Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium.

Hugues Goosse[1]
Université Catholique de Louvain,
Institut d'Astronomie et de Géophysique G. Lemaître
Chemin du Cyclotron, 2, B-1348 Louvain-la-Neuve, Belgium
Phone: 32-10-47-32-98, Fax: 32-10-47-47-22, e-mail: hgs@astr.ucl.ac.be

Hans Renssen

Faculty of Earth and Life Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

Axel Timmermann

IPRC, SOEST, University of Hawaii, 2525 Correa Road, Honolulu, HI 96822, USA

Raymond S. Bradley

Department of Geosciences, University of Massachusetts, 611 North Pleasant Street Amherst, MA 01003-9297, USA

Michael E. Mann

Department of Meteorology and Earth and Environmental Systems Institute (EESI), Pennsylvania State University, University Park, PA 16802-5013

Submitted in revised form to Climate Dynamics

15-12-2005

[1] corresponding author

Abstract

We present and describe in detail the advantages and limitations of a technique that combines in an optimal way model results and proxy-data time series in order to obtain states of the climate system consistent with model physics, reconstruction of past radiative forcing and proxy records. To achieve this goal, we select among an ensemble of simulations covering the last millennium performed with a low-resolution 3-D climate model the ones that minimise a cost function. This cost function measures the misfit between model results and proxy records. In the framework of the tests performed here, an ensemble of 30 to 40 simulations appears sufficient to reach reasonable correlations between model results and reconstructions, in configurations for which a small amount of data is available as well as in data-rich areas. Preliminary applications of the technique show that it can be used to provide reconstructions of past large-scale temperature changes, complementary to the ones obtained by statistical methods. Furthermore, as model results include a representation of atmospheric and oceanic circulations, it can be used to provide insights into some amplification mechanisms responsible for past temperature changes. On the other hand, if the number of proxy records is too low, it could not be used to provide reconstructions of past changes at a regional scale.

1. Introduction

Previous simulations covering the last millennium using climate models were performed with two types of models. In the first group of models, the high frequency variability of the atmosphere is not taken into account (e.g., Crowley 2000; Bertrand et al. 2002; Gerber et al. 2003; Bauer et al. 2003). This means that those models have very low internal variability on interannual, decadal and even centennial time-scales. The variations simulated during the last 1000 years are thus to a very large extent related to the external forcing applied. The second type of model includes more sophisticated atmospheric and oceanic components that allow for a representation of the variability of the climate system on interannual to centennial times-scales (e.g., Cubasch et al. 1997; Waple et al. 2002, Widmann and Tett 2003; Gonzalez-Rouco et al. 2003; Goosse et al.

2005ab). Some of those models have a coarse resolution and include simplifications that induce an underestimation of the variance associated with some modes of variability, in particular in the tropics (e.g. Goosse et al. 2005ab). Nevertheless, for this group of models, the simulated climate evolution includes, as in the real world, a contribution from both internal and forced variations.

The advantage of the first class of models is that they are generally very efficient and thus numerous sensitivity tests can readily be made. Furthermore, as the internal variability is very weak, the response to the forcing can be studied easily (e.g., Crowley 2000; Bertrand et al. 2002). On the other hand, as those models do not include natural variability, model-data comparisons are difficult to perform, particularly at regional scale where the role of internal variations is very important (e.g., Stott and Tett 1998, Goosse et al. 2005b).

Analyzing the contribution of the forcing in the simulations performed with the second group of models is less straightforward. If an ensemble of simulations is available for a particular model, the mean of this ensemble can be used to isolate the forced response of the system, as the averaging tends to remove the internal variability (e.g., Goosse et al. 2005a). The statistics of the simulations can also be compared to the climate reconstructions (e.g., Jones et al. 1998; Collins et al. 2002) and the mean spatial pattern of the response to a particular forcing could be analyzed (e.g., Cubasch et al., 1997; Rind et al. 1999; Waple et al.,2002; Shindell et al. 2001). However, using only the model results, it is not possible to determine which realization was actually 'selected' by the climate system among all the possible ones. As a consequence, a more detailed comparison between the observed and simulated climate evolution at regional scale could not be performed in this framework. In addition, it is not easy to propose a mechanism that could explain the observed evolution in a region of interest during a particular period because different processes could be dominant for different members of an ensemble of simulations.

The same type of problem occurs for the very recent past. When performing a simulation of the present-day climate with a climate model, only the model climatology and its statistics can be compared to observations. In order to remain as close as possible to the observed evolution, complex data assimilation techniques are needed, as in the reanalysis project that covers the last 40-50 years or at smaller time scales for weather prediction (e.g., Kalnay et al. 1996). The goal of those data assimilation techniques is to constrain in an optimal way the model evolution using observational data.

For the last 50 years, a relatively large amount of observations is available. Consequently, it is possible to force the model trajectory to be close to the observed one on a daily basis. This is not possible for the last 1000 years as the number of records is much lower and those records cover only a small fraction of the Earth. Furthermore, they provide information on seasonal (annual) time-scales at best. Because of this large difference in the temporal and spatial resolution, the technique used in the reanalysis or in weather forecasting could not be used directly for the last millennium.

To overcome these problems, Jones and Widmann (2003) proposed a nudging method in which the atmospheric circulation is modified at every time step in order to remain close to a pattern reconstructed from observations. van der Schrier and Barkmeijer (2005) proposed a different technique in which an artificial forcing is added into the model. The goal of this additional forcing is to ensure that, averaged over the simulation, the atmospheric circulation is close to a reconstructed one while leaving the high frequency atmospheric variability free to evolve according to model dynamics.

Both of those techniques introduce a constraint on the model evolution through the atmospheric circulation. They are thus only useful if a reconstruction of the atmospheric circulation is available. As a consequence, these methods have only been used for selected regions or periods when a large amount of data is available like the North Atlantic/European Sector during the years 1790-1820 (van der Schrier and Barkmeijer 2005) or the very recent past (Jones and Widmann 2003).

Here, we describe a different method that does not need such a reconstruction of the atmospheric circulation as a pre-requisite. As a consequence, it is not dependent on the hypothesis used to obtain those reconstructions and it could be used even for data-sparse periods. The principle of the method is to select among a relatively large ensemble of simulations that has been performed with a global 3-D climate model the one that is the closest to the observed climate. This selection is performed by comparing each simulation to available reconstructions. In this ensemble, the model is driven by different reconstructions of the forcing during the last millennium in order to include the uncertainties associated with the forcing (see table 1).

We thus propose to search, in a library of model simulations, for analogues to past conditions recorded in the proxy records. This is similar to the identification of analogues in meteorology (or looking for very close states of the system for two different periods) which has been widely used in different areas such as atmospheric predictability (e.g. Lorenz 1969), seasonal forecasting (e.g., Livezey et al 1994) and statistical downscaling (Zorita et al 1999). Studies of analogues in meteorology have shown that it is possible to find mediocre analogues but finding a very good analogue of a meteorological situation at a global scale is extremely unlikely because of the large number of degrees of freedom of the system (e.g. Lorenz 1669, van den Dool 1994, Nicolis 1998). In that case, van den Dool (1994) estimates that we need to wait the order of 10^{30} years to have an accurate match.

Obtaining a single experiment that would be able to simulate the observed evolution in the different regions over the entire millennium would then require a prohibitively large number of experiments. As a consequence, the procedure here is applied for the state of the climate system averaged over a period that ranges from 10 to 50 years. For each of those periods, a different experiment can be selected as the “optimal", providing the best agreement between simulation and the reconstructions. As a consequence, the proposed method would not give a real time series of climatic variables over the whole millennium. If the technique is successful, it would rather provide for each period a state of the climate system that is consistent with model dynamics, with reconstruction of the forcing and with climate reconstructions. A future extension of the technique to provide a nearly continuous reconstruction is however relatively straightforward as discussed in section 4.

Although the method could also be used to analyze interannual variability, we are using an averaging period of 10 to 50 years because the number of degrees of freedom of the climate system is lower at the decadal time scale than at the interannual timescale (Jones et al. 1997), making higher the chance to find good analogues (van den Dool 1994). Secondly, the correlation between observations and proxies is generally better on decadal time-scales than on interannual times-scales (Jones and Mann 2004). It is thus more likely to find a good approximation of the real past evolution of the system with a reasonable number of numerical experiments for such decadal variations.

The main goals of the present study are firstly to test if the proposed technique could result in a good fit between model results and reconstructions at a reasonable cost and secondly to illustrate how this technique could be an efficient tool in the study of past climate changes. The model and its forcing are briefly presented in section 2. In section 3, the method used to select the best simulations is described and its advantages and disadvantages are discussed. In section 4, the general skill of the method is tested by comparing local model results with a small number of proxy-based reconstructions of the climate in various regions of the Northern Hemisphere. Thanks to the small number of proxies, the model results can be compared in detail to each proxy time series providing clear and precise information on the technique. Using the proxy-based reconstructions described in section 4, an application of the method is then proposed in section 5 where the simulated temperature averaged over the Northern Hemisphere is analyzed and compared to the recent reconstructions of Mann and Jones (2003) and Moberg et al. (2005). In section 6, the method is applied using the temperature reconstruction of Luterbacher et al. (2004) in Europe for the period 1500-2000 AD. This allows testing the skill of the method during a data-rich period and to compare them to the results described in section 4 for a data-sparse configuration. The study ends with some concluding remark and perspectives.

2 Model description and experimental design

Our version of ECBILT-CLIO-VECODE is identical to the one of Renssen et al. (2005) and Goosse et al. (2005a,b). The atmospheric component is ECBILT2 (Opsteegh et al. 1998; Selten et al. 1999), a T21, 3-level quasi-geostrophic model, with simple parameterizations for the diabatic heating due to radiative fluxes, the release of latent heat, and the exchange of sensible heat with the surface. The model contains a full hydrological cycle that is closed over land by a bucket model for soil moisture. Synoptic variability associated with weather patterns is explicitly computed. Each bucket is connected to a nearby ocean grid point to define the river runoff. Accumulation of snow over land occurs in case of precipitation when the land temperature is below zero. Cloud cover is prescribed following a seasonally and geographically distributed climatology (D2 monthly data set of the International Satellite Cloud Climatology Project, ISCCP, Rossow et al. 1996).

The CLIO model (Goosse and Fichefet 1999) comprises a primitive equation, free-surface ocean general circulation model coupled to a thermodynamic-dynamic sea ice model. The representation of the vertical growth and decay of sea ice is based on a 3-layer model (Fichefet and Morales Maqueda 1997). In the computation of ice dynamics, sea ice is considered to behave as a viscous-plastic continuum. The horizontal resolution of CLIO is 3 degrees in latitude and longitude and there are 20 unevenly spaced vertical levels in the ocean.

ECBILT–CLIO is coupled to the VECODE model that simulates the dynamics of two main terrestrial plant-functional types, forest and grassland, and desert as a third dummy type (Brovkin et al. 2002). It should be noted that the computed vegetation changes only affect the land-surface albedo in ECBILT–CLIO, and have no influence on other processes, e.g., soil hydrology.

The coupled model includes realistic topography and bathymetry. The only flux correction in ECBILT–CLIO-VECODE is an artificial reduction of the precipitation over the Atlantic and over the Arctic. The corresponding water is redistributed homogeneously over the North Pacific (Goosse et al. 2001). The model simulates relatively well the climate outside tropical regions (Goosse et al. 2001; Renssen et al. 2002). Its sensitivity to a CO_2 doubling is 1.8°C, which is in the low range of coupled atmosphere–sea-ice–ocean general circulation models. Thanks to the relatively coarse resolution and the simplified parameterization used in the atmospheric model the coupled model is one to two orders of magnitude faster than a state-of-the-art atmosphere-ocean general circulation model. More information about the model and a complete list of references is available at the address http://www.knmi.nl/onderzk/CKO/ecbilt-papers.html.

A total of 105 simulations have been performed covering the last millennium (or longer). All the experiments are driven by the observed variations of greenhouse gas concentration and aerosol load due to human activities during the period 1750-2000, the influence of sulfate aerosols being taken into account through a modification of surface albedo (Charlson et al. 1991). The observed evolution of greenhouse gases (based on a compilation of ice cores measurements, J. Flueckiger, Pers. Com., 2004) is imposed over the pre-industrial period and the forcing due to change in land-use are taken into account (Ramankutty and Foley, 1999). The latter forcing is applied in the model trough a reduction of the area covered by trees and an increase in grassland as VECODE does not include a specific vegetation type corresponding to cropland. The simulations take into account forcing due to variations of orbital parameters following Berger (1978) as well as forcing associated with tropospheric ozone changes. The latter forcing is applied on both longwave and shortwave fluxes as a time-varying external radiative forcing, constant at hemispheric scale, using values deduced from Berntsen et al. (2000). Furthermore, the evolution of solar irradiance and the effect of volcanism are prescribed using different combinations of the reconstructions that have been used up to now in three-dimensional simulation of the climate of the last millennium (Table 1). Nevertheless, it should be kept in mind that the large uncertainties in the reconstruction of those forcings could have a strong impact on the large-scale temperature patterns simulated by the model (e.g. Bertrand et al. 2002, Goosse et al. 2005a).

For each combination of the forcings, the various simulations differ only in their initial conditions. Those initial conditions were taken from model states in a long control simulation and in previous experiments covering the period 1000-1750. In addition, some simulations covering the last 2000 years were started from a model state simulated with ECBILT-CLIO-VECODE around Year 1AD in an experiment covering the whole Holocene but only driven by change in orbital parameters and greenhouse gas concentration (Renssen et al. 2005).

3. Selection of the optimal simulations

The "optimal" simulation of the ensemble for each period is selected as the one that has the minimum of a cost function *CF*:

$$CF_k(t) = \sqrt{\sum_{i=1}^{n} w_i \left(F_{obs}(t) - F_{\mathrm{mod}}^{k}(t)\right)^2}$$

where $CF_k(t)$ is the value of the cost function for each experiment *k*, for a particular period *t*. *n* is the number of reconstructions used in the model/data comparison. F_{obs} is the reconstruction of a variable *F*, based on observations, in a particular location for the period t. F_{mod}^{k} is the value of the corresponding variable *F* simulated in experiment *k* in the model grid box that contains the location of the proxy-record. If the proxy record represents the evolution averaged over an area larger than one model grid box, F_{mod}^{k} is the average of model results over the grid boxes overlapping this area. As the proxies are generally influenced by climate conditions during a particular period of the year (generally winter, summer or annual average), F_{mod}^{k} is evaluated over the months corresponding to the information recorded in the proxies. w_i is a weight factor, characterizing the statistics and reliability of proxy data. We use the additional constraint:

$$\sum_{i=1}^{n} w_i = 1$$

so that adding new observations in the computation of *CF* would not artificially increase the value of *CF*. w_i could be chosen as constant for all *i*. Alternatively, w_i could be related to the spatial distribution of the proxies or to the correlation of a proxy with observations during the instrumental records.

$CF_k(t)$ is first evaluated for all the simulations and all the time periods. The minimum for each *t* over all *k* experiments is then selected and the corresponding experiment corresponds to the "best" one for this period. The result of this procedure is, for instance, that the simulation that minimizes CF for the period 1000-1010 is experiment number *k*=7, for the period 1010-1020, the simulation *k*=63, for the period 1020-1030, the simulation *k*=31 and so on. In the following, the selected states are grouped together in order to form a pseudo-time series for an easier comparison with the observed climate evolution. In this framework, a pseudo-simulation for the whole millennium is obtained by grouping the best simulations for all the periods *t*. For simplicity, it will be referred to as the "best pseudo-simulation". It has a time resolution identical to the averaging period but it does not consist in a continuous time series since a jump is possible between each 10-year (or 25-year or 50-year) time step because different time steps could be associated with different members of the ensemble of simulations.

Any kind of physical variable (e.g. temperature, precipitation, atmospheric pressure, ocean salinity) can be used in the computation of CF. As our main goal here is to test the technique rather than produce a comprehensive reconstruction including all available data, we will use here, for simplicity, only the data sets proposed in two recent studies. In a first step (section 4 and 5), to test the method in a data sparse period, we will use the proxy temperature series in the Northern Hemisphere that have been carefully selected recently by Mann and Jones (2003) and Jones and Mann (2004) (Table 2). In a second step (section 6), the temperature reconstruction of Luterbacher et al. (2004) is employed to illustrate the skill of the method in a smaller region where a large-scale, gridded reconstruction is available.

We only consider here the extra-tropical Northern Hemisphere since more data are available in this region of the globe. In addition, this alleviates the need to take into account the response of tropical modes of variability to changes in the forcing (e.g., Mann et al. 2005a) that would not be well represented in our model because of its coarse resolution and some simplifications in the physics. As in several previous studies, (e.g., Mann and Jones 2003; Jones and Mann 2004; Moberg et al. 2005), before computing *CF*, we will subtract from the time series of both temperature reconstruction and model results their mean over a reference period and divided the obtained anomalies by the standard deviation of the times series over the same reference period. This mean and standard deviation are obtained after performing the averages over the 10, 25 and 50 periods. The years 1856-1980 have been often chosen as a as reference period (Mann and Jones 2003; Jones and Mann 2004). Nevertheless, here, we are using a longer one (1601-1950, except in section 5 where other values are tested) because we are performing averages over time scales up to 50 years and a longer reference period gives then a larger sample to estimate the mean and the standard deviation.

The first advantage of the proposed technique is that it is a very simple one. It uses existing simulations so there is no risk that introducing a data assimilation technique would require extensive technical developments and a lengthy method validation as for instance in the techniques proposed by Jones and Widmann (2003) or van der Schrier and Barkmeijer (2004). Secondly, the method does not modify the model physics. So the optimal simulation is precisely consistent with the model physics and the forcing for a period *t*. Third, the method only requires the output of the simulations. It is thus possible to perform a large number of tests of the method at a very small computer cost since new proxy data can be included as they become available (or excluded) very easily, without the need to produce a new ensemble of experiments. In particular, different types of cost function could be tested readily.

Nevertheless, the technique has also some disadvantages. First, the number of simulations required to reach a good agreement between model results and proxies is not known *a priori* and this number could be quite large. This is probably not a major problem with the model we are using as a number of simulations of the order of one to a few hundred is clearly affordable. Furthermore, those simulations could also be used for other purposes such as the analysis of the impact of the uncertainties of the forcing or to study the variability of the simulations around the ensemble mean for various regions (e.g. Goosse et al. 2005ab). Nevertheless, this could be prohibitive for more elaborate, computationally intensive models. The second disadvantage is that the method does not provide a continuous record of the evolution of the climate system. This is fine for the fast components of the climate system as an average over a decadal period already provides a large amount of information. Nevertheless, the evolution of the low frequency components of the system could not be studied easily because of the jumps between the different experiments selected as the best one for each period *t*. Using different time windows for the averaging could help to test the importance of this problem, but this would not suppress it.

An alternative approach would be to use the information contained in the proxies directly during the simulations. For instance, it would be possible to run an ensemble of simulations for 1 or 10 years, then select the best of this ensemble as the one that minimizes a cost function CF. The end of this best simulation will be used as the initial state for a new ensemble of simulations over the same period and so on for the whole millennium (e.g., Collins 2003). This would offer the advantage of providing a continuous simulation of the past climate that would be as close as possible to the observed evolution and would avoid the jumps between the selected states obtained when constructing our best pseudo-simulation. Nevertheless, this procedure is far more expensive in computer time since it requires a large number of new experiments before making the first analysis as well as new experiments for each test. As such a method has never been tested up to now, it is first necessary to estimate the skill associated with different hypotheses and to select the data set that provides the largest amount of information, using a less powerful but much cheaper technique, as proposed here. After this initial step allowing to set up the method, the continuous reconstruction would be a next step that might provide for the past millennium an analogue for the model reanalyses performed over the instrumental era of the past 50 years. We can thus consider that here we could provide information on processes whose time scale is smaller than the averaging period, using the reasonable hypothesis that those processes are relatively independent from one averaging period to the other. Analysing the links between two averaging periods or low frequency processes would only be possible when a continuous reconstruction will be available.

The proposed method does not allow for any conclusions regarding the model climatology as only anomalies are used to compute the cost function. However, this is not our goal here. Nevertheless, by selecting among all the possible realizations the simulated evolution that is the closest to the observations, the technique could help to address possible deficiencies in the model. For instance, if the model underestimates the response of some mode of internal variability to a particular forcing (e.g. Robock 2000; Shindell et al. 2001;2003; 2004), the technique could help to select a member of the ensemble in which this mode of variability is optimal. Furthermore, it is possible to check if systematic biases are present in the simulation, such as the best simulations being always warmer/colder than the ensemble mean for a long period in a particular region. This would help to identify some problems in the model and to propose solutions.

4. General skill of the method

In order to test the skill of the method using proxy-based reconstructions from various regions of the Northern Hemisphere extratropics, it is possible to examine the cost function directly (see below). Nevertheless, analysing the actual pseudo-time series obtained from the suite of states selected by the minimisation of the cost function is easier to interpret. For simplicity, we have chosen to give the same value to all the weights w_i in the computation of CF. w_i is thus equal to $1/n$ where n is the number of reconstructions (see section 3). The influence of this choice for w_i on our conclusions is discussed in Appendix 1.

The average correlation between an individual simulation gridpoint and the associated proxy reconstruction is rather low (first column of table 3). The correlation is indeed smaller than 0.3, except for two records (LCV and CHY). The values are even negative for CSC and WGV! For the ensemble mean, the values are still low, although they are generally higher than for the individual members. On the other hand, when the method proposed in section 3 is used to obtain the best pseudo-simulation, the results are far improved. Indeed, the correlations between proxy-based reconstructions and the best pseudo-simulation selected by the technique match the correlations between proxy-temperature estimates and actual temperature measurements calculated during the instrumental era. For some records (e.g. CSC, WGV and WUB), the improvement is spectacular since the correlation increases from nearly zero for the individual simulations to > 0.6 for the best pseudo-simulation.

Plotting the time series of the temperatures during the last millennium allows differentiating two types of improvement associated with the selection of the best pseudo-simulation (Fig. 1). For the majority of the records, the selected state is sometimes warmer than the ensemble mean, and sometimes colder but always well within the range of the variability of the ensemble. It is thus reasonable to consider that the selection procedure just picks up the model simulation that displays internal variability that is the most consistent with the observed one. On the other hand, for one record (CSC), the best state is colder than the ensemble mean for all but three 10-year averages during the period 1000-1350 AD. If the CSC proxy-record represents an unbiased estimate of the large-scale temperature at that time (the correlation of this record with instrumental data is 0.32), this systematic difference could be related to a model deficiency that is cured by the selection procedure. On the other hand, as this record is known to be related to a fairly narrow spring season of temperature variability (Cronin et al. 2003), the differences may relate to the non-representativeness of summer mean temperatures by the proxy record.

The correspondence between model predicted series and the associated proxy series is generally good overall, as illustrated by the high correlations diagnosed, but the agreement is often not as good in a particular region at a particular time, as illustrated on Fig.1. In addition, the time series of the minimum of the cost function indicates that there are some time intervals that appear well reproduced by the model (e.g. 1125-1150, 1375-1425 or 1675-1700 in the example shown on Fig. 2). For those periods, the value of the cost function is of the order of 0.5. On the other hand, for other intervals (e.g. 1175-1200 and 1625-1650), the value of CF for the best pseudo-simulation is higher than 0.8. Those higher values could be related to an unusual state of the real climate system. Such a non-frequent pattern of internal variability would then require additional model simulations in order to have a larger sample of model variability and thus a higher chance to reproduce the mode recorded in the proxies. Alternatively, those higher values of CF could be due to uncertainties in the forcing applied, to model deficiencies that would prevent a reasonable simulation of the climate state observed at this time or to larger uncertainties in the proxy record during that period. Regardless, this observation underscores the fact that the agreement between the best simulation and the reconstruction could be very different for different periods. Consequently, a detailed interpretation of the results of even the best simulation must thus be performed with caution.

The state selected as optimal is generally well within the range of the ensemble for all individual records (Fig.1). However, by construction, the state is located in the tail of the distribution of CF. One single simulation very rarely therefore displays low values of CF for neighbouring periods (Fig. 2). As a consequence, it is not possible to select the same simulation as an optimal one for periods longer than the averaging period used to compute CF.

In order to examine the time-scale dependence of the compatibility between simulation and observations, it is useful to compute the value of CF using a 50-year averaging period for the best pseudo-simulation obtained for the smaller (10 year) averaging period. The time average of the minimum of CF for 50-year means is 0.66 (Fig. 3). When using the temperatures reconstructed for the best pseudo-simulation with a

10-year averaging period, this minimum value of CF for 50-year average is 0.64. On the one hand, this shows that the time evolution at lower frequencies is also well represented when using a 10-year averaging period, suggesting some robustness to the technique. On the other hand, one must keep in mind that, when using a 10-year average, the state averaged over 50 years is actually a blend of up to 5 experiments that are not necessarily compatible with each other, in particular for variables that are not constrained by the proxy records used, such as in the ocean interior.

Table 3 and Fig. 1 show that the technique proposed here using an ensemble of 105 simulations provides a reasonable representation of the observed climate at the location where proxy records were obtained. It is worth examining, however, whether such an ensemble size is either larger than necessary, or too small. To address this, the evolution of the mean of CF over the 1000-year period is analysed as a function of the number of experiments (Fig 3). The decrease is large for the first 30-40 simulations. Additional experiments provide some additional improvement in misfit but the decrease is much slower. For instance, the changes associated with the last 50 (25) realisations are of 0.041 (0.016) while only 10 realisations were needed between realisation #20 and #30 to decrease the cost function by 0.05. An (admittedly optimistic) linear extrapolation of our results would appear to imply that about 150 additional experiments would be required to reduce the cost function by 0.1 and about 1000 to reach values of CF close to zero. Such a large ensemble, if required, is still possible using our model, though it would be time consuming.

Whatever averaging period is used, the evolution of CF as a function of the number of experiments is similar (Fig. 3a). On Fig3a, it also appears that values of CF are always smaller for 10-year averages than for 50-year averages. This is due to the fact that for low frequencies, the internal variability of the model is lower, all the simulations of the ensemble providing nearly the same results. The technique proposed here has thus a smaller impact in the model-data comparison. Besides, the number of degrees of freedom at interannual time-scales is higher, making the search of a good analogue more difficult. Overall, in our experiments, the values of CF reach thus a minimum for time-scales around 10 years, but the improvements for this time-scale compared to the other ones are small.

Interestingly, in our experiments, the number of experiments needed to achieve reasonable skill levels is also relatively independent of the number of series used to compute the cost function (Fig. 3b). Indeed, using only 2 series for each region (i.e. 8 series in total, see table 3) has nearly no impact on the cost function compared to the full 12 series. Nevertheless, this result is clearly dependent on the series that are removed. For instance, suppressing series (e.g. CSC) that display a relatively low correlation between model data and proxies induces a decrease of the cost function while suppressing series with high correlation induces an increase of the cost function. For 4 data points, one in each region, the signal is clearer as the values of the cost function are much lower than for 8 or 12 data series. For such a low amount of data, the correlation between proxy time series and model results at locations where the model is constrained is very high. For instance, in the example selected in Table 4, this correlation is always higher than or equal to 0.88. However, the shape of the cost function is the same for 4,8, or 12 data series, with a large decrease of the cost function for the first 30 realisations and a much smaller one for the last 75 realisations. This is in good agreement with Fig. 2 that displays no particular trend in the minimum of the cost function despite the fact that less data are used to compute CF at the beginning of the second millennium than after 1600 AD (Fig.1 and Jones and Mann 2004).

From table 3 we can also see that the correlation between the best pseudo-simulation and a particular proxy record decreases markedly when this particular proxy is not used in the evaluation of the cost function. The density of the proxy information using only 12 time series is obviously very low and the correlation between the different proxies, for one, is low. As a consequence, when a proxy series is not used to constrain the choice of the best pseudo-simulation, the information contained in the other proxies does not allow for a skillful reconstruction of the observed time series at this location, thus yielding a low correlation at this unconstrained location. When only 4 data records are used, there is a significant risk of over fitting of the model results to the proxy data constraints employed. In such a case, the technique provides a close representation of the climate at the locations where the proxies are used in the computation of CF while, at other locations, the best pseudo-simulation may in fact even be worse than the ensemble mean prediction. Table 3 also illustrates that, as in any assimilation technique, in order to obtain a reasonable description of the state of the system in a particular region, it is necessary to have adequate data representation in that region. Extrapolating the results towards regions where no data are available is thus very risky.

The quality of the best pseudo-simulation will depend strongly on the quality of the proxy data as the best pseudo-simulation will tend to follow the signal contained in this proxy series, and the information at one location cannot alleviate data quality problems at other locations when only a few proxies are used. Nevertheless, the main troubles related to the extrapolation to regions where no proxy is available (or not used in the computation of CF) does not appear to be related to the proxy themselves or to errors in the forcing. When performing idealised experiments in which instead of proxies, pseudo-proxies derived from an independent simulation are used in the computation of CF to choose the best simulations, the conclusions are similar (see Appendix 2). This means that, according to our results, whatever the quality of the proxy record, forcing and model, it is necessary to have a reasonable number of proxies in order to make a good spatial reconstruction of the past temperature variations. The tests performed here have shown that 12 is clearly not enough.

In Fig. 3c, we examine the evolution of CF for the simulations of the group K, the group C, the group D or a blend including 5 simulations of each group. For the forcings described in Table 1, the convergence of the method as well as the values of the cost function are nearly identical if we use one or several different forcings. Furthermore, the values of the cost function are very similar for the different forcing scenarios. This arises from the fact that, in our simulations, the natural variability is large enough to mask for the differences in the forcings applied (see also Goosse et al. 2005a). As a consequence, the comparison of proxy series with model results in the present framework thus does not allow for a determination of which forcing history is associated with the greatest consistency between model results and proxy data.

5. Simulation of the hemispheric mean temperature

The analyses performed above serves to illustrate the details of the technique, but the amount of data used is too low to obtain a reasonable estimate of the regional pattern of the climate evolution during the last 1000 years. However, it is possible to analyse the evolution at the hemispheric scale. It is worthwhile, in particular, to compare our results to recent reconstructions of annual Northern Hemisphere mean temperature. We first consider the reconstruction of Mann and Jones (2003), as all the proxy records used here are also included in their study. We note, nonetheless, that most alternative reconstructions agree well with this reconstruction within its estimated uncertainties (see Jones and Mann 2004 Fig. 5). One limitation in the present analysis is that, no proxy information in the tropical Northern Hemisphere is used (see Table 2) and future generalisations of this study will seek to incorporate tropical proxy records (e.g. long coral series in the tropical Pacific, Indian, and Atlantic oceans) as constraints. At this stage, it should be recalled that the reconstruction of Mann and Jones (2003) is not used in the computation of CF. Our selection of the best pseudo-simulation is thus independent of the statistical technique used by Mann and Jones (2003) to reconstruct the hemispheric mean temperature.

It should first be noted that the agreement between an individual member of the ensemble with this proxy-based reconstruction is much higher for the hemispheric mean series than for the regional times series presented in table 3. This is expected, as the role of natural variability relative to the forced variability is reduced at hemispheric scales (Goosse et al. 2005b). For averaging periods of 10 to 50 years, this correlation with individual members has a mean between 0.60 and 0.72. The best pseudo-model simulation has even a better correlation: it is larger than 0.78 for all the time scale and for all the cost functions selected here. The best correlation (0.84) is reached for a 25-year average.

As shown on Fig. 4, the best pseudo-simulation is not very different from the ensemble mean. This illustrates the important role of the radiative forcing used in the simulations. Without a reasonable forcing, it would not be possible to produce a simulation of the temperature at hemispheric scale that closely matches the observations, even if a technique such as that proposed here is used. Nevertheless, some differences between the best pseudo-simulation from the ensemble mean do produce an even closer match with the Mann and Jones (2003) reconstruction. For instance, the warmest temperature before the 20^{th} century in the ensemble mean occurs in the middle of the 12^{th} century while this period is colder than the one before and the one after it in both the reconstruction and the best simulation. Some differences between the model results and the Mann and Jones (2003) remain but they are well within the error bars of both model results and of the reconstruction.

The comparison in Fig. 4a has been performed for normalised anomalies but a similar conclusion would have been obtained for the temperature anomalies themselves (Fig. 4b). This is not a consequence of the technique used here as all the time series are divided by their standard deviation before computing CF. This is rather due to the fact that our simulations have nearly the same standard deviation as the Mann and Jones (2003) reconstruction (Goosse et al. 2005b). For instance, for the best pseudo-simulation using an averaging period of 10 years, the standard deviation reaches 0.117°C while in the value for the Mann and Jones (2003) reconstruction is 0.103°C. When using a longer interval to scale their record (the years 1856-1995), the standard deviation of the reconstruction of Mann and Jones (2003) is about 25% higher (see ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/jones2004/jonesmann-nhrecon-rescale.txt), which is still close to the model values. We have also computed the best pseudo-simulation using 1860-1980 instead of 1600-1950 as a reference period in order to be closer to the one chosen by Mann and Jones (2003). The influence on our results is minor, the changes due to the choice of a different reference period being difficult to see by eye while the modification of the variance being of the order of 10 % (not shown).

The reconstruction of Mann and Jones (2003) is based on statistical methods calibrated using recent temperature data, assuming thus some stationarity of the links between proxies and climate at the hemispheric scale during the whole millennium. This assumption seems in good agreement with our simulations as very similar results are obtained using the comparison between the results of a 3-D climate model and the proxy reconstruction. Nevertheless, as the range of temperature simulated by various models over the last millennium is quite large (Jones and Mann 2004), the agreement could be model dependent and further experiments are needed to examine this point.

The technique proposed here gives thus a better representation of the climate not only at locations where proxy data are used to constrain the model evolution, but also at the hemispheric scale, which is not constrained by observations. As a consequence, the improvement at regional scales does not in general come at the expense of favourable comparisons at the hemispheric scale. This is not, however, true in all cases. For instance, using only 4 proxy series to select the best pseudo-simulation, the correlation between the best pseudo-simulation at the hemispheric scale and the Mann and Jones (2003) reconstruction is lower than the one for individual members, illustrating again the over fitting of the model results to available data in this case.

In Fig 4b, we have also displayed the recent reconstruction of Moberg et al. (2005). The comparison with our best pseudo-simulation is less straightforward compared to Mann and Jones (2003) reconstruction, as different proxy-data have been selected by Moberg et al. (2005). Nevertheless, the shape of this reconstruction is very similar to our best reconstruction for the pre-industrial periods, the curves diverging only after 1950 (i.e., the last 30 years of the reconstruction of Moberg et al. (2005). Overall, the correlation between our best pseudo-simulation and Moberg et al. (2005) reconstruction for 10-year averages is 0.68. On the other hand, the magnitude of the changes is very different, the standard deviation of the reconstruction of Moberg et al. (2005) (0.20 K for the period 1600-1950, using a 10 year average) being about 2 times larger than the one of our best pseudo-simulation or than in Jones and Mann (2003) (see for instance Mann et al. 2005b for a discussion of the differences between various reconstructions). This illustrates clearly that the technique used here could bring some useful information about the relative temperature changes during the last millennium but could not help in determining the magnitude of those changes. To increase or decrease the magnitude of the model temperature variations, it would be necessary to modify the amplitude of the model response to external forcing and thus would require some parameter changes or some modifications of the representation of physical processes in the model.

6 Model-data comparison of temperature in Europe

In order to compare model results with the European temperature reconstruction of Luterbacher et al. (2004) for 1500-2000 AD, their data have first been interpolated on the model grid. Then, the cost function has been evaluated for each model run and each period, using the formula described in section 3. In the computation of CF, the sum is performed over all the model grid points for which data are available and the differences in the surface of the different grid boxes is taken into account through the weights w_i. The computation is done once per period (for instance once for a 10-year period), including both summer (JJA) and winter (DJF), meaning that the same simulation is selected as the best one for all the seasons.

The evolution of CF as a function of the number of experiments is similar to the one described when using 12 proxy records in section 4: the value of CF decreases quickly for the first 30 experiments while the decrease is much smaller for the following experiments. The value of CF obtained after 105 experiments is 0.85 for 25-year averages, i.e. about 0.2 higher than for the 12 proxy-data (section 4). As expected, the model seems thus to have more trouble to reproduce the temperature at grid scale than at a larger scale. Nevertheless, even for such a stringent test at grid scale, the correlation between the reconstruction and the best pseudo-simulation is reasonable. For decadal averages (Fig. 5), the correlation in winter is higher than 0.7 in the centre of the domain for the period 1500-2000 AD. Smaller correlation is found near the boundaries, probably because of the influence of the unconstrained regions outside of Europe. Those smaller values might also reflect a lower quality of the reconstruction in regions where a small amount of proxy records is available. In summer, the correlation is lower, in particular in Eastern France/South-western Germany. The minimum correlation corresponds to the Alps in the model that culminates at less than 1000 m because of the coarse resolution of the model. This bad representation of the topography as well as of the local soil characteristics and of the particular physical processes important for mountainous regions leads to a poor agreement between model and reconstructions. For reasons that need to be investigated, the problems are apparently more important in our model in summer since winter anomalies are relatively well reproduced there.

The agreement between the best pseudo-simulation and the reconstruction of Luterbacher et al. (2004) is encouraging but a model could not be used to provide information at grid-scale. In any model, one should rather analyse patterns that cover at least a few grid points in order to have a chance to simulate the important physical processes at this scale. Fortunately, constraining model results at grid-scale also provide good agreement at large-scale as described on Fig. 6. Averaged over Europe, the major observed variations are quite well reproduced by the best pseudo-simulation, as for instance the very cold condition in winter during the late 17th Century or the warming in summer during the 1940's-1950's and the following cooling during the 1960's-1970's. Fig. 6 displays changes in the normalised temperature, because the technique is designed to obtain optimal information on this variable. Nevertheless, the agreement on absolute temperature is also good, as our model has nearly the same standard deviation as the reconstruction of Luterbacher et al. (2004) at decadal scale (Goosse et al. 2005b).

Up to now, we have not considered the interannual variations but we mention that subject briefly here as the data of Luterbacher et al. (2004) provide a better estimate of the changes at this frequency than the proxies used in section 4. The evolution of the cost function for interannual changes is similar to the one described above for 10-, 25- and 50-year averages. The correlations between model results at interannual-scale have the same magnitude as for 10-year averages in winter and are even slightly higher in summer. As for the 10-year average, the large-scale agreement between model results and reconstruction is fine when averaged over Europe, both for the interannual variations themselves and for the low frequency changes (Fig.7). We must recall that those low frequency changes of the best pseudo-simulation are not easy to interpret since a 25-year period in the best pseudo-simulation could come from 25 different experiments. Nevertheless, in the present framework, we wanted just to underline that the good agreement at interannual scale was not obtained at the expenses of the low frequency changes. Furthermore, those results illustrate that the technique described here could be used for interannual variations at a reasonable cost, if one is interested in phenomena occurring during one particular year.

Constraining the model results with 12-proxy distributed in the Northern Hemisphere could provide an alternative reconstruction of temperature changes at hemispheric scale (section 6). This is not the case when using the Luterbacher et al. (2004) reconstruction as no data are available outside Europe. Furthermore, Luterbacher et al. (2004), thanks to the relatively high amount of observational records, provide already a very good spatial distribution of the temperature changes. The best pseudo-simulation brings thus very little additional information on temperature in Europe.

On the other hand, the best pseudo-simulation could be used to analyse in detail changes during a particular period. For instance, Fig. 8b, display the surface temperature in the best pseudo-simulation for the period 1690-1700, the coldest decade in Europe (Fig. 6). Actually, this decade was also the coldest in spring in Europe (Xoplaki et al 2005) while autumn and summer did not show exceptional conditions (Luterbacher et al 2004, Xoplaki et al 2005). On fig. 8, the temperature changes that are locally smaller than one standard deviation computed for the 10 simulations with the lowest value of CF are masked. Indeed, this means that among all the simulations that have a low value of CF, very different temperature changes could be found and thus the changes displayed on the average are not significant.

In addition to the information contained in the reconstruction (Fig 8a), Fig. 8b suggest that the temperatures in the eastern Atlantic where also lower than the mean during the period 1690-1700, a region where no data are available in the Luterbacher et al. (2004) reconstruction. Significant cooling is also noticed in the best pseudo-simulation during this period off the eastern US and Japan coasts as well as in tropical areas (not shown). Nevertheless, the later results must be taken with caution because of the poor representation of tropical dynamics in the model.

Furthermore, the technique could be used to analyse the atmospheric and oceanic circulation during this cold period over Europe. The advantage of analysing the European area is that model results could be compared to the reconstruction of sea level pressure of Luterbacher et al. (2002). We must insist here that this reconstruction is not used to constrain model results so the comparison really provides a test of the quality of the technique on variables that are not directly constrained. Instead of sea level pressure, we have shown on Fig. 9 the geopotential at 800 hPa because it is the model dynamic variable that is the closest to sea level pressure. Fig. 9 shows that the agreement between the best pseudo-simulation and the reconstruction is quite good for this period. Both display a maximum of the anomaly close to Iceland and a minimum over Southern Europe, leading to a decrease in the intensity of the westerly winds over the Atlantic and Europe. Some local differences are also present but as the large-scale changes are reproduced in a very satisfactory way by the best pseudo-simulation, it is possible to use model results to analyse the atmospheric processes responsible for the temperature changes. In particular, the changes in atmospheric circulation are in good agreement with the hypothesis that the particularly cold conditions of the late 17th Century are related to a smaller inflow of warm air from the Atlantic towards Europe (e.g., Luterbacher et al. 2001; Shindell et al. 2001;2003).

No reconstruction of the ocean current during the late 17th Century is available to compare with model results. Nevertheless, it is instructive to note a clear decrease in the intensity of the Gulf Stream/North Atlantic Drift at surface. This results in a slower northward transport of warm oceanic water and suggest that both oceanic and atmospheric circulation play a role in the cold conditions observed during this period. Such a role of the oceanic circulation is in agreement with the recent results of van der Schrier and Barkmeijer (2005).

Analysing the best pseudo-simulation during the 1690-1700 period suggest that the technique described here could help in understanding the mechanisms responsible for the temperature changes during a particular period of the past. Nevertheless, we have chosen on purpose an example favourable to the method, in order to show its potential. The 1690-1700 period has a clear and strong signal (which is due to a large extent in the model to internal variability, see Goosse et al. 2005b) that is reasonably well reproduced by the model. On the other hand, during some periods, no robust changes could be obtained from the simulations, the experiments displaying low value of CF presenting different changes in atmospheric and/or oceanic circulation. As a consequence, model results could only be used to describe some reasonable circulation patterns that could lead to the changes noticed in the reconstruction. Only additional data could help in selecting among all the proposed hypothesis the one that is the more likely. In this framework, the technique proposed here could help in determining the optimal location of the new data, corresponding to regions where the different hypothesis leads to large differences.

7 Discussion and conclusions

We have proposed a method to select among a relatively large ensemble of simulations covering the last millennium those that are the closest to proxy temperature data. This procedure provides then a state of the climate system that is consistent with estimates of external forcing, the proxy-records and with model physics. The only inconsistency with model physics are the jumps that are obtained when grouping the different states selected in a pseudo-simulation. Unfortunately, this forbids the analysis of low frequency changes in the deep ocean but allows for studying changes in atmospheric and ocean surface circulation, as done for instance by van der Schrier and Barkmeijer (2005) for a 30 year period. Furthermore, the technique could be easily improved at a later stage to solve this problem (Collins 2003).

The proposed technique is very simple as it only requires the computation of a cost function that provides a quantitative estimate of the difference between the model results and the proxy-data. Nevertheless, the tests presented here demonstrate that the procedure allows for a generally favourable comparison between model and reconstructions, both in configurations where only a few data are available as well as in cases where a large amount of data is used to constrain model results.

This high level of skill in the method is achieved using a modest number of simulations, with 30 experiments being a reasonable lower limit. In the tests performed here, the ensemble size required to yield good agreement between model results and proxy-data appears to be largely independent of the time scale, the number of proxy series, or the cost function used. The technique is thus clearly practical for 3-D climate models such as the one used in this study.

Our main goal here was to test precisely the advantage and limitations of the technique, to prove that the technique could bring very useful information in different conditions on a wide variety of spatial and temporal scales and thus to encourage future use of this technique. Nevertheless, in the two examples proposed here, the proxies contain only a small number of degrees of freedom, because the number of proxies itself is small (sections 4 and 5) or because the focus is on a small area (Europe, section 6). In such a case, the probability to find a good analogue in a sample of relatively small size is good (van den Dool 1994). Besides, if we were using a large number of proxy records, covering the whole earth, the number of degrees of freedom would increase significantly. This could also be the case if we had selected additional proxies recording other variables than surface temperature such as precipitation, sea surface salinity, wind direction and or wind magnitude. In order to find an analogue in an affordable number of simulations, it would then likely be necessary to reduce the number of freedom of the system, for instance by spatial filtering or by focussing on a few leading empirical orthogonal functions (van den Dool 1994). This will be tested in the near future.

The required ensemble size could also be model-dependent as a higher resolution model could, for instance, contain more degrees of freedom. On the other hand, in our model, the response of the dynamical modes of variability to changes in forcing is generally modest, as these modes exhibit mainly internal, rather than externally-forced, variability. The observed changes in those modes of variability can thus only be reproduced through changes in model internal variability. By contrast, some models that incorporate a more sophisticated representation of certain atmospheric physical and chemical processes (e.g. Shindell et al. 2001) simulate a stronger response of the dynamical modes of variability to the forcing. If these responses are representative of true climate responses, fewer experiments would be required by such a model to get a good agreement between model and observation, as the forced response would already include a significant part of the changes in the dynamical modes.

Using a small number of proxies, the uncertainties could be very large in areas that are not included in the evaluation of the cost function. As a consequence, the best pseudo-simulation selected by the technique proposed here could not be used to provide a reliable estimate of the regional distribution of the changes. It can only provide various patterns that are all compatible with the available proxies. On the other hand, the technique could be used to propose a reconstruction at large scale that is complementary to the ones obtained using statistical methods, as illustrated here for hemispheric averages.

In regions where more data are available, the technique could be used to make some extrapolation to nearby areas not covered by the proxy-data, but this must be done with great care as the quality of the extrapolation depends on the location and the period investigated. In each future application of the technique, it will thus be necessary to test precisely the robustness of such an extrapolation. In addition, when enough data is available, the method could also provide information on variables that are not recorded on the proxies such as wind pattern or ocean current. This is very useful to study the processes or mechanisms responsible for the anomalies recorded in the proxies or conversely the impact of the temperature anomalies on other variables. Even if it is not possible to find robust mechanisms in the different simulations selected, the method will provide some indications of possible mechanisms that could be tested with more widespread proxy data networks.

Finally, as proxy-data are also influenced by non-climatic factors, if a dense network of proxy data is used to constrain model results, the method proposed here could be valid to isolate problematic proxy series. Indeed, the technique selects a state that is in reasonable agreement with all the neighbouring proxies and with model dynamics. As a consequence, if one or a small subset of proxy records suffer from non-climatic biases, the selected optimal state might be used to identify discrepancies of these records from the predicted state.

Acknowledgements.

We would like to thank J. Luterbacher for sending us his reconstruction of European temperatures. M. Collins and two anonymous referees made very useful comments on an earlier version of this manuscript. H. Goosse is Research Associate with the Fonds National de la Recherche Scientifique (Belgium). H. Renssen is sponsored by the Netherlands Organization for Scientific Research (N.W.O). A. Timmermann is supported by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) through its sponsorship of the International Pacific Research Center. M.E. Mann was supported by the NOAA- AND NSF-supported "Earth Systems History" program. This study was carried out as part of the Second Multiannual Scientific Support Plan for a Sustainable Development Policy (Belgian Federal Science Policy Office, contracts EV/10/7D and EV/10/9A) and the Action Concertée Incitative Changement Climatique (project Changement Climatique et Cryosphère) from the French Ministry of Research.

References

Bard E, Raisbeck G, Yiou F, Jouzel J (2000) Solar irradiance during the last 1200 years based on cosmogenic nuclides. Tellus 52B: 985-992

Bauer E, Claussen M, Brovkin V, Huenerbein A (2003) Assessing climate forcings of the Earth system for the past millennium. Geophys Res Let 30(6): 1276 doi 10:1029/2002GL016639

Berntsen T, Wyhre G, Stordal, F, Isaksen, I (2000) Time evolution of tropospheric ozone and its radiative forcing J Geophys Res, 105: 8915-8930

Berger A (1978) Long-term variations of daily insolation and Quaternary climatic changes. J Atmos Sci 35: 2363-2367

Bertrand C, Loutre MF, Crucifix M, Berger A (2002) Climate of the last millennium: a sensitivity study. Tellus 54A: 221-244

Bradley RS (1999) Paleoclimatology, reconstructing climates of the quaternary. International Geophysics Setries, 64 (2nd edition), Academic Press 613 pp.

Briffa KR, Jones PD, Bartholin TS, Eckstein D, Schweingruber FH, Karlèn W, Zetterberg P, Eroren M (1992a) Fennoscandian summers from AD 500: temperature changes on short and long timescales. Clim Dyn 7: 111-119

Briffa KR, Jones PD, Schweingruber FH (1992b) Tree-ring density reconstruction of summer temperature patterns across western North America since 1600. J Clim 5: 735-752

Briffa KR, Jones PD, Schweingruber FH, Shiyatov G, Cook EF (1995) Unusual twentieth-century summer warmth in a 1,000-year temperature record from Siberia. Nature 376: 156-159

Briffa KR (2000) Annual variability in the Holocene: interpreting the message of ancient trees. Quat Sci Rev 19, 87-105

Brovkin V, Bendtsen J, Claussen M, Ganopolski A, Kubatzki C, Petoukhov V, Andreev A (2002) Carbon cycle, vegetation and climate dynamics in the Holocene: experiments with the CLIMBER-2 model. Global Biogeochem Cycles 16: DOI: 10.1029/2001GB001662

Charlson RJ, Langner J, Rodhe H, Leovy CB, Warren SG (1991) Perturbation of the Northern Hemisphere radiative balance by backscattering from anthropogenic sulfate aerosols. Tellus 43 AB: 152-163

Collins M, Osborn TJ, Tett SFB, Briffa KR, Schweingruber FH (2002) A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities. J Clim 15: 1497-1515

Collins M (2003) Quantitative use of paleo-proxy data in global circulation models. Geophysical Research Abstracts 5: 0114

Cronin TM, Dwyer GS, Kamiya T, Schwede S, Willard DA (2003) Medieval Warm Period, Little Ice Age and 20th century temperature variability from Chesapeake Bay. Global and Planetary Change 36: 17-29

Crowley TJ (2000) Causes of climate change over the past 1000 years. Science 289, 270-277

Crowley TJ, Baum SK, Kim KY, Hegerl GC, Hyde WT (2003) Modeling ocean heat content changes during the last millennium. Geophys Res Let 30 (18): 1932

Cubasch U, Voss R, Hegerl GC, Waszkewitz J, Crowley TJ (1997) Simulation of the influence of solar radiation variations on the global climate with an ocean-atmosphere general circulation model. Clim Dyn 13: 757-767

D'Arrigo R, Jacoby G, Frank D, Pederson N, Cook E, Buckley B, Nachin B, Mijiddorj R, Dugarjav C (2001) 1738 years of Mongolian temperature variability inferred from a tree-ring width chronology of Siberian pine. Geophys Res Lett 28: 543-546

Fichefet T, Morales Maqueda MA (1997) Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. J Geophys Res102:12609-12646

Fisher DA, Koerner RM, Kuiviner K, Clausen HB, Johnsen, SJ, Steffensen JP, Gundestrup N, Hammer CU (1996) Inter-comparison of ice core and precipitation records from sites in Canada and Greenland over the last 3500 years and over the last few centuries in detail using EOF techniques. In “Climatic variations and forcing mechanisms of the last 200 years", NATO ASI Ser; I, vol.41, edited by P.D. Jones, R.S Bradley and J. Jouzel, pp. 297-328, Springer-Verlag, Berlin

Gerber S, Joos F, Brugger P, Stocker TF, Mann ME, Sitch S, Scholze M (2003) Constraining temperature variations over the last millennium by comparing simulated and observed atmospheric CO2. Clim Dyn 20: 281-299

Gonzalez-Rouco F, von Storch H, Zorita E (2003) Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years. Geophys Res Lett 30 (21): 2116

Goosse H, Fichefet T (1999) Importance of ice-ocean interactions for the global ocean circulation: a model study. J Geophys Res 104: 23,337-23,355.

Goosse H, Selten FM, Haarsma RJ, Opsteegh JD (2001) Decadal variability in high northern latitudes as simulated by an intermediate complexity climate model. Annals of Glaciology 33, 525-532.

Goosse H., Crowley T, Zorita E, Ammann C, Renssen H, Driesschaert E (2005a) Modelling the climate of the last millennium: what causes the differences between simulations? Geophys Res Let 32 L06710, doi:10.1029/2005GL022368

Goosse H, Renssen H, Timmermann A, Bradley RS (2005b) Internal and forced climate variability during the last millennium: A model-data comparison using ensemble simulations. Quat Science Rev 24, 1345-1360

Jacoby GC, D'Arrigo RD (1989) Reconstructed northern hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. Clim Change 14: 39-59

Jones J, Widmann M (2003) Reconstructing large-scale variability from paleoclimatic evidence by means of Data Assimilation Through Upscaling and Nudging (DATUN). In Fischer, H., Lohmann, G., Flser, G., Miller, H., von Storch, H. and Negendank J.F. (editors), The KIHZ project: towards a synthesis of Holocene proxy data and Climate models: 171-193. Springer, Berlin

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

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: 455-471

Jones PD, ME Mann (2004) Climate over past millenia. Rev Geophys 42(2): RG2002, DOI 10.1029/2003RG000143

Kalnay E and XXI others (1996) The NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77: 437-471

Lean J, Beer J, Bradley R (1995) Reconstruction of solar irradiance since 1610: implications for climate change. Geophys Res Lett 22: 1591-1594

Livezey RE, Barnston AG, Gruza GV, Ran'Kova EY (1994) Comparative skill of two analog seasonal temperature prediction systems: objective selection of predictors. J Climate 7: 608-615

Lorenz EN (1969) Atmospheric predictability as revealed by naturally occurring analogues. J Atmos Science 26: 636-646

Luckman BH, Briffa KR, Jones PD, Schweingruber FH (1997) Tree-ring based reconstruction of summer temperatures at the Comubia Icefield, Alberta, Canada, A.D. 1073-1983. The Holocene, 7: 375-389

Luterbacher J, Rickli R, Xoplaki E, Tinguely C, Beck C, Pfister C, Wanner H (2001) The Late Maunder Minimum (1675-1715)-A key period for studying decadal scale climatic change in Europe. Climatic Change 49(4): 441-462

Luterbacher J, Xoplaki E, Rickli R, Gyalistras D, Schmutz C, Wanner H (2002) Reconstruction of Sea Level Pressure fields over the eastern North Atlantic and Europe back to 1500. Clim Dyn 18: 545-561

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H (2004) European seasonal and annual temperature variability, trends, and extremes since 1500. Science 303: 1499-1503

Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 378: 266-270

Mann ME, Jones PD (2003) Global surface temperatures over the past two millennia. Geophys Res Lett 30(15): 1820, doi:10.1029/2003GL017814.

Mann ME, Cane MA, Zebiak SE, Clement A (2005a) Volcanic and solar forcing of the tropical Pacific over the past 1000 years. J Clim 18: 447-456

Mann ME, Rutherford S, Wahl E, Ammann C (2005b) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J Clim 18: 4097-4107

Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlén W (2005) Highly varaible northern hemisphere temperatures reconstructed from low- and high-resolution proxy data. Nature 433: 613-617

Nicolis C (1998) Atmospheric analogs and recurrence time stsistics: toward a dynamical formulation. J Atmos Science 55:465-475

Opsteegh JD, Haarsma RJ, Selten FM, Kattenberg A (1998) ECBILT: A dynamic alternative to mixed boundary conditions in ocean models. Tellus 50A: 348-367

Ramankutty N, Foley JA (1999) Estimating historical changes in global land cover: croplands from 1700 to 1992. Glob Biogeoch Cycles 13(4):997-1027

Renssen H, Goosse H, Fichefet T (2002) Modeling the effect of freshwater pulses on the early Holocene climate: the influence of high-frequency climate variability. Paleoceanography 17: 1020, 10.1029/2001PA000649

Renssen H, Goosse H, Fichefet T, Brovkin V, Driesschaert E, Wolk F (2005) Simulating the Holocene climate evolution at northern high latitudes using a coupled atmosphere-sea ice-ocean-vegetation model. Clim Dyn 24: 23-43

Rind D, Lean J, Healy R (1999) Simulated time-dependent response to solar radiative forcing since 1600. J Geophys Res 104: 1973-1990

Robock A. (2000) Volcanic eruptions and climate. Rev Geophys 38: 191-219

Rossow WB, Walker AW, Beuschel DE, Roiter MD (1996) International satellite cloud climatology project (ISCCP) documentation of new cloud datasets. WMO/TD-No 737, World Meteorological Organisation

Selten FM., Haarsma RJ, Opsteegh JD (1999) On the mechanism of North Atlantic decadal variability. J Clim 12: 1956-1973

Shindell DT, Schmidt GA, Mann ME, Rind D, Waple A (2001) Solar forcing of regional climate change during the Maunder Minimum. Science 294: 2149-2152

Shindell DT, Schmidt GA, Miller R, Mann ME (2003) Volcanic and Solar forcing of Climate Change During the Pre-Industrial era. J Clim 16: 4094-4107

Shindell DT, Schmidt GA, Mann ME, Faluvegi G (2004) Dynamic winter climate response to large tropical volcanic eruptions since 1600. J Geophys Res 109: D05104, doi: 10.1029/2003JD004151

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

Van den Dool HM (1994) Searching for analogues, how long must we wait? Tellus 46A 314-324.

van der Schrier G, Barkmeijer J (2005) Bjerknes' hypothesis on the coldness during 1790-1820 AD revisited. Clim Dyn 25 (5): 537-553

van Engelen AFV, Buisman J, Ijnsen F (2001) A millennium of weather, winds and water in the low countries, in "History and climate: memories of the future ?", edited by P.D. Jones et al. pp. 101-124, Kluwer Acad., Dordrecht, The Netherlands

Vinther BM, Johnsen SJ, Andersen KK, Clausen HB, Hansen AW (2003) NAO signal recorded in the stable isotopes of Greenland ice cores. Geophys Res Lett 30(7): 1387, doi:10.0129/2002GL016193

Waple AM, Mann ME, Bradley RS (2002) Long-term patterns of solar irradiance forcing in model experiments and proxy based surface temperature reconstructions. Clim Dyn 18: 563-578

Widmann M, Tett SFB (2003) Simulating the Climate of the last millennium. Pages News 11: 21-23

Xoplaki E, Luterbacher J, Paeth H, Dietrich D, Steiner N, Grosjean M, Wanner H (2005) European spring and autumn temperature variability and change of extremes over the last half millennium. Geophys Res Lett 32: L15713 (DOI:10.1029/2005GL023424

Yang B, Braeuning A, Johnson KR, Yafeng S (2002) General characteristics of temperature variations in China during the last two millennia. Geophys Res Let 29(9) 1324: doi:10.1029/2001GL014485

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Climate 12: 2474-2489

Appendix 1. Influence of the choice of the cost function.

In the main text, in the computation of the cost function (hereafter referred to as CF1), all the weights w_i have the same value (see section 4). Here, we study the influence of the choice of the cost function by selecting different w_i. In the first additional CF (CF2), w_i is proportional to the correlation between the proxy-based reconstruction and instrumental temperature data (Table 2). The goal is to give more weight in the evaluation of the cost function to the proxy record that provides a better reconstruction of the temperature. Finally, in the second additional CF (CF3), the data density is taken into account in a simple way: the weight of a record is proportional to its correlation with instrumental data as for CF2 but, in addition, is divided by the number of records in the same region. As WUM, WUB, JTL (Table 2) all provide information on the Western United states and Canada, their weight is divided by 3 while the weight of WGF and WGV (Western Greenland) is divided by two.

As expected, the correlations using CF2 are generally lower than with CF1 for the records that exhibit a low correlation with instrumental data (e.g. CSC), while they are greater for the records that exhibit a high correlation with instrumental data (e.g. PUB and LCV) (Table A1). For CF3, the correlations tend to be lower for the records located in data-rich regions than with CF2 (e.g. WUM, JTL) while the other correlations do not change significantly.

The convergence of the method is similar for the three cost function tested. Nevertheless, Fig A1 shows that using CF2, the value of the cost function is lower than using CF1. This means that the model is generally in closer agreement with proxy records at locations where the proxy has itself a good correlation with instrumental data and shows less good agreement at locations where this correlation is lower (e.g. CSC or MOD-see table 3). Nevertheless, this is not universally true. For example, the agreement between model results and proxy records is good for record CHY though the correlation with instrumental data is low. On the other hand, CF3 yields nearly the same correlation as CF2. As a consequence, taking into account the density of the proxies for the different regions does not seem to have an important influence on the value of the cost function in the present framework.

Appendix 2. Testing the technique using synthetic time series.

A lack of consistency between the proxy records and the best pseudo-simulation selected by the technique proposed here could be due to different factors. First, this could be due to a too small number of simulations that does not allow finding a good analogue to the real evolution of the climate system. Second, this could be related to model deficiencies, the model not being able to simulate correctly some processes at large-scale or at the scale recorded by the proxy. Third, the uncertainties on the forcing evolution are quite large with a potential impact on the simulated results. Fourth, proxies do not only record changes in climatic conditions, they are also affected by non-climatic factors. The interpretation of proxy-record is also sometimes very difficult (e.g. Bradley 1999, Briffa 2000, Jones and Mann 2004). Among all those possible sources of discrepancy between model and observations, only the first one is related to the technique itself.

In order to make tests focussed only on the technique, we have made one additional simulation with ECBILT-CLIO-VECODE driven by both natural and anthropogenic forcings as in the ensemble of 105 simulation analysed here. From this simulation, we have extracted 12 time series at the same locations as the ones used in section 4. Using, those 12 time series as pseudo-proxies, we have then tried to find the simulations among the ensemble of 105 which were the closest to the additional experiment. In this case, all the differences between the pseudo-proxies and the best pseudo-simulation will be related to the technique itself. Model errors would have no impact, as the same model is applied in all the simulations, the same forcing is used in the ensemble and in the additional simulation and the pseudo-proxies are perfect indicators of model temperature.

Table A2 provides results that are very similar to table 3 when real proxies were used. The correlation is better for pseudo-proxies as expected but the difference is not very large. When using the pseudo-proxies, the cost function for 12 data using a 25-year averaging period, has a value of 0.60, while it was 0.65 for real proxies. When the number of proxies used to constrain the choice of the best simulation is reduced, the correlation between pseudo-proxies and the best pseudo-simulation is lower at locations not used in the computation of CF. For some locations, the decrease is small. For instance, the correlation between the pseudo-proxy record for WUB and the best pseudo-simulation using 8 data (2 in North America) is 0.70. This contrasts with the real proxy for which the correlation drops to 0.33 in the same type of experiment. Apparently, the constraint in the model results is still high enough for the Northern America when using 2 pseudo-proxies. In particular, the area corresponding to WUM and WUB are partly overlapping (Jones and Mann 2004). On the other hand, when using real proxies, we must take into account that the correlation between the proxy records WUB and WUM is quite weak. In such a case, constraining model results using WUM is thus of little help in obtaining a good correlation between WUB and the best pseudo-simulation. For some other locations, like WGV, the constraint in nearly the same area but for a different season (WGF) is not sufficient to give a good correlation between the pseudo-proxy records and the best pseudo-simulation (Table A2). For TOB, when using 4 proxies, the correlation is even smaller than 0.20. On average, the correlations using 4 proxies are not very high at locations not used in the computation of CF, but still higher than using the real proxies.

We have also performed the same type of test, but this time using a higher density of proxies in a particular region (Table A3). To do so, in addition to the pseudo-proxy records used above, we have selected 8 additional pseudo-proxies: Low Countries in summer (LCS) and in winter (LCW) (in order to avoid duplicate information LCV is not used anymore), Western Russia in summer (RUS) and Winter (RUW), Switzerland in summer (SSB) and in winter (SWM), Eastern France in summer (BSC), Czech lands in summer (CLS) and winter (CLW).

Again, in some case, the correlation between the pseudo-proxy and the best simulation can be very good for some pseudo-proxies that are not used in the computation of CF. For instance, the correlation for Eastern France (BSC) is very high, the information form Switzerland (SSB, SWM), an adjacent grid box compare to the one corresponding to eastern France in the model, providing a strong constraint on the choice of the best simulation. On the other hand, for TOB, the correlation is very low as soon as this record is not used in the computation of CF, the pseudo proxies in the nearby Western Russia being of little help.

Those experiments using pseudo-proxies show thus results similar to the ones obtained using real proxies. Extrapolating toward regions where no proxy is available should thus be performed with great care. In some cases, the information contained in nearby location could be sufficient but it is not a general rule. It is necessary to test in each case if this extrapolation is reasonable, providing robust results in the model world, as for instance done in section 6 (Figure 8). The potential inadequacies in the forcing, model formulation or in the interpretation of the proxies provide additional source of trouble in this extrapolation.

On the other hand, for annual mean hemispheric temperature, the correlation between the best pseudo-simulation and the one of the additional simulation is higher than 0.9, if more than 8 pseudo-proxy records are used. For an average over Europe, the correlation between the best pseudo-simulation and the additional simulation is higher than 0.80 for annual, summer and winter mean temperatures, except in the cases where only 4 pseudo-proxy records are used in Europe. This shows that using the best pseudo-simulation to make an average over a relatively large region is a much safer procedure than the extrapolation to region where no data is available.

Figures Caption.

Figure 1. Time evolution of the normalised temperature anomalies in 12 regions in the extra-tropical Northern Hemisphere. The reference period covers the years 1600-1950. The times series plotted are 10-year averages. The black line corresponds to the mean over the 105 simulations while the grey lines are the ensemble mean plus and minus two standard deviation. The blue line is the pseudo-time series derived by the succession of the states that produce the minimum of the cost function (i.e., the best pseudo-simulation) obtained by using all 12 proxies of table 2. The red lines are the proxy reconstruction for the same region (see Table 2 for the list of references).

Figure 2. Value of the cost function as a function of time using a 25-year time-scale for the best pseudo-simulation using 12 proxy records (black) and for a particular, typical simulation (red).

Figure 3. Evolution of the cost function averaged over the last 1000 years as a function of the number of experiments (a) for different averaging periods using 12 proxy records, (b) as a function of the number of experiments when using 12, 8 or 4 proxy data series in the computation of CF for 25-year averages and (c) as a function of the number of experiments for the experiments of group K, C, D and a blend of all the groups (see table 1).

Figure 4. (a) Time evolution of the normalised temperature anomaly averaged over the Northern Hemisphere. The times series plotted are 10-year averages. The black line corresponds to the mean over the 105 simulations while the grey lines are the ensemble mean plus and minus two standard deviations. The blue line corresponds to best pseudo-simulation while the red line is the reconstruction of Mann and Jones (2003). (b) Time evolution of the anomaly of annual mean temperature (in Kelvin) averaged over the Northern Hemisphere for the best pseudo-simulation (left vertical axis). The red line is the reconstruction of Mann and Jones (2003) (left vertical axis) and the light blue one the reconstruction of Moberg et al. (2005) (right vertical axis).

Figure 5. Correlation over the years 1500-2000 AD between the best pseudo-simulation and the reconstruction of Luterbacher et al. (2004) in winter (DJF, top) and in summer (JJA, bottom) for a 10-year averaging period.

Figure 6. Time series of the normalised temperature anomaly for a 10-year averaging period in Europe in the best pseudo-simulation (blue) and in the reconstruction of Luterbacher et al. (2004) (red) in winter (DJF, top) and in summer (JJA, bottom).

Figure 7. Time series of the normalised temperature anomaly in winter for a 1-year averaging period in Europe in the best pseudo-simulation (blue) and in the reconstruction of Luterbacher et al. (2004) (red). On the bottom panel, a 21-year running mean has been applied to the time series.

Figure 8. Normalised temperature anomaly in winter (DJF) during the period 1690-1700 AD in the reconstruction of Luterbacher et al. (2004) (top) and for the best pseudo-simulation (bottom). Not significant model results are masked (see text for more details).

Figure 9. Anomaly of winter (DJF) sea level pressure during the period 1690-1700 AD in the reconstruction of Luterbacher et al. (2002) (in hp, top), anomaly of 800 hp geopotential height (in dam, middle) and anomaly in surface ocean current (bottom) in the best pseudo-simulation.

Figure A1. Evolution of the cost function using 12 proxy records averaged over the last 1000 years as a function of the number of experiments for different cost functions.

Table 1: Description of the experiments.

Number of experiments	Symbol of the group	Starting date	Forcing Solar	 Volcanic
25	K	1000 AD	Lean et al (1995) / Bard et al.[1] (2000)	Crowley (2000)
25	C	1000 AD	Crowley (2000)	Crowley (2000)
25	D	1 AD	Crowley[2]-2004	Crowley[2]-2004
15	B	850 AD	Lean et al (1995) / Bard et al. [1] (2000)	Crowley[2]-2004
15	M	850 AD	Bard et al. [1] (2000)	Amman[3]-2004

[1] We are using the reconstruction of Bard et al. (2000) scaled to match the Maunder Minimum irradiance reduction derived by Lean et al. (1995).

[2] updated from Crowley et al (2003).

[3] As described in Jones and Mann (2004).

Table 2. Proxy series used in the model-data comparison at hemispheric scale (modified from Jones and Mann 2004).

Location/site	Abbreviation	Reference	Decadal correlation with instrumental data
Western North America			
Northern tree line [a]	NTJ	Jacoby et D'Arrigo (1989)	0.71
Western United States [a]	WUM	Mann et al (1998)	0.61
Western United States [b]	WUB	Briffa et al. (1992b)	0.66
Jasper [b]	JTL	Luckman et al. (1997)	0.49
North Atlantic			
Chesapeake Bay [a]	CSC	Cronin et al. (2003)	0.32
Greenland [a]	WGF	Fisher et al. (1996)	0.75
Greenland [c]	WGV	Vinther et al. (2003)	0.78
Europe			
Tornetrask (Fennoscandia) [b]	TOB	Briffa et al. (1992a)	0.54
Polar Urals [b]	PUB	Briffa et al. (1995)	0.85
Low Countries [a]	LCV	van Engelen et al. (2001)	0.83
Eastern Asia			
Mongolia [a]	MOD	D'Arrigo et al. (2001)	0.40
China [a]	CHY	Yang et al. (2002)	0.22

[a] correspond to a reconstruction of annual mean temperature
[b] correspond to a reconstruction of summer temperature
[c] correspond to a reconstruction of winter temperature.

Table 3. Correlation between model simulations and proxy–reconstruction for a 25 year averaging period. The best pseudo-simulation is obtained using 12, 8 and 4 proxy records in the computation of CF for column 4,5 and 6, respectively. The correlation for the locations that are used to compute the CF are in bold.

Location (Abbreviation)	Mean over the experiments [a]	Ensemble mean	Best simulation using 12 data	Best simulation using 8 data	Best simulation using 4 data
NTJ	0.26	0.43	**0.82**	0.51	0.19
WUM	0.27	0.61	**0.58**	**0.59**	**0.91**
WUB	0.19	0.31	**0.71**	0.33	0.43
JTL	0.24	0.35	**0.78**	**0.76**	0.22
CSC	-0.07	-0.17	**0.64**	**0.62**	-0.01
WGF	0.14	0.23	**0.47**	**0.63**	**0.88**
WGV	-0.02	-0.06	**0.61**	-0.01	0.29
TOB	0.24	0.51	**0.71**	0.26	0.20
PUB	0.08	0.29	**0.76**	**0.83**	-0.02
LCV	0.53	0.68	**0.71**	**0.83**	**0.94**
MOD	0.19	0.33	**0.69**	**0.67**	**0.92**
CHY	0.62	0.75	**0.80**	**0.85**	0.60

[a] the correlation is first performed for all the simulations and then the average over all the correlation is computed

Table A1. Correlation between model simulations and proxy–reconstruction for a 25 year averaging period.

Location (Abbreviation)	Best simulation using CF1	Best simulation using CF2	Best simulation using CF3
NTJ	0.82	0.85	0.86
WUM	0.58	0.81	0.51
WUB	0.71	0.79	0.80
JTL	0.78	0.61	0.41
CSC	0.64	0.28	0.52
WGF	0.47	0.59	0.63
WGV	0.61	0.76	0.51
TOB	0.71	0.70	0.77
PUB	0.76	0.80	0.86
LCV	0.71	0.86	0.84
MOD	0.69	0.66	0.73
CHY	0.80	0.65	0.80

Table A2. Correlation between the 105 model simulations and an additional independent simulation performed with the model for a 25 year averaging period. The best pseudo-simulation is obtained using 12, 8 and 4 pseudo-proxy records deduced from the additional simulation in the computation of CF for column 4,5 and 6, respectively. The correlations for the locations that are used to compute the CF are in bold.

Location (Abbreviation)	Mean over the experiments	Ensemble mean	Best simulation using 12 data	Best simulation using 8 data	Best simulation using 4 data
NTJ	0.48	0.69	**0.71**	0.52	0.70
WUM	0.25	0.49	**0.77**	**0.83**	**0.87**
WUB	0.38	0.66	**0.78**	0.70	0.55
JTL	0.57	0.77	**0.85**	**0.89**	0.52
CSC	0.31	0.61	**0.86**	**0.87**	0.42
WGF	0.36	0.53	**0.83**	**0.79**	**0.85**
WGV	0.09	0.18	**0.73**	0.40	0.41
TOB	0.32	0.64	**0.78**	0.54	0.19
PUB	0.17	0.51	**0.68**	**0.80**	0.27
LCV	0.65	0.87	**0.93**	**0.92**	**0.97**
MOD	0.31	0.57	**0.83**	**0.90**	**0.93**
CHY	0.66	0.81	**0.92**	**0.92**	0.82

Table A3. Correlation between the 105 model simulations and an additional independent simulation performed with the model for a 25 year averaging period. The results are similar to table A2, except that 8 additional pseudo-proxies have been selected in Europe. The best pseudo-simulation is obtained using 20, 16 and 12 pseudo-proxy records deduced from the additional simulation in the computation of CF for column 1,2 and 3, respectively. The correlations for the locations that are used to compute the CF are in bold. Only the locations in Europe are displayed. Locations outside Europe were all used in the computation of CF, in each case.

Location (Abbreviation)	20 data	16 data	12 data
LCS	**0.82**	0.65	0.61
LCW	**0.80**	0.61	0.65
RUS	**0.89**	**0.83**	0.53
RUW	**0.73**	**0.75**	0.27
SSB	**0.87**	**0.82**	**0.83**
SWM	**0.82**	**0.83**	**0.85**
TOB	**0.62**	0.25	0.20
BSC	**0.82**	0.78	0.79
CLW	**0.69**	**0.86**	**0.56**
CLS	**0.80**	**0.92**	**0.66**

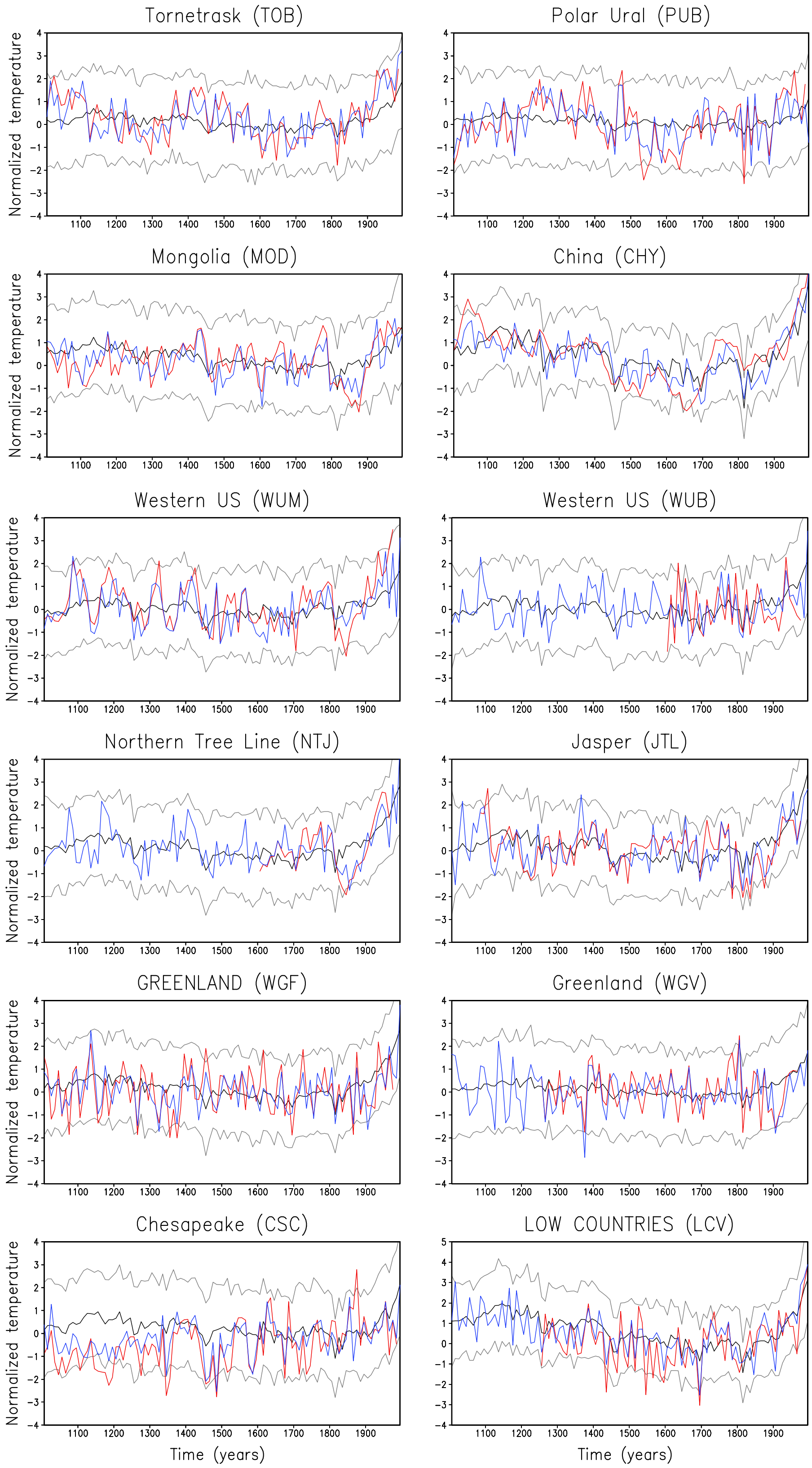
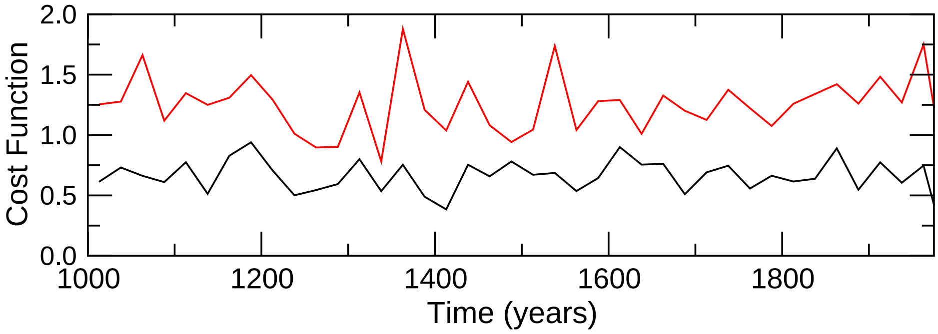

Tornetrask (TOB)
Polar Ural (PUB)
Mongolia (MOD)
China (CHY)
Western US (WUM)
Western US (WUB)
Northern Tree Line (NTJ)
Jasper (JTL)
GREENLAND (WGF)
Greenland (WGV)
Chesapeake (CSC)
LOW COUNTRIES (LCV)
Normalized temperature
Time (years)

Figure 1



2.0
1.5
1.0
0.5
0.0
Cost Function
1000
1200
1400
1600
1800
Time (years)

Figure 2

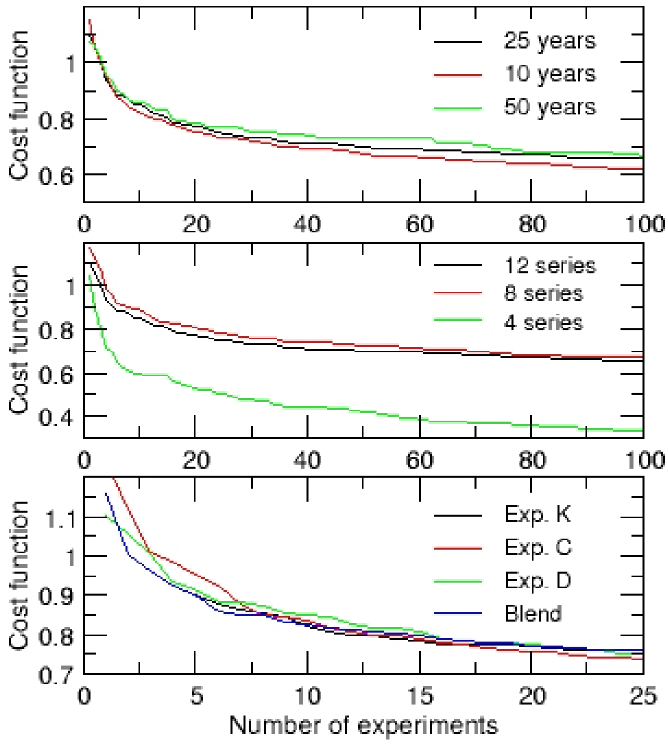

25 years
10 years
50 years
Cost function
12 series
8 series
4 series
Exp. K
Exp. C
Exp. D
Blend
Number of experiments

Figure 3

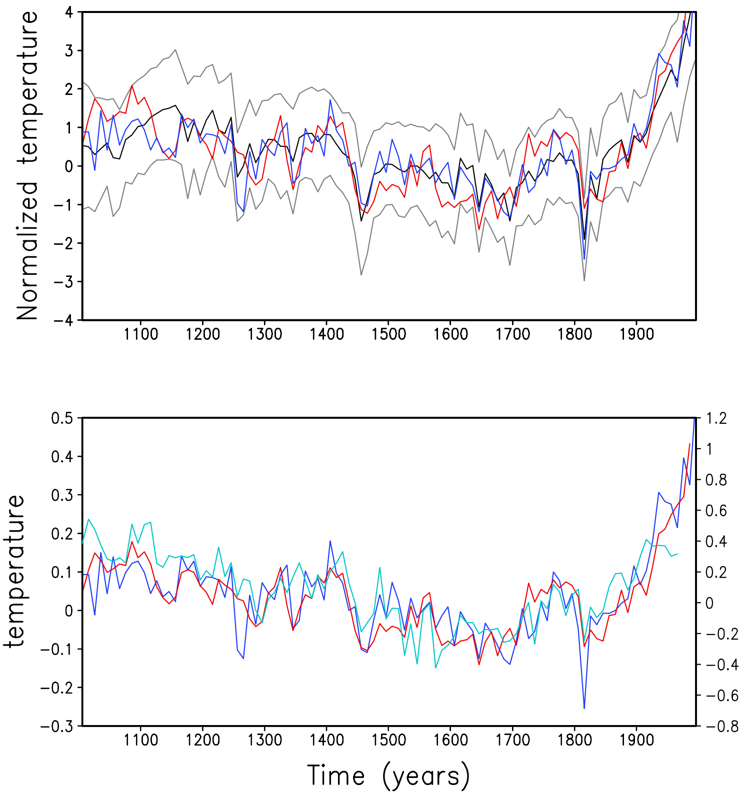


Figure 4

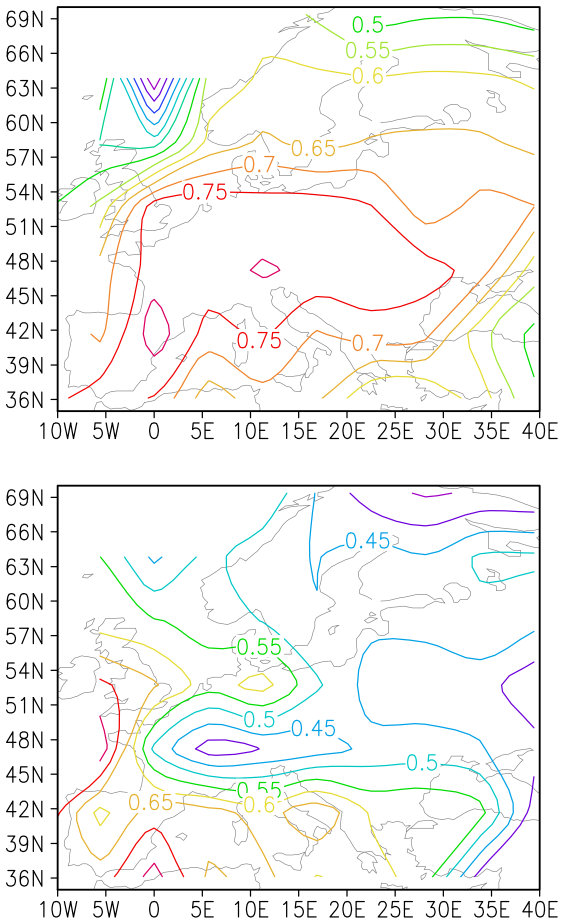

69N
66N
63N
60N
57N
54N
51N
48N
45N
42N
39N
36N
10W 5W 0 5E 10E 15E 20E 25E 30E 35E 40E
0.5
0.55
0.6
0.65
0.7
0.75
0.75
0.7
69N
66N
63N
60N
57N
54N
51N
48N
45N
42N
39N
36N
10W 5W 0 5E 10E 15E 20E 25E 30E 35E 40E
0.45
0.55
0.5
0.45
0.5
0.55
0.65
0.6

Figure 5

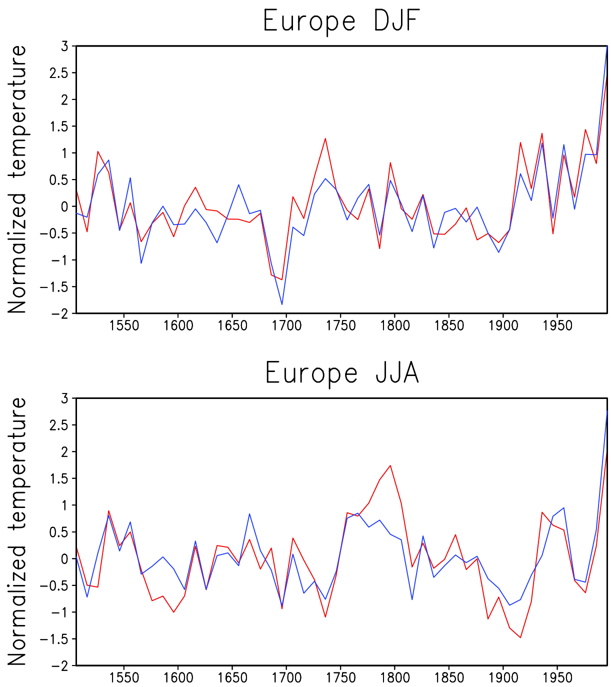
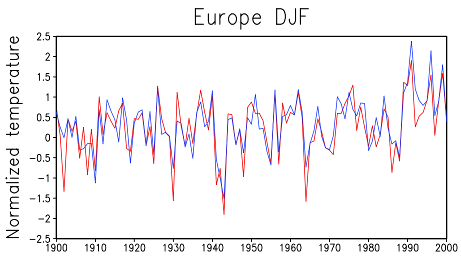

Europe DJF
Normalized temperature
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1550
1600
1650
1700
1750
1800
1850
1900
1950
Europe JJA
Normalized temperature
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1550
1600
1650
1700
1750
1800
1850
1900
1950

Figure 6


Europe DJF
Normalized temperature
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000

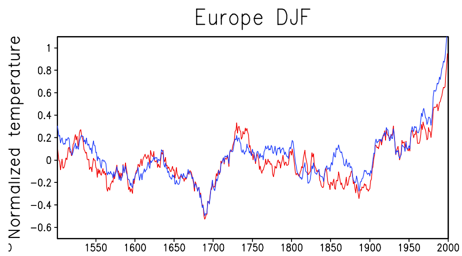

Europe DJF
Normalized temperature
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1550
1600
1650
1700
1750
1800
1850
1900
1950
2000

Figure 7

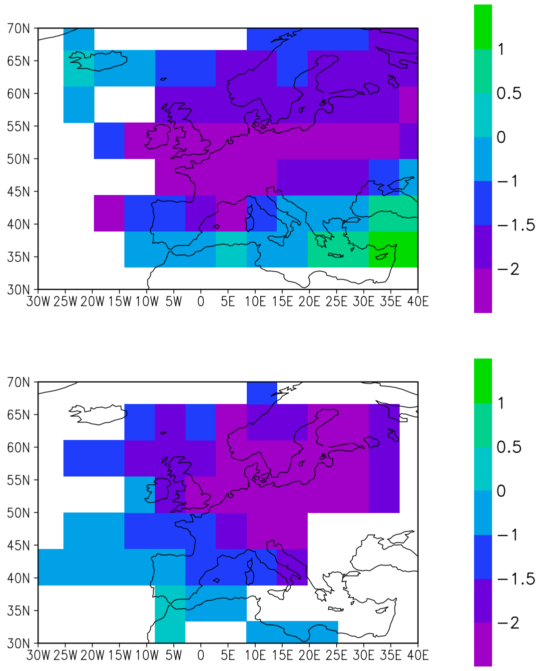

70N
65N
60N
55N
50N
45N
40N
35N
30N
30W 25W 20W 15W 10W 5W 0 5E 10E 15E 20E 25E 30E 35E 40E
1
0.5
0
−1
−1.5
−2
70N
65N
60N
55N
50N
45N
40N
35N
30N
30W 25W 20W 15W 10W 5W 0 5E 10E 15E 20E 25E 30E 35E 40E
1
0.5
0
−1
−1.5
−2

Figure 8

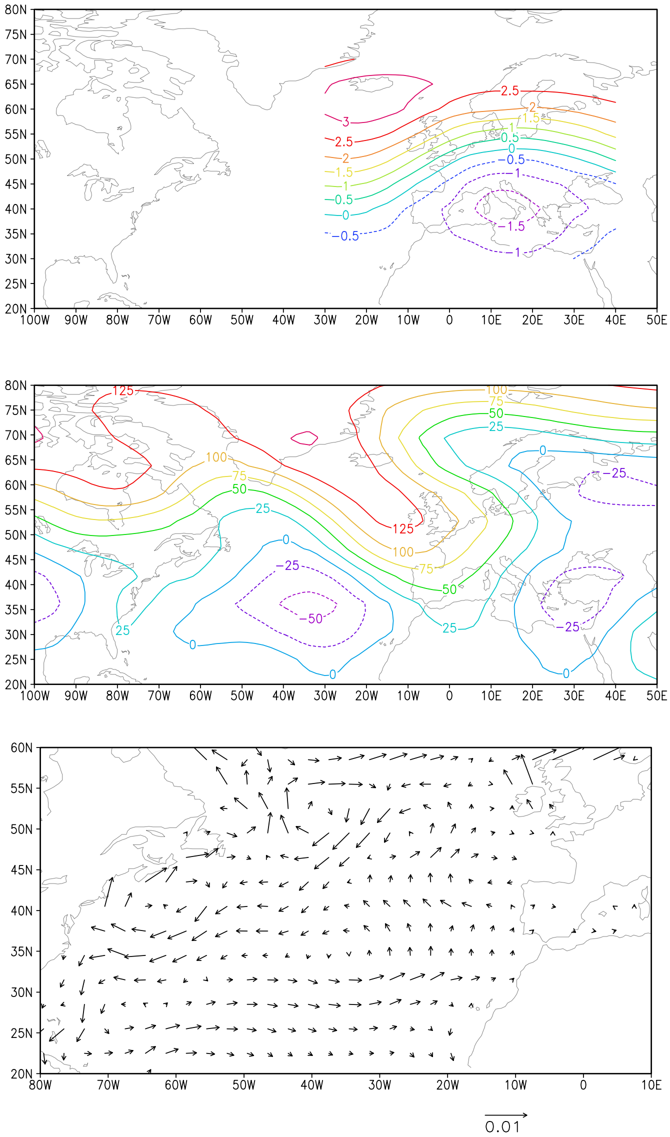

80N
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
20N
100W
90W
80W
70W
60W
50W
40W
30W
20W
10W
0
10E
20E
30E
40E
50E
60N
55N
50N
45N
40N
35N
30N
25N
20N
80W
70W
60W
50W
40W
30W
20W
10W
0
10E
0.01

Figure 9

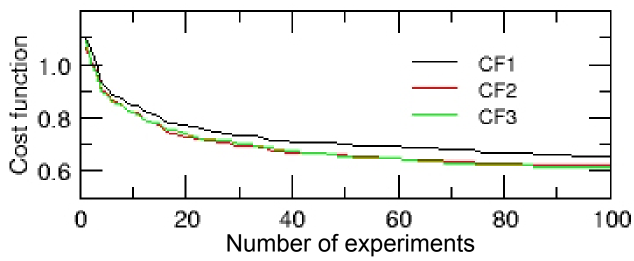

Cost function
1.0
0.8
0.6
0
20
40
60
80
100
Number of experiments
CF1
CF2
CF3

Figure A1.