simulated temperature proxies.'' This approach neglects all the sources of errors inherent in the statistical transfer functions between the proxy record and the local climate conditions. These errors can be quite large and strongly dependent on the location and the nature of the proxy. The focus of this study lies, however, in testing if a reasonable number of temperature-sensitive proxy records is, in principle, sufficient for realistic reconstructions of the global temperature. At this particular point, this approach is at odds with

the data used by MBH98. These authors also use other types of proxy records that are not sensitive to annual mean temperature variations, but to precipitation or seasonal temperatures. To mimic such type of proxies for the estimation of the global mean temperature in a sensible manner is, as stated in the introduction, not straightforward and would probably complicate the interpretation of the results of this analysis.

The model is calibrated in the reference period between the simulation years 920 and 1000. First, an empirical orthogonal function (EOF) analysis (von Storch and Zwiers 1999) of the standardized annual temperature anomalies is carried out:

$$T_{\rm sim}(x_i, t) = \sum_{j=1}^{N} \mathbf{f}_j(x_i)\boldsymbol{\alpha}_j(t), \qquad (1)$$

where $T_{\rm sim}(x_i, t)$ is the standardized annual temperature anomaly simulated at grid point x_i in the calibration period, N is the number of grid points, $\mathbf{f}_j$ is the jth EOF, and α_j is the corresponding PC. The spatial patterns $\mathbf{f}_j$ are the eigenvectors of the weighted correlation matrix of the temperature field, where the weights represent the latitudinally varying surfaces of the grid boxes. Since the patterns $\mathbf{f}_j$ are eigenvectors, their normalization is arbitrary. Here, the normalization is chosen so that the standard deviation of the principal components α_j in the calibration period is unity.

In a second step, a multiple linear regression equation linking the leading PCs (as predictors) and the standardized anomalies of local ''proxy'' data (as predictands) is set up:

$$p_k(t) = \sum_{j=1}^{n} a_k^j \alpha_k(t) + \epsilon_k(t), \qquad (2)$$

where $p_k(t)$ is the standardized anomaly of the kth proxy, a_k^j are the regression coefficients, and ϵ are the residuals. Since the time mean of the proxies and of the PCs in the calibration period is zero, no independent term is needed in the regression model. The number of proxies in this study ranges between 98 and 15, whereas the number n of retained EOFs is of the order of 10 or less, so that the regression model defined in Eq. (2) is always well posed. The value of the coefficients a_k^j can be estimated by minimizing the total variance of ϵ_k. The number of PCs included in (2) depends on the number of pseudoproxies used, as explained in the following section.

Once the coefficients a_k^j have been estimated in the calibration period, the model can be reversed to estimate temperature deviations when the value of the proxy data (in our case the temperature at the proxy grid points) is known. As in MBH98, the values of the PCs in Eq. (2) $\alpha(t)$ are estimated for each year by minimizing the amplitude of the error term $\epsilon_k(t)$. In a second step, the standardized temperature field can be reconstructed by setting in the estimated values of $\alpha(t)$ in (1). Finally, the physical temperature field is obtained by multiplying at each grid point the reconstructed standardized field by the gridpoint standard deviation estimated in the period of calibration. It should be noted that this statistical method is somewhat different from the conventional regressional approach, in which the value of the PCs would be determined by reversing the model in Eq. (2), so that the influence of, say, proxy k onto PC j is directly given by the regression coefficient α_k^j. Instead, MBH98's method yields an estimation of the value of the temperature PCs that is optimal for the set of climate indicators as a whole, so that the estimations of individual PCs cannot be traced back to a particular subset of indicators or to an individual climate indicator. This reconstruction method offers the advantage that possible errors in particular indicators are not critical, since the signal is extracted from all the indicators simultaneously.

It is noted that in MBH98 the estimation of the PCs is achieved somewhat differently from the method we use. These authors use the standardized monthly temperature anomalies to estimate the correlation matrix and subsequently the PCs are annually averaged. The reason for this procedure is to increase the number of independent samples, that leads to a better estimation of the EOFs. This approach has also been tested with the model results, but the EOF patterns thus obtained were closely related to the EOFs of the annual temperature. As will be mentioned in following sections, the spatial structure of leading EOFs of the annual temperature in other periods of the simulation also remain in essence unchanged.

Several measures can be used to assess the quality of the reconstructed temperature field. Measures that keep the spatial information can be the local time correlation between the simulated and reconstructed temperature or the Brier skill score, also known as the percentage of explained (or resolved) variance, defined as

$$\beta_i = 1 - \frac{\sum_t [T_{\rm sim}(x_i, t) - T_{\rm rec}(x_i, t)]^2}{\sum_t T_{\rm sim}(x_i, t)^2}, \qquad (3)$$

where $T_{\rm sim}$ is the simulated temperature deviations from the mean of the calibration period, $T_{\rm rec}$ is the reconstructed temperature deviations, and i is the gridpoint index.

4. Spatially decimated versus complete temperature fields

The statistical model of MBH98 is calibrated using available gridded temperature fields derived from observations. For the model calibration these authors retain only those grid points that are continuously covered in the period 1901–80. These grid points are located mainly, though not exclusively, at low and midlatitudes in the Northern Hemisphere, on the continental landmasses, and in coastal areas. The area covered by these grid points is about 30% of the global surface. There-

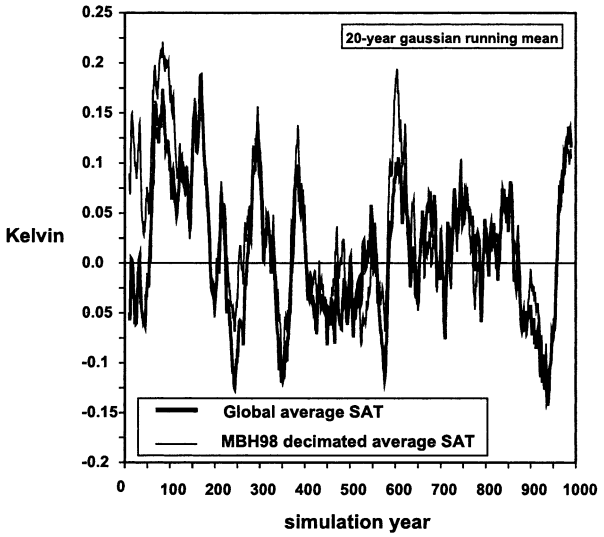

FIG. 1. Annual near-surface temperature simulated by the ECHO-G model, globally averaged, and averaged over the areas used by MBH98 for the calibration of their statistical model.

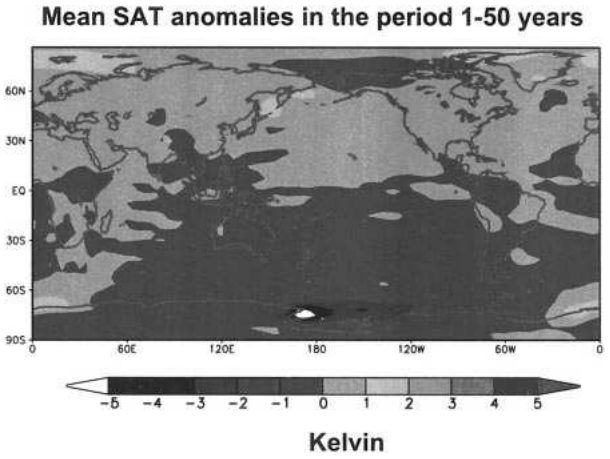

FIG. 2. Mean near-surface temperature anomalies in the first 50 yr of the climate simulation.

fore, the actual temperature reconstructions in MBH98 are limited, in principle, to those areas and the question arises to what extent the reconstructed temperature variations are representative of the global average temperature. Figure 1 shows the temperature simulated by the model ECHO-G, averaged over all model grid points, and averaged only over the grid points used by MBH98 in their calibration. This figure demonstrates that in general both curves are highly coherent at all timescales (correlation 0.85), so that it seems that the average of the MBH98 decimated temperature fields could be considered a good indicator of the global temperature. However, at the beginning of the simulation, in the period approximately 1–50 yr, both evolutions differ markedly. Figure 2 shows the temperature anomalies of these first decades. In this figure, a contrast between the Northern and Southern Hemispheres is apparent, together with strong cold anomalies in the Ross Sea. This pattern of temperature anomalies cannot be captured by the either the MBH98 set of proxy records or the gridded temperature set used for the calibration. Furthermore, any statistical reconstruction based on pseudoproxies mimicking the MBH98 set will also fail to reproduce this particular spatial anomaly pattern. The possibility that this temperature pattern arises from an incomplete spinup of the coupled model cannot be excluded. Actually, it does not appear again in the rest of the simulation. If the incomplete spinup were the reason, it would not have consequences for the validity of the MBH98 approach in the real world. However, this should be tested in climate simulations with other models.

The use of spatially incomplete gridded fields for the calibration could also affect the estimation of the temperature EOFs. Within the climate model simulation one can compare the results of an EOF analysis of the complete field and of the MBH98 decimated temperature field, shown in Fig. 3. The correlation matrix between the PCs (Table 1) helps associate the individual EOFs in both datasets. The leading EOF of the decimated field shows a positive sign almost everywhere, with stronger anomalies in northern North America. The global counterpart also shows large areas with the same sign, but its magnitude differs in some regions, especially in the Indian Ocean (strong positive anomalies) and in northern Europe (negative anomalies). It also shows areas with alternating signs in the Southern Hemisphere that are not covered in MBH98. The second EOF of the decimated fields shows the known temperature pattern associated with the Arctic Oscillation (Thompson and Wallace 1998), with a temperature contrast between the continents and oceans in the Northern Hemisphere, and a temperature seasaw between Greenland and western Europe. This EOF can be partially recognized as the third EOF of the complete field. The third decimated EOF shows strong positive anomalies in the eastern equatorial Pacific (at the few grid points located there), moderate positive anomalies in the northern Pacific, and negative anomalies over eastern Greenland. The corresponding EOF in the complete field (the second) shows the characteristic Southern Oscillation signal. Its structure deviates somewhat from the decimated EOF in the south-central tropical Pacific. Only the three leading principal components are correlated with the global mean temperature ($r = 0.75$, $r = 0.34$, and $r = -0.36$, respectively). Therefore, together they can describe about 80% of the variability of the global annual mean temperature.

The number of PCs retained in Eq. (2) depends on the number of pseudoindicators. From the leading 15 those that were sufficiently resolved by the pseudoproxies in the calibration period were retained. However, the number of retained PCs had only a minor influence on the reconstructed mean global temperature, provided that the leading three were included. That was

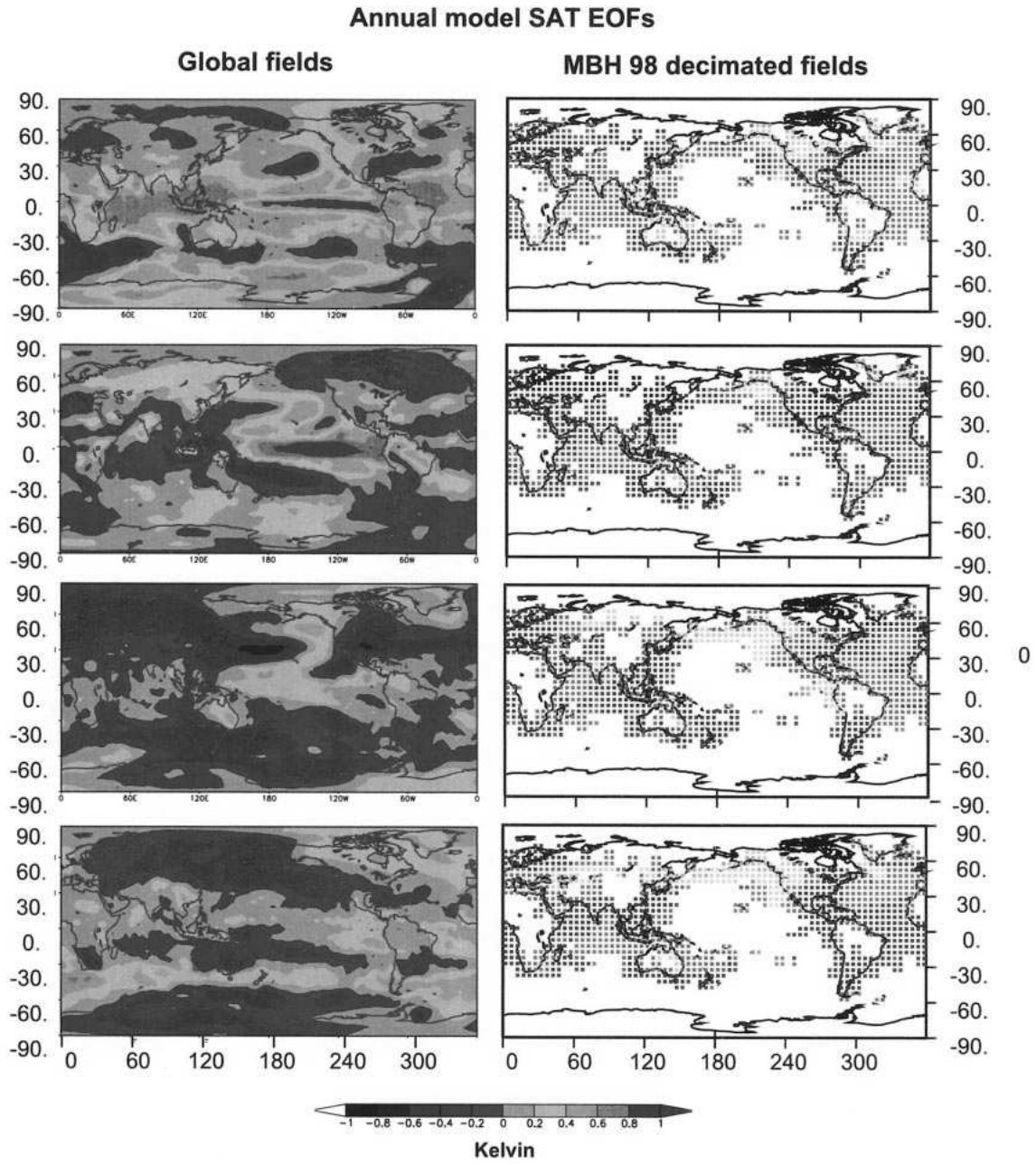

FIG. 3. The four leading EOFs of the mean annual temperature in the control simulation with ECHO-G for the global and MBH98 decimated fields. The patterns were derived from standardized temperature time series, but in this figure they have been rescaled by the local standard deviation.

TABLE 1. Correlation matrix of the leading PCs of the complete temperature fields and the MBH98 decimated fields. Outstanding correlations are boldfaced.

Complete fields	Decimated fields			
	PC1	PC2	PC3	PC4
PC1	**0.81**	0.10	0.09	0.14
PC2	0.36	0.39	**0.66**	0.04
PC3	0.18	**0.84**	0.25	0.04
PC4	0.23	0.17	−0.18	−0.30

mostly the case, except for a very low number of pseudoproxies (15–20) and unrealistic pseudoproxy configurations, but these were not excluded from the following analysis. The leading PC was always retained as predictor.

In contrast to the EOFs derived in MBH98 from gridded observations, the ECHO-G EOF analysis shows no patterns with essentially the same sign over the whole globe. This fact is presumably due to the lack of external forcing in this simulation. In the real world at least two

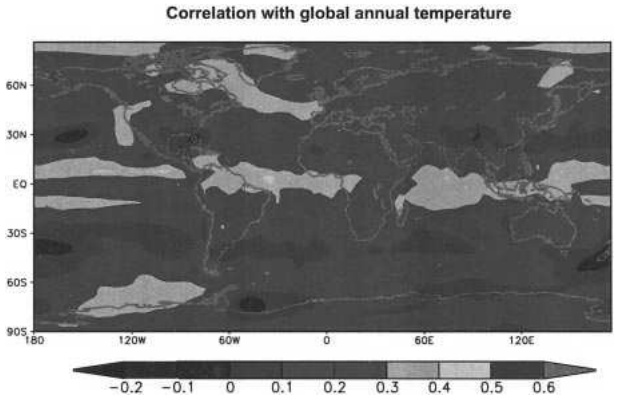

FIG. 4. Correlation pattern between the gridpoint annual temperature and the annual global mean temperature in the ECHO-G simulation.

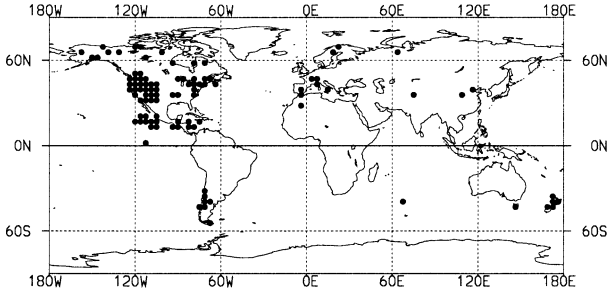

FIG. 5. Position of the grid points in the climate model ECHO-G at a resolution T30 most closely located to the proxy indicators in MBH98.

factors, greenhouse gas atmospheric concentrations and solar variability, contribute to a global in-phase variations of temperature, albeit not in the interannual frequency range. Other control simulations with constant external forcing performed with other models (Stouffer et al. 2000) do show a leading air temperature EOF with an almost uniform sign over the globe.

A question related to the spatial coverage of the data used to calibrate the model, is the optimal location of the (pseudo) proxy indicators. This question has been addressed by Bradley (1994), within the framework of a long simulation with the GFDL coupled climate model. Perhaps the most straightforward method to identify the areas more closely connected to the global mean temperature is to calculate the correlation pattern between the local annual temperatures and the global annual mean. This is shown in Fig. 4 for the simulation with ECHO-G. This figure can be directly compared with Fig. 2a in Bradley (1994). In general, the correlation coefficients are lower in the ECHO-G simulation than in the GFDL simulation. Similar features in both simulations include the low correlations in the South Atlantic, Indian, and South Pacific Oceans, and generally high correlations in the tropical regions, albeit the ECHO-G model results do not show the clear correlation pattern in the tropical Pacific as the GFDL model does. However, an important difference is that, in the ECHO-G simulation, higher correlations are also found at higher latitudes in the Northern Hemisphere and in the east Pacific sector of the Antarctic Ocean. This latter region plays an interesting role in the ECHO-G simulation, as it will be described later.

Therefore, it seems that the search for optimal sites for temperature reconstructions is model dependent and that a direct translation to the real world has to be considered with caution.

5. Dependence of the temperature reconstructions on the number of proxies

As stated before, the role of climate proxy data will be played by the simulated air temperature. When choosing the location of the pseudoproxies, the proxy dataset used by MBH98 has been taken as a guideline, but for the reasons mentioned above not all the real proxies of MBH98 will be considered. Those not sensitive to temperature will not be retained in the present study. The number of proxies used by MBH98 depend on the period when the reconstruction is attempted, since not all proxy data are available throughout the whole period back to 1400 A.D. In the most recent period a number of 112 indicators was used. Since the horizontal resolution of the ECHO-G model in this simulation is rather coarse (3.75°) some of the locations of the temperature-sensitive indicators have to be associated with the same grid point in the model. Therefore, in this study the maximum number of pseudoindicators is limited to 98. The position of the grid points playing the role of climate indicators is shown in Fig. 5. Most pseudoindicators are situated in the Northern Hemisphere over land. Those originally located on islands now correspond to ocean grid points. This difference should not be critical, as the relationships between the local temperature and the large-scale temperature field would probably be very similar to the case that small islands were resolved in the climate model. No pseudoindicator is located in Antarctica.

The basic calibration period was chosen to be the last 80 yr of the simulation. This choice should be arbitrary as far as it is not influenced by the spinup phase of the simulation. Our choice was somewhat dictated by a relatively sharp increase of global mean temperature seen in the last decades of the simulation, which in some sense could play the role of the global warming observed in the real world in the last decades, although of a smaller magnitude. The calibration period in MBH98, 1900–80 A.D., obviously lies within the presently observed global warming phase.

The simulated annual temperature field throughout the whole simulation can be reconstructed once a set of pseudoindicators has been chosen. To estimate the influence of the number of pseudoproxies used in the reconstruction, the reconstruction method was applied using a decreasing number of pseudoindicators n_{pi}, ranging from the original number $n_{pi} = 98$ down to $n_{pi} = 15$. In the real reconstruction by MBH98 the set of proxy

data remaining backward in time is given by external and independent factors, such as archeological investigations on favorable tree-ring sites, and therefore this subset is presumably a random selection of the original proxy dataset. To account for this source of uncertainty, ensemble reconstructions of 50 members each were carried out, where for each member of the ensemble n_{pi} pseudoindicators were chosen at random from the original set of 98.

The reconstruction exercise can now be carried out with several setups. The calibration can be performed using either the EOFs derived from the complete fields or from the decimated fields; and also, the reconstructed average temperature can be compared to the global average or to the decimated average. The basic setup that would mimic the error in the MBH98 setup would be a calibration with the EOFs of the decimated fields and verification against the global average. This setup includes the possible errors of the statistical method and the errors derived from the limitations of the observational record. Therefore, it could be also interesting to assess the influence of the use of the complete EOFs, and also to estimate the possible overestimation of the true skill in the MBH98 approach when the reconstructed temperature is verified against the decimated temperature field. Therefore, in the present exercise, a verification against the decimated average temperature can give an indication of the level of this overestimation.

Of all possible combinations, three seem especially interesting: 1) calibration with decimated EOFs and verification against either the global mean temperature or 2) against the decimated mean temperature; 3) the use of the complete EOFs, verifying against the global average temperature. The latter can be used as a benchmark for the previous two setups.

Figure 6 shows the ensemble root-mean-squared error and the ensemble mean explained variance as a function of the number of pseudoindicators. The explained variance increases, as expected, with the number of pseudoindicators. The smallest error is found in the reconstruction of the decimated average temperature with the decimated field EOFs, a result that appears reasonable, as the decimated fields contain less degrees of freedom than the complete fields and the decimated EOFs contain enough information to reconstruct at least the decimated temperature field. The error is larger when comparing the reconstruction based on decimated field EOFs with the global average temperature (the equivalent to the MBH98 reconstruction). This result could be also expected, since the information of the areas not included in the decimated field is only partially retained in the decimation. This latter error is similar to the one attained with complete EOFs, indicating that the EOFs derived from decimated fields already contain more or less the same information as the complete EOFs, as far as the global average temperature is concerned.

Figure 6 indicates that, with 98 pseudoproxies, the use of the full mean instead of the decimated mean

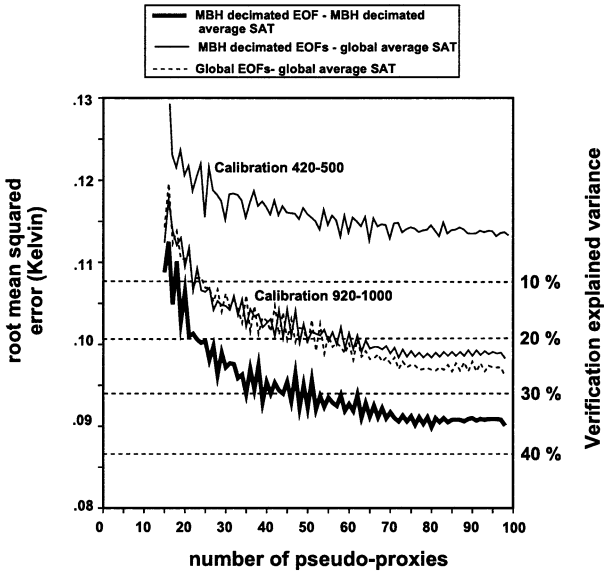

FIG. 6. Root-mean-square error and explained variance of the reconstructed annual global temperature in the verification period as a function of the number of pseudoindicators for different reconstructions setups. The calibration periods are indicated.

temperature causes a drop from 35% to 25% in the explained variance of the global average temperature.

The minimum number of proxies required for a meaningful temperature reconstruction depends on the error that one is willing to accept. The results show that the amount of explained variance levels off with about 70 pseudoproxies, so that the marginal increase in explained variance is small when more pseudoproxies are included. Therefore, it can be concluded that a set of 70 indicators is almost optimal for reconstructing the global mean. On the other hand a set of about 15–20 pseudoproxies can be expected to reconstruct approximately 15%–20% of the temperature variance, which amounts to a correlation between two standardized time series of about 0.4. It is noted that other temperature reconstructions of the last millennium (i.e., Jones et al. 1998) are also based on a quite limited number of climate indicators. If the error due to the local transfer function were included in these reconstructions, they might be resolving a fairly small amount of variance at the interannual timescale. The situation is, however, better at longer timescales, as will be described in section 7. This is consistent with the previous findings of Jones et al. (1997), who tried to estimate the number of effective independent degrees of freedom of the near-surface air temperature field, in model simulations and in observations. The number of degrees of freedom was found to be dependent on both the model and the timescales. At interannual timescales, four climate models yielded numbers between 21 and 52, so that 15 pseudoproxies may be too small a number. At decadal timescales, the number of degrees of freedom decreases to a range of 7–27.

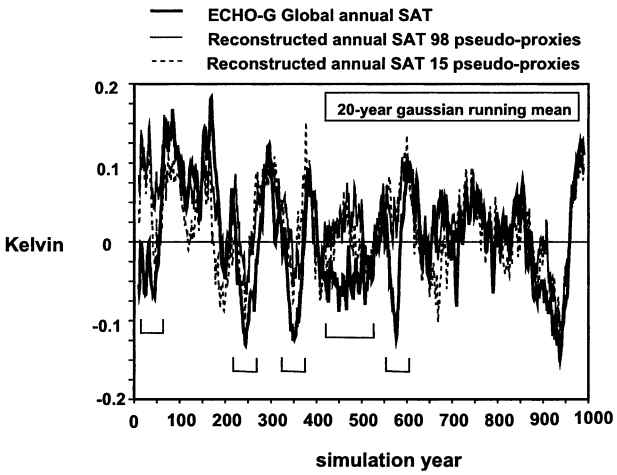

FIG. 7. Simulated global mean annual temperature and a statistical temperature reconstruction using the decimated fields EOF and 98 and 15 pseudoindicators, respectively. The statistical model was calibrated in the period 920–1000.

We now turn our attention to the question of whether the errors of the temperature reconstructions are distributed uniformly over time or whether there are some periods where the errors tend to be larger than in other periods, regardless of the number of pseudoindicators. This would indicate that the method captures certain climatic configurations better than others. Figure 7 displays the simulated global mean temperature of one member of the $n_{pi} = 98$ and of the $n_{pi} = 15$ ensembles. The most evident errors in the reconstructions are, apart from the first 100 yr of the climate simulation already mentioned, the periods 240–60, 340–60, and 555–75, marked in Fig. 7, when the reconstructions tend to underestimate global cooling. An additional period of disagreement occurs between the years 400 and 500. This latter period is longer than the others and also the character of the reconstruction errors is different. This can be seen in Fig. 8, where the mean temperature anomalies of these four periods are displayed. In the first three periods, strong negative anomalies around Antarctica and south of Greenland are found, whereas in the fourth

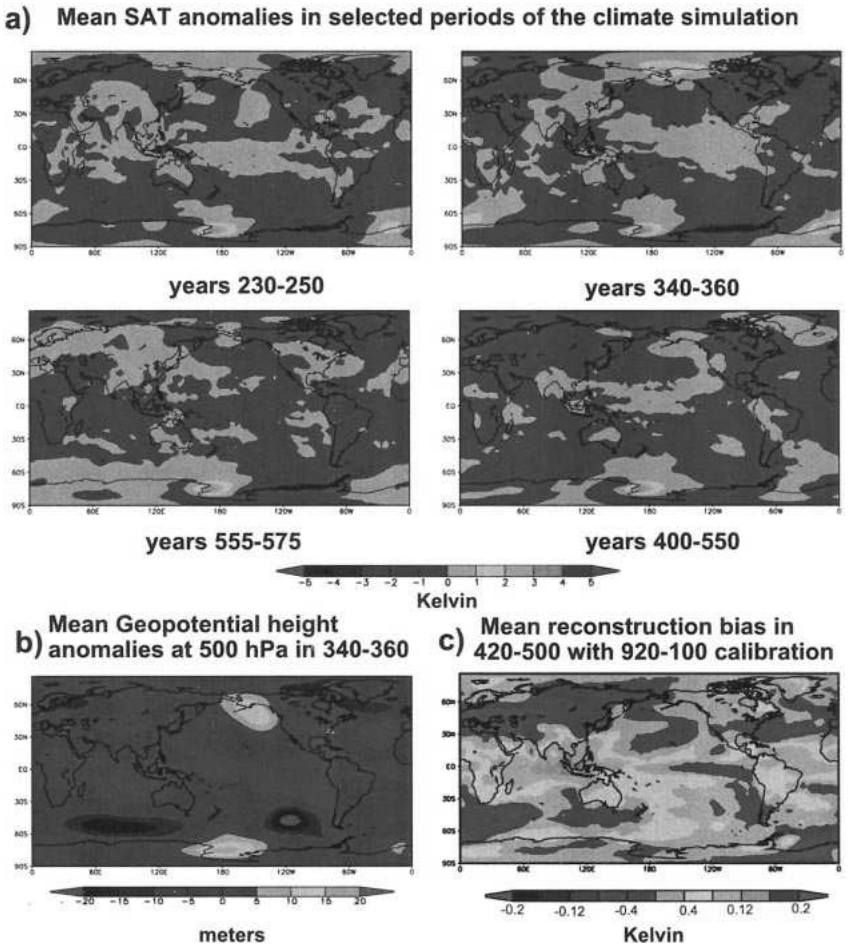

FIG. 8. (a) Near-surface temperature anomalies in the climate simulation in selected periods (see Fig. 7); (b) anomalies of the geopotential height in the period 340–360; (c) mean temperature reconstruction bias in the period 420–500 using the model calibrated in 920–1000.

period these anomalies are not present. Figure 8 also illustrates the anomalies of the geopotential height at 500 hPa for the years 340–360 (for the other periods the anomaly pattern is quite similar). The geopotential height displays anomalies at high latitudes in both hemispheres in areas that could be physically related to the temperature anomalies. The geopotential anomaly pattern does not bear a clear resemblance to any of the annular modes of atmospheric circulation (Thompson and Wallace 2000).

It seems therefore that a temperature anomaly configuration in the ECHO-G model exists, which is also associated to consistent variations of the geopotential height. This configuration is not properly grasped by the network of pseudoproxies, mainly located in the Northern Hemisphere over land. It should not be forgotten that this variability mode could be just a feature of the ECHO-G model and not necessarily a variability mode of the real climate. A more thorough physical analysis of this variability mode is beyond the scope of this paper.

6. Influence of the choice of calibration period

As noted before, the period between the years 400–500 appears to be somewhat special. The reconstructed temperatures show global average temperatures steadily above the mean of the calibration period, whereas the simulated global temperature is lower than normal for a fairly long period of time. The skill of the reconstruction if the calibration had been carried out in this period is investigated in this section. Only the results of the main reconstruction setup are shown, namely, calibration using the decimated field EOFs and comparison to the global average temperature. The two leading EOFs of the decimated temperature field look very similar to the EOFs derived in the simulation years 920–1000 (not shown) with spatial correlations of 0.8 and 0.6, respectively. Also, the scores of the EOFs, calculated as projections of the EOF patterns onto the anomalies in the other period (e.g., the 920–1000 EOF projected onto the 420–500 anomalies, and vice versa) are also correlated for the two leading EOFs (0.9 and 0.5, respectively). However, the skill of the 420–500 model is clearly lower, as can be seen in Fig. 9, where the temperature reconstruction with 98 pseudoproxies is displayed, and in Fig. 6, where the explained variance of this latter model as a function of the number of pseudoproxies is also shown. The lower skill of the statistical model in this period is observed not only in the verification period, but also in the calibration period. For instance, for 98 pseudoproxies the model reaches a calibration explained variance of 18%, compared to an calibration explained variance of 35% in the period 920–1000.

Perhaps a more serious problem is that the 420–500 model is also unable to reproduce the low-frequency evolution of the global temperature anomalies in its own

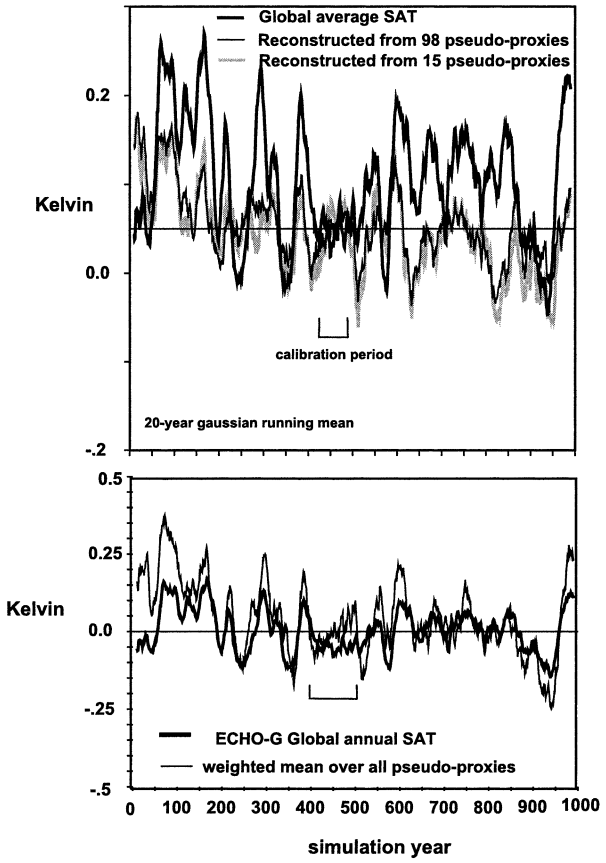

FIG. 9. (top) Global average near-surface temperature simulated by ECHO-G and temperature reconstructions based on 98 and 15 pseudoindicators using decimated fields EOFs. The calibration period is 420–500. (bottom) Annual temperature averaged over the 98 pseudoproxies, compared to the global average temperature.

calibration period (Fig. 9), although the reconstructed and simulated temperatures are still relatively highly correlated at high frequencies. This is illustrated in Fig. 10, where the time correlation between the low-pass-filtered simulated and reconstructed global mean is shown. It shows that despite the opposite behavior of both curves at long timescales, they show a fairly high correlation at timescales shorter than several years. The reason for this striking behavior lies in the different low-frequency temperature evolution of the pseudoproxies, compared to the global or even hemispheric means. The average temperature of the pseudoindicators is also displayed in Fig. 9. The different evolution of the pseudoproxy temperature is mainly due to the North American and west European locations (not shown), and these are able to strongly influence most of the reconstructions. From these results it can be speculated that in the period 400–500 some low-frequency process modulating the temperature contrast between the landmasses and the ocean in the Northern Hemisphere is more active than in other periods. The special character of the period 400–500 is also reflected in the evolution of the Arctic

FIG. 10. Correlation between the simulated global average temperature and the reconstructed temperature based on the decimated fields EOFs with 98 pseudoproxies, in the period 400–500, as a function of timescale filtering. The high-pass and low-pass filter contributions to the correlation are shown. The calibration period was 420–500, however, a similar picture is obtained when the statistical model is calibrated in 920–1000.

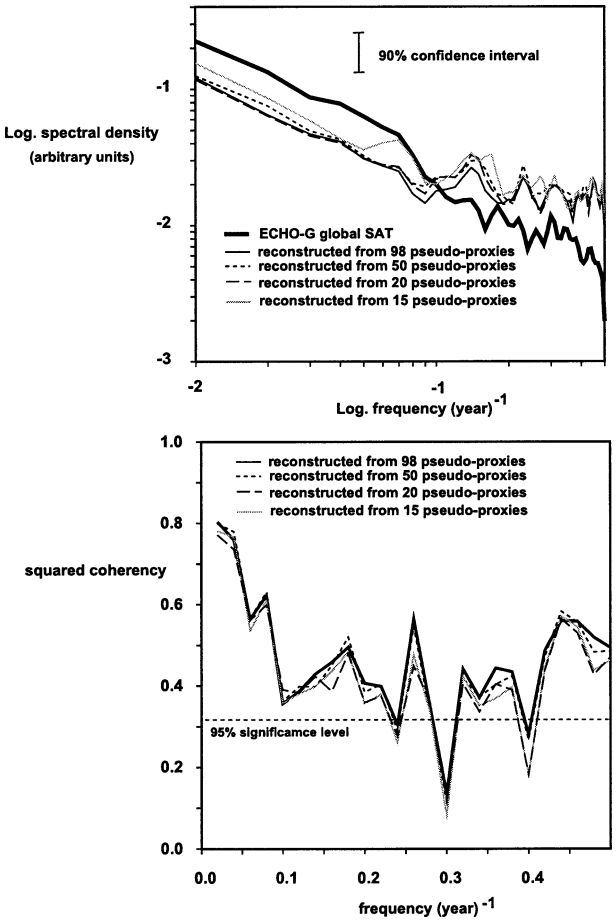

FIG. 11. (top) Spectra of the global average temperature (SAT) and of the reconstructed temperatures using 98, 50, 20, and 15 pseudoindicators. (bottom) Coherence between the global average SAT and the reconstructed temperature using simulated 98, 50, 20, and 15 pseudoproxies. The significance level was estimated by Monte Carlo simulations assuming independent red noise processes.

Oscillation index (I. Fischer-Bruns 2002, personal communication). To gain insight into the nature of such a process the mean error of the reconstructions in 420–500 using the calibration period 920–1000 is shown in Fig. 8c. The results strongly point to some type of ENSO-like behavior in the tropical Pacific–Indian Ocean sector and in the tropical Atlantic. The interannual root-mean-squared error of the reconstruction shows basically the same spatial structure. A detailed dynamical analysis is certainly needed to clarify the model behavior in this period. In any case, it seems that the ECHO-G simulation exhibits periods where the statistical relationships at decadal timescales cannot be simply extrapolated from the interannual timescale. The choice of calibration period is therefore relevant.

7. Spectral characteristics of the temperature reconstruction

The exercise shown in this paper aims at the reconstruction of the global mean temperature at interannual timescales, based on the annual time series of pseudoindicators. However, in many paleoclimatic studies the focus lies usually on longer periods, decadal and longer. At these longer timescales, the fraction of variance explained by the temperature reconstruction will probably be larger, since the process responsible for lower-frequency temperature variations will presumably have a large-scale character.

Also relevant is the level of variability of the reconstructions at different very long timescales. For instance, there are now several reconstructions of temperature indices of the last centuries that show different levels of variability at long timescales. Jones et al. (1998) presented a reconstruction of a global temperature index, that primarily reflects the summer temperature variations, based on considerably fewer climate indicators than MBH98. This reconstruction displays climate deviations at centennial timescales almost twice as large as those present in the MBH98 reconstruction. Furthermore, temperature reconstructions based on borehole measurements (Huang et al. 2000) seem to indicate even cooler conditions in the past centuries and a stronger warming trend toward the twentieth century.

Figure 11a shows the spectra of the temperature reconstructions based on the decimated field EOFs for different numbers of pseudoproxies. The spectrum of

the global average temperature (SAT) is also displayed. In general, all reconstructions show a higher level of variability than the simulated global mean for periods shorter than about 10 yr. This might be due to the fact that the reconstructions are based on a much smaller number of degrees of freedom than the global average and the sampling variability should therefore be larger. However, at lower frequencies the global temperature systematically shows a level of variability a factor of 2 larger than in the reconstructions. Although the individual differences with the spectrum of the global SAT lie almost within the 90% confidence interval, the fact that all reconstructions systematically show a smaller level of variability may be indicative that processes that are responsible for the low-frequency variability are not adequately captured in the statistical reconstruction. This is, to some extent, not surprising since the calibration of the statistical model reflects the high-frequency relationships between the pseudoproxies and the temperature field. Further analysis with other statistical methods is probably needed to assess the role of the statistical model in the level of reconstructed variance and its relevance in establishing the significance of the global warming observed in the twentieth century.

Figure 11b shows the ensemble mean coherence spectra between the reconstructed temperatures and the simulated global temperature. The coherence reconstruction with 98 pseudoindicators remains more or less at a constant level (around 0.5, corresponding to an explained variance of 25%) from the interannual timescale up to timescales of approximately 5 yr. For longer timescales the coherence increases with decreasing frequency. For the longest timescales the coherence reaches about 0.8 (about 60% of explained variance). A similar behavior is observed for the reconstructions with a lower number of pseudoproxies, although the coherence is, as expected, slightly lower than for 98 pseudoproxies. This result underlines the potential for the estimation of past natural global temperature variability with a relatively small number of proxy indicators at timescales of decades, which is the relevant timescale for studies of detection of climate change in the observational record. These results indicate that at decadal timescales the number of indicators required for a reasonable reconstruction of the global mean temperature can be lower than at interannual timescales, and are in line with previous findings by Jones et al. (1997).

8. Concluding remarks

The aim of empirical models linking local proxy indicators and large-scale climate indices, such as the annual global mean temperature, is to estimate the past evolution of these large-scale indices. In this paper a reconstruction exercise has been performed in the world simulated by a global climate model, where the role of the temperature pseudoindicators was played by air temperature simulated at selected grid points.

There are at least two sources of uncertainty in the real climate reconstructions: one is given by the empirical transfer function between the real proxy indicators and the real temperature field; the second has its origin in the statistical reconstruction model itself. In our exercise only the last source of uncertainty can be quantified; in the real world the reconstructions errors are likely to be higher.

In the climate simulated by the ECHO-G model without external forcing, the global annual mean temperature can be reasonably well reconstructed with a few pseudoproxy indicators placed in the locations of a reasonable network of real proxy indicators, especially at decadal timescales. However, some deficiencies related to the irregular distributions of pseudoindicators remain. In some simulation periods, lasting several decades, the temperatures in the Southern Hemisphere do not evolve in line with their counterparts in the Northern Hemisphere. Since the network of pseudoproxies is located mainly on the Northern Hemispheric landmasses, they are unable to reproduce the global temperature anomalies in this period. This result supports the warning put forward by Barnett and Jones (2000) on the present lack of Southern Hemispheric proxies. In another period of about 100 yr, the low-frequency relationship between the pseudoproxies and the global temperature was not coherent with their high-frequency correlation, resulting again in a poor reproduction of the global average. The third important caveat found in this study is related to shorter periods of cooling of just two decades long, possibly related to oceanic processes along the Antarctic and Greenland coast. These cooler periods are underestimated by the reconstructed temperature.

It could be tempting to relate the deficiencies in the reconstructions to the question of the search of optimal proxy locations. However, comparison of our Fig. 4 with other model results (Bradley 1994) suggests that the ideal proxy locations could be model dependent. Though there are large areas where both models agree to point high correlations between local temperatures and the global averages, important differences remain, so that optimal proxies for one model could not behave optimally in the real world. The climate simulation presented here may even indicate that the optimal proxy location may also not be stationary in time.

It was also found that the skill of the statistical model increases at longer timescales. The coherence between the simulated and reconstructed global annual temperature attains values of approximately 0.8 at timescales of 10 yr, even when only 15 pseudoindicators are used. This result suggests that if the focus lies in climate reconstructions at decadal timescales, one may speculate that a network of only a few proxy indicators with a high signal-to-noise ratio might prove sufficient to reconstruct the global temperature. However, the present study also indicates that the proxy datasets should aim at covering both hemispheres, since their temperature evolutions in this control simulation are not necessarily

strong at all times. This simulation, however, lacks the external forcing influencing the real climate. The analysis of an externally forced simulation of the past five centuries, now under way, might show a higher correlation between both hemispheres.

Acknowledgments. We thank the editor, Prof. Michael Mann, and three anonymous reviewers for their critical comments that greatly contributed to improve this manuscript. This work was partially funded by the Helmholtz Society through the project KIHZ (Climate variations in historical times). Dennis Bray provided editorial assistance.

REFERENCES

Appenzeller, C., T. F. Stocker, and M. Arklin, 1998: North Atlantic oscillation dynamics recorded in Greenland ice cores. *Science,* **282,** 446–449.

Barnett, T. P., and P. D. Jones, 2000: Reply. *Bull. Amer. Meteor. Soc.,* **81,** 2990–2992.

——, and Coauthors, 1999: Detection and attribution of recent climate change: A status report. *Bull. Amer. Meteor. Soc.,* **80,** 2631–2659.

Bradley, R. S., 1994: Are there optimum sites for global paleotemperature reconstruction? *Climate Variations and Forcing Mechanisms of the Last 2000 Years,* P. D. Jones et al., Eds., NATO ASI Series, Vol. 41, Springer-Verlag, 603–624.

——, M. K. Hulme, and M. E. Mann, 2000: Comments on ''Detection and attribution of recent climate change: A status report.'' *Bull. Amer. Meteor. Soc.,* **81,** 2987–2990.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn, 1998: Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature,* **393,** 450–455.

——, T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov, 2001: Low-frequency temperature variations from a northern tree ring density network. *J. Geophys. Res.,* **106,** 2929–2941.

Cook, E. R., R. D. D'Arrigo, and K. R. Briffa, 1998: A reconstruction of the North Atlantic Oscillation using tree-ring chronologies from North America and Europe. *Holocene,* **8,** 9–17.

Crowley, T. J., 2000: Causes of climate change over the last 1000 years. *Science,* **289,** 270–277.

Cubasch, U., R. Voss, G. C. Hegerl, J. Waszkewitz, and T. J. Crowley, 1997: Simulation of the influence of solar radiation variations on the global climate with an ocean–atmosphere general circulation model. *Climate Dyn.,* **13,** 757–767.

Huang, S., H. N. Pollack, and P. Y. Sten, 2000: Temperature trends over the past five centuries reconstructed from borehole temperatures. *Nature,* **403,** 756–758.

Jones, P. D., T. J. Osborn, and K. Briffa, 1997: Estimating sample errors in the large-scale temperature averages. *J. Climate,* **10,** 2548–2568.

——, 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.

Legutke, S., and R. Voss, 1999: ECHO-G, The Hamburg Atmosphere–Ocean Coupled Circulation Model. Deutsches Klimarechenzentrum Rep. 18, 62 pp. [Available from Deutsches Klimarechenzentrum, Bundesstrasse 55, D-20146 Hamburg, Germany, or online at http://www.dkrz.de/forschung/reports.html.]

Luterbacher, J., C. Schmutz, D. Gyalistras, E. Xoplaki, and H. Wanner, 1999: Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys. Res. Lett.,* **26,** 2745–2748.

Madden, R. A., D. J. Shea, G. W. Branstator, J. J. Tribia, and R. O. Weber, 1993: The effects of spatial and temporal sampling of estimates of the global mean temperature: Experiments with model data. *J. Climate,* **6,** 1957–1066.

Mann, M. E., and S. Rutherford, 2002: Climate reconstructions using pseudo-proxies. *Geophys. Res. Lett.,* **29,** 1391–1394.

——, R. S. Bradley, and M. K. Hughes, 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature,* **392,** 779–787.

——, ——, and ——, 1999: Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.,* **26,** 759–762.

Pfister, C., and Coauthors, 1999: Documentary evidence on climate in sixteenth-century Europe. *Climatic Change,* **43,** 55–110.

Stouffer, R. J., G. C. Hegerl, and S. Tett, 2000: A comparison of surface air temperature variability in three 1000-year coupled ocean–atmosphere model integrations. *J. Climate,* **13,** 513–537.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25,** 1297–1300.

——, and ——, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000–1016.

von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.