Climatic Change
DOI 10.1007/s10584-010-9824-7

European temperature records of the past five centuries based on documentary/instrumental information compared to climate simulations

Eduardo Zorita · Anders Moberg · Lotta Leijonhufvud · Rob Wilson · Rudolf Brázdil · Petr Dobrovolný · Jürg Luterbacher · Reinhard Böhm · Christian Pfister · Dirk Riemann · Rüdiger Glaser · Johan Söderberg · Fidel González-Rouco

Received: 17 December 2009 / Accepted: 30 December 2009

Abstract Two European temperature reconstructions for the past half-millennium, January-to-April air temperature for Stockholm (Sweden) and seasonal temperature for a Central European region, both derived from the analysis of documentary sources and long instrumental records, are compared with the output of climate

---

E. Zorita (✉)
GKSS Research Centre, Geesthacht, Germany
e-mail: eduardo.zorita@gkss.de

A. Moberg · L. Leijonhufvud
Department of Physical Geography and Quaternary Geology, Stockholm University,
Stockholm, Sweden

A. Moberg
e-mail: anders.moberg@natgeo.su.se

L. Leijonhufvud
e-mail: lotta.leijonhufvud@ekohist.su.se

R. Wilson
School of Geography & Geosciences, University of St Andrews, St Andrews, Scotland
e-mail: rjsw@st-andrews.ac.uk

R. Brázdil · P. Dobrovolný
Institute of Geography, Masaryk University, Brno, Czech Republic

R. Brázdil
e-mail: brazdil@sci.muni.cz

P. Dobrovolný
e-mail: dobro@sci.muni.cz

J. Luterbacher
Department of Geography; Climatology, Climate Dynamics and Climate Change,
Justus-Liebig University, Giessen, Germany
e-mail: juerg@geogr.uni-giessen.de

simulations with the model ECHO-G. The analysis is complemented by comparisons with the long (early)-instrumental record of Central England Temperature (CET). Both approaches to study past climates (simulations and reconstructions) are burdened with uncertainties. The main objective of this comparative analysis is to identify robust features and weaknesses in each method which may help to improve models and reconstruction methods. The results indicate a general agreement between simulations obtained with temporally changing external forcings and the reconstructed Stockholm and CET records for the multi-centennial temperature trend over the recent centuries, which is not reproduced in a control simulation. This trend is likely due to the long-term change in external forcing. Additionally, the Stockholm reconstruction and the CET record also show a clear multi-decadal warm episode peaking around AD 1730, which is absent in the simulations. Neither the reconstruction uncertainties nor the model internal climate variability can easily explain this difference. Regarding the interannual variability, the Stockholm series displays, in some periods, higher amplitudes than the simulations but these differences are within the statistical uncertainty and further decrease if output from a regional model driven by the global model is used. The long-term trend of the Central European temperature series agrees less well with the simulations. The reconstructed temperature displays, for all seasons, a smaller difference between the present climate and past centuries than is seen in the simulations. Possible reasons for these differences may be related to a limitation of the traditional 'indexing' technique for converting documentary evidence to temperature values to capture long-term climate changes, because the documents often reflect temperatures relative to the contemporary authors' own perception of what constituted 'normal' conditions. By contrast, the amplitude of the simulated and reconstructed inter-annual variability agrees rather well.

R. Böhm
Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
e-mail: reinhard.boehm@zamg.ac.at

C. Pfister
Oeschger Centre for Climate Change Research and Institute of History,
University of Bern, Bern, Switzerland
e-mail: pfister@hist.unibe.ch

D. Riemann · R. Glaser
Institute of Physical Geography, University of Freiburg, Freiburg, Germany

D. Riemann
e-mail: dirk.riemann@geographie.uni-freiburg.de

R. Glaser
e-mail: ruediger.glaser@geographie.uni-freiburg.de

J. Söderberg
Department of Economic History, Stockholm University, Stockholm, Sweden
e-mail: Johan.Soderberg@ekohist.su.se

F. González-Rouco
Universidad Complutense de Madrid, Madrid, Spain
e-mail: fidelgr@fis.ucm.es

1 Introduction

Global climate models are important tools to estimate future climate change caused by changes in the radiation balance of the Earth linked to increasing concentrations of greenhouse gases. There has been considerable progress in the understanding of the relevant processes that give rise to low-frequency climate variability and of those that determine the sensitivity of the climate to external forcing. However, in spite of this progress, global climate models have difficulties in simulating the present climate, specifically at regional scales, and uncertainties still remain related to their ability to simulate climates different from those of the recent past (see Bader et al. 2008, for a review of the skill of current climate models). One test for climate models is to compare climate simulations with climate reconstructions, based on natural proxies or on documentary records, over the same periods. Most previous comparative studies have been devoted to the Holocene Optimum (e.g. Wohlfahrt et al. 2008; Wanner et al. 2008), or to the last glacial maximum (e.g. Ramstein et al. 2007). Such comparisons are useful for improving the understanding of climate dynamics under conditions that may be substantially different from those of today, in particular for the Last Glacial Maximum and they are also important to assess changes in the simulated climate sensitivity to changes in external forcings. However, the centennial-scale climate variability and climate sensitivity in those periods could have been different from the present.

In this paper we compare climate simulations with local-to-regional climate reconstructions for the last few centuries, i.e. a period when the external forcing and climate were much closer to the present state than in the Last Glacial Maximum, for instance. Over this period the number of comparative studies between simulations and reconstructions is rather limited (Jones and Mann 2004; Jones et al. 2009). Nevertheless, a variety of studies are available, of which some have focused on regional scales (e.g. Casty et al. 2005; Goosse et al. 2005, 2006; Raible et al. 2006; Wilson et al. 2006; Luterbacher et al. 2006, 2009, 2010; Gouirand 2007). Here, the comparison not only focuses on assessing the skill of the models in simulating the long-term trends, which are presumably to a large extent related to the long-term variation in the external forcing in these centuries, but also examines the amplitude of multi-decadal variations around the long-term trends. The amplitude of these variations, which are more related to internal variability, can indicate how strongly regional climates can deviate—and for how long—from long-term global warming or cooling trends. These variations thus provide some information on how regional climates might deviate over the coming decades from the anthropogenically-forced global warming trend. We also compare the amplitude of interannual temperature variations, which are even more strongly dominated by internal variability. The model simulations may also be useful for evaluating the proxy-based climate reconstructions, which may suffer from biases in low- as well as high-frequency variations. Such limitations might be illuminated when a set of climate reconstructions from nearby regions are compared with simulated data for corresponding areas.

In comparing the output of global simulations with reconstructions, one should not expect complete agreement at interannual timescales, even if the global climate model and the climate reconstructions are both perfect (Yoshimori et al. 2005). The evolution of the unforced variations, i.e. those variations not related to external forcing but caused by internal non-linear climate processes—in the simulations and in the observed past climate will not necessarily agree in spatio-temporal detail. For instance, since inter-annual climate variations are essentially due to internal random variability, a very low correlation should be expected between climate simulations and reconstructions at inter-annual timescales, possibly excluding the sudden temperature variations caused by volcanic eruptions (Fischer et al. 2007). At longer timescales internal random fluctuations are filtered out more effectively and the signal of the external forcing can be identified more easily. Those are the timescales at which model and reconstructions should agree. However, the appropriate degree of filtering required for the externally forced regional temperature signal to emerge from the internal fluctuations is not known.

Currently, the problem is compounded by the limited spatial resolution of global climate models that hinders a faithful representation of the relevant local processes, such as topography, coast lines, land-sea, land-use, etc., such that the disagreement between model and reconstruction can also be due to deficiencies in the model and not only to internal random variations. A potential solution is the use of regional climate models which have a more faithful representation of regional factors. A comparison between local or regional climate reconstructions and regional climate simulations would be more meaningful than with the output of global models. Unfortunately, the computational requirements of regional models are also large and multi-century-long regional simulations are only just beginning to be performed.

Bearing in mind the potential problems when comparing local climate reconstructions with the output of coarse-resolution global models, we undertake an initial joint assessment of model simulations with three long-term records derived from historical documentary evidence and long instrumental temperature observations. These records represent different aspects of European climate over the past centuries. The millennial simulations were performed with the global climate model ECHO-G (see e.g. Zorita et al. 2005, 2007). One of the temperature reconstructions is also compared with output from the regional climate model RCA3 (Kjellström et al. 2005) designed for the Scandinavian region. This regional model was driven by one of the global simulations with ECHO-G, and should represent the coastlines, topography, and ocean heat capacity of the actual region more realistically (Moberg et al. 2006). However, the regional ocean sub-model used in the regional simulation is simplified and does not represent heat advection by the Baltic Sea currents. Despite those limitations in the global and regional model comparisons, we expect to find robust features among all time series, as well as identifying sources of disagreement that could indicate further ways for improvement of reconstruction and modeling approaches.

This study is part of a European project (MILLENNIUM) specifically aimed at studying climate in Europe over the last millennium (Gagen et al. 2006). This paper is a contribution to a special issue devoted to documentary climate evidence, hence we restrict our model-proxy data comparison only to documentary data and long instrumental records. Documentary evidence has previously been used for developing a number of climate reconstructions for different parts of Europe (for more details see e.g. Brázdil et al. 2005), and has also been used, together with natural proxy data and long instrumental records, to derive climate reconstructions at the European continental scale (e.g. Luterbacher et al. 2004, 2007; Xoplaki et al. 2005; Pauling et al. 2006).

The first step in deriving a documentary based temperature reconstruction usually involves the compilation of a time series of temperature indices into an ordinal index scale (Brázdil et al. 2005). Such an index series can then be calibrated in terms of past temperatures based on statistical analysis of the temperature indices and observed temperatures in an overlapping period (Dobrovolný et al. 2009a). In addition, some types of documentary data (e.g. phenological data, starting dates of shipping seasons, etc.) directly reflect environmental conditions that are closely related to temperature. This kind of documentary data (Chuine et al. 2004; Guiot et al. 2005; Meier et al. 2007; Rutishauser et al. 2008; Leijonhufvud et al. 2009) allow a more direct statistical calibration to temperatures using approaches similar to the standard reconstruction methods applied to physical proxies, such as tree ring data. In our study we utilize one temperature reconstruction of each kind, one based on temperature indices for Central Europe and another based on continuous data related to the ice-free shipping season from Stockholm—which allows a comparison of the two approaches, and a third series, the Central England temperature series (Manley 1974) is almost exclusively based on instrumental observations although also it does include some information from documentary sources in its earliest part (before AD 1720). Details of the reconstructed/instrumental temperature series and the models are given in the following sections of this paper and the reader is referred to the companion papers in this special issue for more information about the reconstruction procedures (Leijonhufvud et al. 2009 for the Stockholm record and Dobrovolný et al. 2009b for the Central Europe record).

2 Simulated and reconstructed temperature series

The global climate model ECHO-G is an atmosphere-ocean coupled model that has been widely used, for instance, for estimations of future climate change by the Intergovernmental Panel on Climate Change (Min et al. 2005; Meeh et al. 2007). The atmosphere is represented by the atmosphere model ECHAM4, with a horizontal resolution of 3.75° × 3.75° and 19 levels covering the troposphere and the lower stratosphere. The ocean model is HOPE-G, with a horizontal resolution of about 2.8° × 2.8°, with an increasingly finer resolution towards the equator in the tropical regions. The ocean model contains 20 levels irregularly spaced at various ocean depths, with higher vertical resolution in the upper ocean layers. A flux correction is applied to the coupling of the atmosphere and ocean models to avoid climate drift. This flux correction is constant in time and has a zero global average (Legutke and Voss 1999).

The model ECHO-G was driven by estimations of external forcing in the past millennium taken from existing literature, as described in Zorita et al. (2005). The solar irradiance was derived from proxy-based estimations of solar radiative forcing taking into account the geometry of the Earth and its albedo. The changes in the solar irradiance between the AD 1960–1990 mean and the mean in the Late Maunder Minimum (AD 1680–1710) are, in these simulations, −0.3%. The short-wave volcanic forcing was derived from estimations based on acidity of ice layers in Greenland and Antarctica ice cores. This forcing was translated to an effective change of the solar constant in the model. This change takes place in the model along the whole year when an eruption occurs and is, for all eruptions, maximum in the northern summer. Estimates of atmospheric concentrations of carbon dioxide and methane were obtained from measurements of air bubbles in ice cores.

More details about the simulations of the past millennium with ECHO-G are given by Zorita et al. (2005, 2007) and González-Rouco et al. (2006). A discussion of the uncertainties in the external forcings and the difficulties of comparing the ECHO-G data with local proxy data are given by Gouirand et al. (2007).

Two simulations (denoted here ERIK1 and ERIK2) were carried out with ECHO-G and with the same external forcing over the past millennium. Obviously the initial conditions for these millennium simulations are unknown, so that a large ensemble of simulations would be required to assess the full range of possible temperature evolutions. However, the computer costs for performing a large ensemble of simulations over the whole millennium are prohibitive. Two simulations with different initial conditions deliver a reasonable approximation about the magnitude of internal model variability. The initial conditions used in year AD 1000 in the ERIK2 simulation were colder than in ERIK1 (see Zorita et al. 2007 for details). The simulated global and northern hemisphere mean temperature in both simulations show higher temperatures at the beginning and end of the millennium with colder temperatures in the central centuries, roughly between AD 1400 and 1800 (Zorita et al. 2005, 2007). From roughly AD1300, both simulations show similar multi-centennial variations. The simulated temperatures from ERIK1 and ERIK2 for the period AD 1500–1990 are used for comparison with the reconstructed/instrumental temperature series. For some particular aspects of this analysis, data from a control simulation also with the model ECHO-G were used. A 1,000-year control simulation was performed with constant greenhouse gas concentrations, a constant annual cycle of solar insolation and no volcanic eruptions.

For comparisons with the Stockholm data, we also use data from one of the first multi-centennial climate simulations performed with a regional climate model. Such data are available for the Scandinavian region for the period AD 1550–1929 (see Moberg et al. 2006 for details). The regional model is the atmospheric RCA3 (Kjellström et al. 2005). This model originates in the numerical weather prediction model HIRLAM and was used in the present study with a horizontal resolution of $1^\circ \times 1^\circ$ and 24 levels in the atmosphere. The regional model was driven at the boundaries of the domain by ECHO-G (using the ERIK1 simulation), and forced by the same external forcings as ECHO-G, with the exception of CH_4. The influence of this forcing on temperature is, however, communicated indirectly to the regional model by the lateral boundary conditions. To represent the Baltic Sea and the lakes in the domain of simulation, the model RCA3 was coupled to the lake model FLAKE (Mironov 2008). FLAKE takes as input radiative and heat fluxes from the atmosphere as well as solid precipitation. Calculations of sea (lake) surface temperatures and ice/snow conditions are performed locally for each grid box. The Baltic Sea flow dynamics are, however, not represented in this model. A fully three-dimensional model of the Baltic Sea flow would have substantially increased computer time requirements, and this was considered too demanding (Moberg et al. 2006). In case of both the global and the regional model, the temperatures simulated in the respective single grid-cell nearest to Stockholm was considered for this study.

It is well known that simulations with global climate models may lack realism at local scales, because of the poor representation of coastlines, land–sea interactions and topographical features. This is quite obvious in the case of the Baltic Sea, where the land connection to North Sea is two grid-cells wide, approximately 700 km. This may cause an unrealistically large advection of heat from the Atlantic Ocean into the Baltic Sea region and bias the simulation of mean temperatures along the Baltic Sea coast. Therefore only the temperature deviations from a reference state (but not the absolute temperatures) are considered. For the regional model, the land/sea mask is more realistic, but still too coarse to resolve local features such as lakes and the archipelago near Stockholm.

The Stockholm temperature reconstruction is a combination of temperature evidence from documentary data and long instrumental data (see Leijonhufvud et al. 2009 for details). The instrumental data are available from AD 1756 onwards and the documentary data from AD 1502 to 1892. The instrumental record is the homogenized series updated from Moberg et al. (2002). The documentary data consist of several sub-series of estimates of the start of the sailing season after each winter, which are normalized in a way that preserve low-frequency variations and combined into a dimensionless average time series (Leijonhufvud et al. 2009). The underlying assumption to use this series as a proxy for air temperature is as follows: the Baltic Sea as far north as Stockholm, including the archipelago, is usually ice-covered during the winter months, and the recorded start of the sailing season indicates when the water became ice-free after each winter. The start of the sailing season (i.e. the approximate date of ice break-up) is found to be strongly correlated with the mean temperature from January to April. This holds very well up to the late nineteenth century until ships became more capable of navigating in adverse conditions (Leijonhufvud et al. 2008). The Stockholm temperature series is developed as a January-to-April mean temperature reconstruction.

In this study two slightly different versions of the reconstructions are used: the first version, covering the period AD 1500–1892, is the result of a calibration of the proxy series to instrumental data by ordinary least squares in the period AD 1756–1892. After AD 1892, the reconstruction is spliced to the instrumental record from Stockholm. In the second version, we used a ‘variance matching approach’ (Esper et al. 2005). The reconstruction consists of the calibrated proxy data series in the period AD 1502–1892, after having adjusted its mean and variance to agree with the instrumental data over the full overlapping period (AD 1756–1892). From AD 1893 onward the reconstructed record is identical to the observations. The variance scaling is applied to avoid loss of variance caused by a regression calibration (Leijonhufvud et al. 2009, see also a more general discussion of variance scaling version regression calibration by Esper et al. 2005).

The Central Europe temperature reconstruction (henceforth CEurT), is a combination of monthly temperatures reconstructed from documentary temperature index data for the period AD 1500–1854 and long instrumental temperature measurements after that date (Dobrovolný et al. 2009b). Altogether 11 instrumental station records from Germany, Switzerland, Austria and the Czech Republic, were used to derive the CEurT areal average. This instrumental average series starts in AD 1760. All station series were homogenized and ten of them originate from the HISTALP database (see Auer et al. 2007 for technical details of the homogenization procedures). The homogenization includes, among other aspects, adjustment for urban warming trends and an adjustment for the early instrumental warm-bias problem (Frank et al. 2007a, b; Böhm et al. 2009). The temperature index data were derived from historical documentary evidence from the same countries, except Austria. First, monthly, seasonal and annual index series were derived for each country, and then averages were formed to obtain a regional series of dimensionless temperature proxy data for the period AD 1500–1854. None of the three national series, however, is complete over the entire period and, therefore, the number of series contributing to the average series varies in time. This introduces inhomogeneities in the variance of the average series that should be corrected. An adjustment was applied using standard methods used in dendroclimatology (Osborn et al. 1997; Frank et al. 2007a, b), where similar problems are usually encountered. Next, the variance-adjusted monthly, seasonal and annual series were calibrated (AD 1771–1816) and verified (AD 1760–1770 and AD 1817–1854) against the homogenized instrumental temperatures (for details see Dobrovolný et al. 2009b). Here, similar to the Stockholm reconstruction, we use two versions of the CEurT reconstruction. In both versions, the proxy data are used for the period AD 1500–1854 and instrumental data from AD 1855 onwards. The only difference is that in one version, the proxy data are calibrated by means of linear regression (AD 1771–1816) and in the other case the same data are adjusted to have the same mean and variance as the instrumental data over AD 1760–1854. By using both variants (in the case of both CEurT and Stockholm) we are able to compare the result of regression versus variance scaling for the calibration with the simulated temperatures (Esper et al. 2005).

The third temperature record is the well known Central England monthly temperature record (CET), starting in AD 1659. This series is exclusively based on instrumental and early-instrumental observations from a set of stations in the Central England region back to AD 1723. Before this year the CET is based on a combination of various short early instrumental observations for sites closer to London, and temperature observations from De Bilt in the Netherlands (AD 1707–1722) as well as a variety of historical documentary evidence (Manley 1953, 1974; Parker et al. 1992; Parker and Horton 2005). Manley (1974) clearly pointed out that data from before around AD 1720 must be considered as less reliable. As discussed by Jones (1999), Manley provided only vague information about how the temperature estimates before AD 1723 were derived. It appears that the available information together with a good understanding of the meteorological and climatological conditions of the area were used to derive an ‘educated best guess’ temperature estimate for each month. For the first decade, these estimates are provided without decimals in the Celsius scale. Hence, the very earliest part of the CET series has some similarities to an ordinal temperature index scale being calibrated to temperatures. The monthly temperatures in the following five decades are mostly given at a precision of 0.5 K, which is closer to the later standard precision of 0.1 K but nevertheless less precise. Despite this uncertainty in the early part of the CET record, comparison with temperature estimates derived from many boreholes on the British Isles show a remarkably good agreement of the long-term evolution of annual mean temperatures over the last three-and-a-half centuries (Jones 1999).

3 Modeled and reconstructed Stockholm winter–spring temperature series

The evolution of regional temperatures may include a considerable amount of random unforced internal variations, which are different across model simulations and also different from the one present in observed temperatures. The internal variations dominate the short term variability. The externally forced variations, theoretically more prominent at longer time scales, ideally should be the same in simulation and reconstructions. To what extent the simulated and observed series should be filtered to highlight the common external signal is a priori unknown. An estimate of the necessary filtering can be provided by comparing the temperatures simulated in both ECHO-G simulations. Figure 1a shows the January–April temperatures for

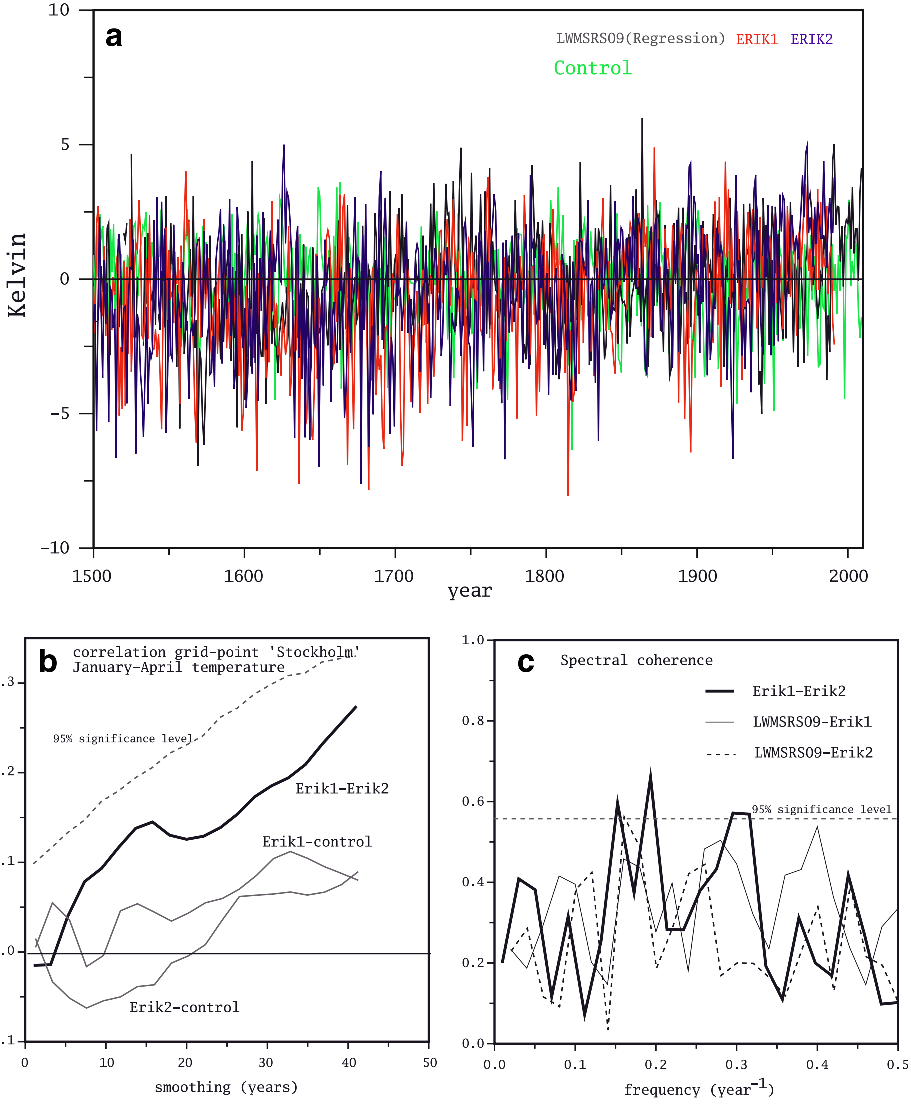

Fig. 1 January to April mean temperature in Stockholm reconstructed from documentary records (Leijonhufvud et al. 2009, LWMSR09) (regression version) and simulated in two runs with the global climate model ECHO-G (ERIK1, ERIK2). *Upper panel* **a** deviations from the AD 1829–1929 mean; *left panel* **b** correlation between the linearly detrended ERIK1 and ERIK2 series after weak to strong smoothing with a Gaussian filter, together with the 95% significance level obtained from Monte Carlo simulations. *Right panel* **c** coherence spectra between the ERIK1–ERIK2 series and between the LWMSR09 (regression version)

the Stockholm grid-cell, together with the reconstructed series (regression version only). We consider deviations from the AD 1829–1929 mean (the last 101 years common to all simulated series). For comparison, a 500-year temperature series for the Stockholm grid-cell from a control simulation for the present climate—with constant external forcing, is included (Zorita et al. 2003).

The temperature variations are dominated by the large interannual variability and it is actually very difficult to identify any coherency among the series. The correlations between the simulated series, after linear detrending and an increasing degree of smoothing is shown in Fig. 1b, together with the 95% significance level estimated with paired random Gaussian noise series smoothed in the same way. The linear detrending is introduced to exclude the effect of the common linear multi-centennial trend and identify the possible common decadal variations. None of the correlations are found to be individually significant. Taken as a whole, the figure does indicate a weak commonality, since all smoothed correlations are positive, this is unlikely to occur by chance. This is supported by the corresponding correlations between the forced simulations (ERIK1 and ERIK2) and the control run, both of which are weaker (Fig. 1b). Nevertheless, the amount of variance shared by ERIK1 and ERIK2 is quite small as indicated by the small magnitude of the correlations.

The coherence spectrum between both simulated series (Fig. 1c), shows significant coherence at the 95% level for three isolated frequencies in the range between 0.02 and 0.5 year^{-1}. This frequency range, which corresponds to time scales of 50 and 2 years respectively, is dictated by the length of the records, about 500 years. The power of lowest frequency cycles can therefore be estimated with ten cycles in the record. The estimation of the power of lower frequencies would be compromised by small number of cycles present. Therefore, the frequency range shown in Fig. 1c does not include the long-term trends present in the records. When estimating the coherency at 50 independent frequencies one would expect about three frequencies to attain the 95% significance level by chance. On average, therefore, the only commonality between the simulated temperatures seems to be the long-term trend across the five centuries (see following paragraph). This result indicates that no agreement between the simulated and the reconstructed series should be expected on average (punctual external shocks such as volcanic eruptions could, however, induce coherent cooling in both records, Wagner and Zorita 2005). The coherence between the simulated series is an ideal case, as both share exactly the same forcing. As expected, the simulated series show no significant coherence with the reconstructed series (Fig. 1c).

The picture becomes clearer when considering the series after a 31-year running-mean smoothing. Both reconstruction versions (regression and variance scaling) are very similar after the 31-year smoothing (Fig. 2a). Reconstructed temperatures start with a declining trend in the early part of the record, reaching a minimum between AD 1550–1650, followed by a prolonged long-term warming until present. The long-term trend along the whole period is 0.37 ± 0.10 K/century (2 x standard error uncertainties). Superimposed on this trend a roughly 80-year long warming episode during AD 1680–1760, peaking in the 1730s, is also observed. Both simulations reproduce the long warming trend along the five centuries with superimposed decadal variations. The decadal variations in the two simulations are uncorrelated (Fig. 1b, c). The evolution of the temperature simulated by the regional model follows, in general, the temperature from the driving global simulation ERIK1 rather closely. Differences are clearly smaller than those between the output from the two global models.

With the exception of the warm episode around AD 1730 seen in the reconstruction, the long-term positive temperature trend agrees well between the reconstructions and simulations: 0.44 ± 0.11 K/century in ERIK1, 0.47 ± 0.12 K/century

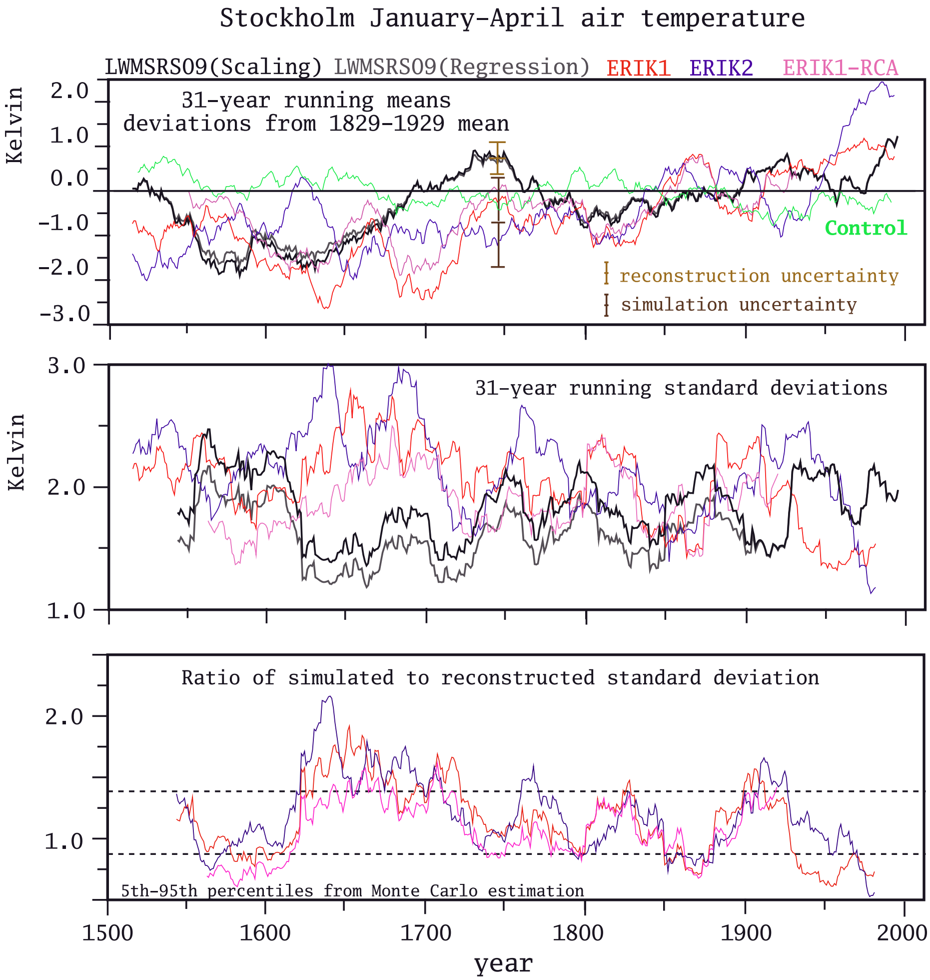

Fig. 2 January to April mean temperature in Stockholm reconstructed from documentary records (Leijonhufvud et al. 2009; LWMSR09) (two versions: variance-scaling and regression calibration) and simulated in two runs with the global climate model ECHO-G (ERIK1, ERIK2) and one run with the regional climate model RCA (ERIK1—RCA). *Upper panel*: deviations from the AD 1829–1929 mean smoothed with a 31-year running mean filter; *middle panel*: 31-year running standard deviations; *lower panel*: ratio of the simulated standard deviations to the reconstructed (variance-scaling) standard deviation

in ERIK1 and 0.31 ± 0.15 K/century in the regional RCA simulation (note the RCA simulation ends in 1929). An interesting aspect is that, although the control simulation displays some variability at multidecadal timescales, it does not show a comparable multi-centennial trend (−0.13 ± 0.08 K/century), as both forced simulations do. The most reasonable interpretation is that this multi-centennial trend, seen in the forced simulations, is mainly caused by the external forcing. Also, it suggests that the amplitude of the centennial trend in the external forcing was reasonably chosen when the simulations were performed (Zorita et al. 2005). In general, this amplitude can primarily be varied by adjusting the amplitudes of the solar and volcanic forcings, which are not well constrained by reconstructions. The greenhouse gas forcing is, in contrast, well constrained by ice-core measurements. The agreement in the trend is therefore relevant because the present uncertainties in the magnitude of the changes of the solar forcing in the last centuries (see e.g. discussion in Gouirand et al. 2007) or an unrealistic model climate sensitivity could easily have caused a mismatch between the simulated and observed long-term trends in the past few centuries.

As already mentioned, a discrepancy is found between the simulations and the reconstructions from about AD 1680 to 1760. The warming episode, on top of the long-term centennial trend and lasting for about 80 years in the reconstruction, is absent in both simulations. This mismatch is difficult to reconcile by invoking the estimated uncertainties in the reconstructed temperature or by internal model variations. Figure 2 illustrates this by showing the amplitude of the uncertainty in the reconstructions at multi-decadal timescales (the 31-year means). These uncertainties have been derived from the root-mean-square error (rmse) between the reconstructed record (regression-based version) and observed temperatures in a validation period (Leijonhufvud et al. 2009). The inter-annual errors in the validation period are serially uncorrelated, so that the approximate 2-sigma uncertainty in the 31-year running means is taken as 2 x rmse divided by $\sqrt{\sum_i w_i^2}$, where w_i are the weights of the filter. The uncertainties associated with the internal variability in the model can be estimated from two methods. One estimate is the amplitude of the simulated temperature variations in the same grid-cell in a control simulation with the global model along 1,000 years. This amplitude should include all sources of variability that are not connected to the external forcing, e.g. internal variations of the North Atlantic Oscillation modulating the advection of maritime air towards Scandinavia, or variations of the sea-surface temperatures in the North Atlantic caused by internal ocean dynamics. The magnitude of the model uncertainty is taken as twice the standard deviation of the smoothed (31-year running mean) simulated Stockholm temperature in the control run. The other method is to calculate the standard deviation of the difference between the two simulated time series in ERIK1 and ERIK2. This standard deviation should be a factor $\sqrt{2}$ larger than the standard deviation associated with the internal variations. This latter method yields a larger value for the internal variations and this is the one shown in Fig. 2a. The internal variations may also depend on the basic mean climatic state. As the control run was performed with present forcing, and not with the external forcing of the years around AD 1730, the estimation based on the control run may be biased. Similar caveats apply to the second estimation method, as the basic mean climate state changes along the 500 years of the simulations. A more accurate estimation of the internal random variations can only be provided by a large ensemble of simulations with the same model and the same external forcing. As stated above this is, at present, computationally very demanding

From these uncertainty estimations as well as estimations of the amplitude of the internal variations, it appears that the sustained warm episode in the reconstructions around AD 1730 is not compatible with the simulations.

It is unlikely that variations in the external forcing as prescribed in the model can explain the early 80-year warming episode in the reconstruction, as this is absent in both global simulations. This may be interpreted as an indication that the multi-decadal internal variability in the ECHO-G model is too small; this could be due to the dynamics of the North Atlantic Ocean circulation being too weak in the model. It may alternatively indicate that the reconstructed Stockholm temperatures are too warm around AD 1730. However, there is independent support from the instrumental temperature measurements at the nearby city of Uppsala that the 1730s were indeed unusually mild (Moberg et al. 2005; Leijonhufvud et al. 2008). Moreover, the Central England temperatures and also the De Bilt temperature in the Netherlands show a notable warmth in the same period (Jones and Briffa 2006) (see also below). The mismatch between proxy and model data for Stockholm during the decades around AD 1730 is best explained, on the basis of the limited evidence available so far, as an incapability of the model to reproduce this warming feature (see also the conclusion section).

The Stockholm inter-annual temperature variability, as expressed by running 31-year standard deviations, shows large inter-decadal changes. The inter-annual variability of the variance-scaled series is, by construction, larger than the regression-based reconstruction. The reconstruction standard deviation largely agrees with the simulations in some periods, but there are also periods when the simulated variability is larger in both simulations, in particular between AD 1620 and AD 1700 (Fig. 2). The ratio of the modeled to reconstructed (variance-scaled) standard deviations is shown also in Fig. 2. The significance of this ratio is assessed by constructing 1,000 pairs of synthetic time series, both with standard deviation unity and the same autocorrelation as the simulated and reconstructed temperatures, respectively. Only in the decades around AD 1620–1700 is the ratio of simulated to reconstructed variance significantly (5% level) larger than could be explained by chance.

The inter-annual variability in the regional model, however, does not significantly differ from the variance-scaled reconstruction in the AD 1620–1700 period. Rather, the regional model variability is essentially consistent with the Stockholm temperature reconstruction everywhere except before AD 1620, when the ratio between the simulated and reconstructed variance is below the 5th percentile derived from the Monte Carlo simulations (and for a few years around AD 1650, when fluctuating above the 95th percentile). As this happens about 10% of the entire time period, one may conclude that the simulated regional inter-annual temperature variance is not significantly different from the reconstruction. Notably, the variance in the regional model is almost consistently smaller than in the corresponding (ERIK1) global simulation, and also closer to the reconstruction than the global simulation. This indicates that the regional model provides a more realistic simulation of the local climate in this area.

All of the records display no clear trend in the inter-annual temperature variance and the differences between time evolution of the global model records are also large for much of the time. From the large differences between the simulated inter-annual variance in the two global simulations, it can be concluded that the external forcing can only account for a small amount of inter-annual variability, which thus must be dominated by internal variability.

4 Modeled and reconstructed Central European Temperature series

In this section we compare the model simulations obtained with ECHO-G with the reconstruction of the Central European air temperature based on documentary data (Dobrovolný et al. 2009b). Although the reconstruction is available for all 12 calendar months, we restrict our analysis to seasonal and annual data because a more robust comparison is achieved using such averaged time series.

As in the case with the Stockholm results, both simulations do not show coherent behaviour at any timescale other than the long-term trend. The results of the correlation between smoothed simulated series and their coherence spectrum, shown in Fig. 1 for Stockholm, also holds for the CEurT (not shown). Figure 3 shows the smoothed annual and seasonal reconstructed CEurT and the corresponding

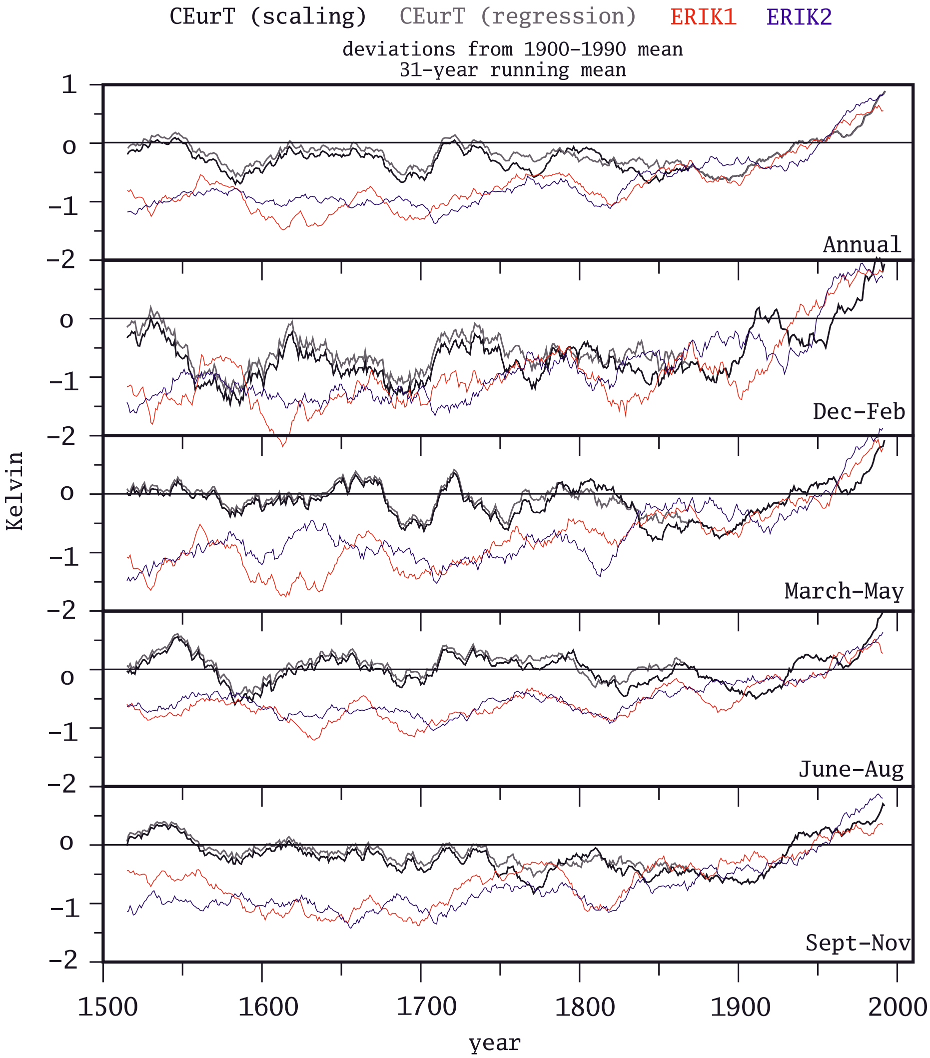

Fig. 3 Annual and seasonal means of the Central European Temperature reconstructions (Dobrovolný et al. 2009b, CEurT scaling, CEurT regression) and of the corresponding grid-cells in two simulations with the global climate model ECHO-G (ERIK1, ERIK2). The series represent deviations from the AD 1900–1990 mean and are smoothed with a 31-year running mean filter

temperatures simulated by the model averaged over eight grid-cells that roughly correspond to the Central European area. Deviations from the AD 1900–1990 mean are presented here. Because of the coarse resolution of the global model, the model topography in this area is rather flat (the Alps are represented by elevations of approximately 600 m). Therefore, the results are not critical to the exact choice of model grid-cells to represent the Central European area. As in the case of the Stockholm reconstruction, the proxy data have been calibrated with two methods: regression calibration, and adjustment of mean and variance to the observations in the calibration period.

Overall, the relationship between reconstructed and simulated temperatures has a different character for Central European temperatures than presented for Stockholm. There is good agreement between model and the reconstruction back to about AD 1850. The Central European reconstructions mostly show systematically warmer temperature deviations than the model before AD 1850, whereas the differences between reconstructed and simulated Stockholm temperature are rather more random in character, and include both positive and negative signs. This is found for all seasonal records and therefore for the annual means as well, although the differences between model and reconstructions are smaller for the winter season.

The seasonal reconstructions show multi-decadal variability with an amplitude similar to that displayed by the model simulations. However, the positions of the maxima and minima do not tend to agree with those of the simulations. This can be attributed to internal regional variability, since the timing of the maxima and minima do not always agree in the two model simulations. A more striking difference between the reconstruction and the simulations is that the long-term warming found over the last five centuries in the simulations is much less pronounced in the reconstruction. There is even a slight cooling trend seen in the reconstruction over the first centuries after AD 1500. In the period AD 1500–1800, both simulations display positive temperature trends for all seasons, which are statistically significant at the 5% level in winter, autumn and annually. For instance, the trends for the annual values are about 0.09 ± 0.07 K/century. By contrast, the reconstructions display non-significant trends which are all negative except for summer.

The inter-annual variance in the reconstructions and simulations (31-year running standard deviations; Fig. 4) shows both similarities and differences depending on the time period and season. In winter and summer, the reconstructions show a large inter-annual variance at the beginning of the record, which diminishes along the past half-millennium. In contrast the spring and autumn records do not show any discernible long-term trends in their inter-annual variance (Dobrovolný et al. 2009b). The spring reconstruction nevertheless displays somewhat smaller inter-annual variance in the twentieth century compared to the previous four centuries. The simulated records show a fairly constant inter-annual variance in all seasons, which stands in contrast to the instrumental data for winter and summer but agrees reasonably well in spring and autumn. Thus, at the beginning of the past half-millennium the reconstructions display larger inter-annual variance than the simulations during winter and especially so during the summer season. The reconstructed pre-instrumental variability could possibly be too high in these seasons (both in the regression and the variance matched version). If so, the reason for this is at present not known. Since the regression residuals for the summer are significantly serially positively correlated according to a Durbin–Watson test (see Dobrovolný et al. 2009b), the inter-annual variance of the reconstruction in summer could be incorrect. This may also affect

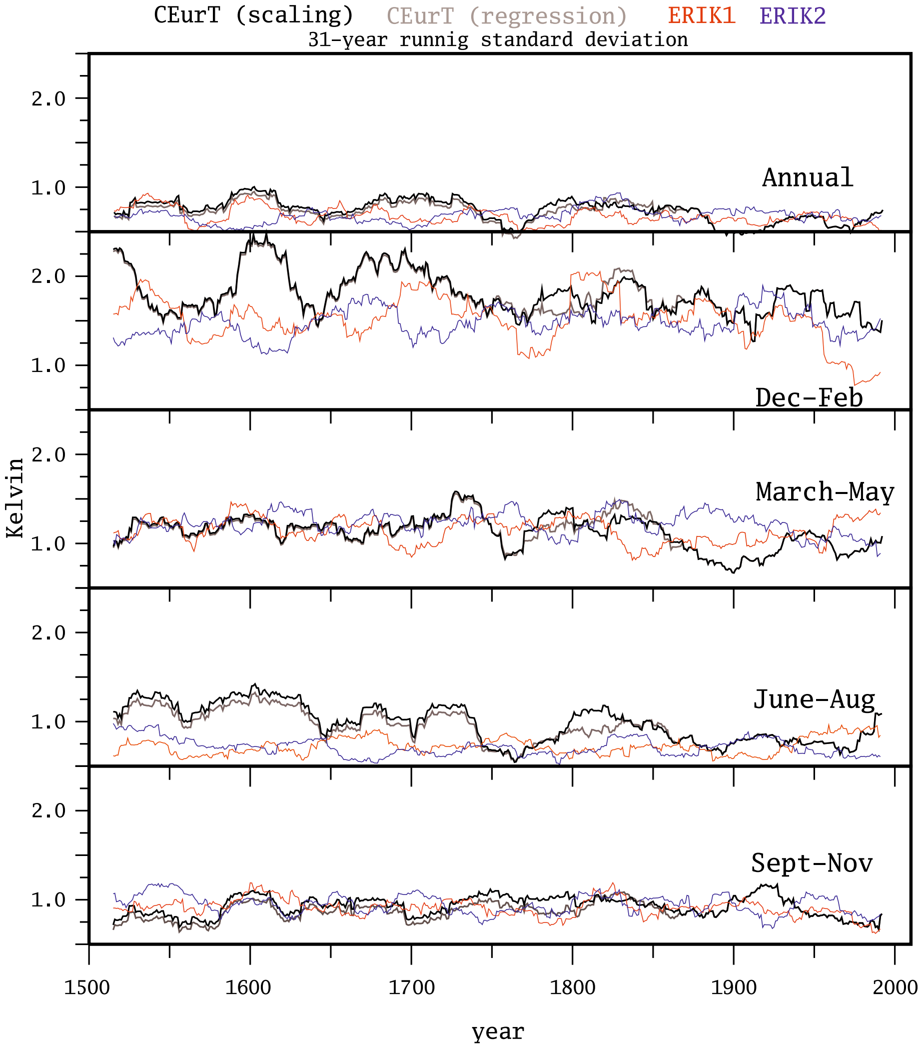

Fig. 4 Running standard deviation (31-year window) of the annual and seasonal temperature derived from the reconstructed Central European temperature and from output of two simulations with the global climate model ECHO-G in the corresponding grid-cells

the annual data, which also exhibit significant positive autocorrelation (Dobrovolný et al. 2009b). The autocorrelated residuals may explain some of the differences in reconstructed and simulated inter-annual temperature variance in summer, but does not explain the differences in winter. One possibility is that the real climate had, on average, more high-frequency temperature variability in summer and winter during the sixteenth to eighteenth centuries, perhaps caused by another circulation mode, and that this is correctly picked up by the reconstruction but not by the simulations. Previous analyses of reconstructed atmospheric circulation features over Europe (500 hPa geopotential height fields) and those simulated with ECHO-G over the last 350 years indicate quite substantial differences in some periods (Casty et al. 2005), which likely are not related to a response to the external forcing, either in the simulation or in the reconstructions (Casty et al. 2007). These differences could therefore be caused by internal variability. In any case, it appears that further studies are needed to understand the differences between simulated and reconstructed interannual variability in the pre-instrumental period.

5 Modeled and observed Central England Temperature series

All proxy-based reconstructions of past temperatures contain uncertainties of a statistical nature, which can be partially quantified, but there could be additional unknown sources of error which have not been taken into account. To shed more light on whether or not the agreements and discrepancies found between the Stockholm and CEurT reconstructions and the model simulations are robust, we carry out a similar comparison between the simulation and the long instrumental CET record. However, it should be emphasized that this 350-year long temperature record can hardly be considered homogeneous over the whole period (recall our brief discussion in Section 2). The results of this comparison are conveyed by Fig. 5. To represent the 'model CET', four model grid-cells located over Great Britain have been selected. Deviations from the AD 1900–1990 mean are considered.

The long-term linear trend of the annual CET series over the period common to the simulations AD 1656–1990 is 0.19 ± 0.05 K/century (2 x standard error uncertainty). This is somewhat lower than the simulated linear trends (0.29 ± 0.04 K/century and 0.32 ± 0.05 K/century, in ERIK1 and ERIK2 respectively). The control simulation, in contrast, displays an insignificant trend (−0.12 ± 0.17 K/century). This again supports the idea that the choice of external forcing, combined with the model climate sensitivity, yields a reasonable picture of long-term temperature evolution over recent centuries. As in the case of the Stockholm temperature series, this long-term trend is very likely also externally forced. As discussed below, the weaker trend for the annual values in the CET series compared to the forced simulations can be partly due to the summer trend. This summer trend might be affected by homogenization problems.

A multi-decadal disagreement is noted around the AD 1710s to 1730s, similar to the one noted earlier for the Stockholm January–April temperature reconstruction. In CET, this mismatch is seen particularly clearly in the annual, March–May and September–November data but also in December–February data. The CET and Stockholm records have been derived independently, and therefore, one has to conclude that the relatively warm decades around AD 1730 is probably a real phenomenon. The agreement between the CET and Stockholm series could indicate that the anomaly might have its origin in the North Atlantic and that it could have been a more widespread phenomenon (see also Luterbacher et al. 2004 and Xoplaki et al. 2005 for European mean temperature; Jones and Briffa 2006). However the anomaly is not as clearly found in the CEurT series, although the reconstructed winters for that region during the AD 1710s to 1740s display modest warmer temperatures.

The long-term evolution of CET summer temperatures over the whole period clearly differs from the model simulations. The trend in the observed summer

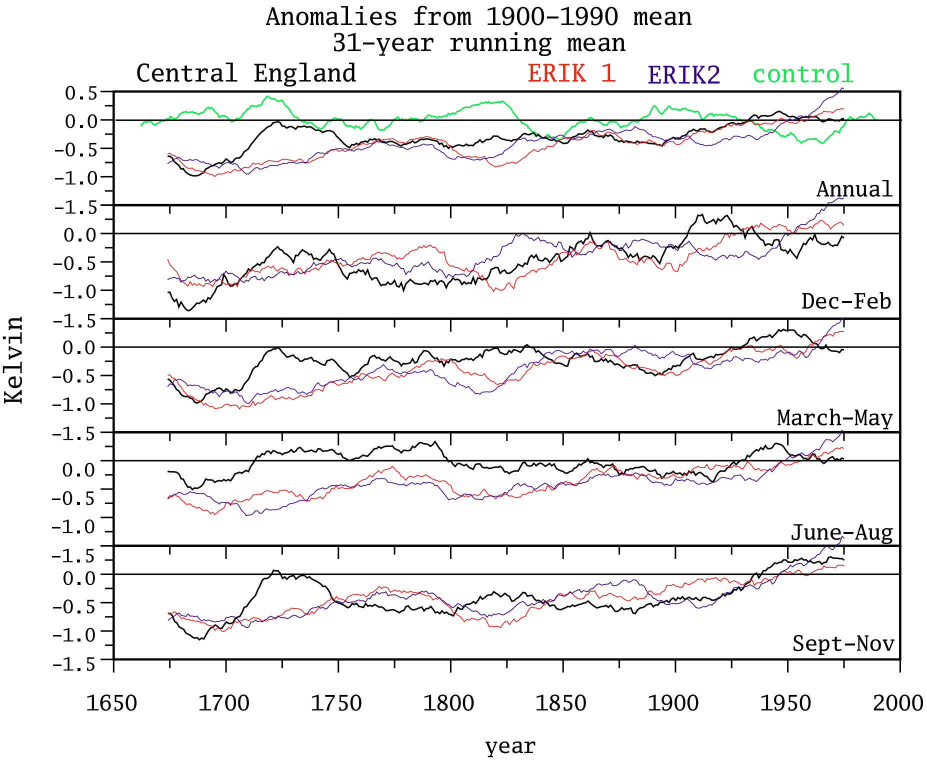

Fig. 5 Annual and seasonal means of the Central England Temperature series and of the corresponding grid-cells in two simulations with the global climate model ECHO-G (ERIK1, ERIK2). The time series represent deviations from the AD 1900–1990 mean and are smoothed with a 31-year running mean

temperatures is very small, and summer temperatures in the 18th century seem to be higher than recent temperatures until AD 1990. This result is not seen for the other seasons, where the observed and simulated trends are comparable. The practically constant observed summer temperatures together with the increasing winter temperatures along the past 350 years lead to a clear long-term decreasing amplitude of the observed annual cycle, which is absent in the simulations (Fig. 6), and which has been reported also for other long instrumental temperature records in the Northern Hemisphere (Jones et al. 2003). However, the constantly warmer instrumental summer temperatures before the late 19th century in the CET data suggest the presence of a warm bias in the early English observational data in this season, similar to what has been concluded for Central European stations (Frank et al. 2007a, b) and also proposed for Stockholm and Uppsala summer temperatures (Moberg et al. 2003). The argument for the presence of such a bias is that the thermometers were less well protected against radiation effects before the introduction of 'modern' screens, which typically occurred in the mid- to late-nineteenth century. This effect is largest in summer. This issue is discussed in more detail in Böhm et al. (2009) for the Great Alpine Region. Böhm et al. (2009)

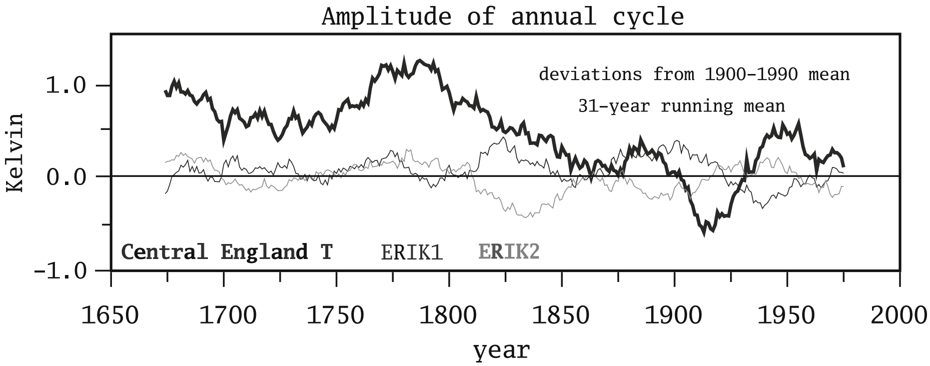

Fig. 6 Evolution of the amplitude of the annual cycle (June–August mean minus December–February mean) in the Central England temperature record and in two simulations with the climate model ECHO-G. The timeseries represent deviations from the 1900–1990 mean and are smoothed with a 31-year running mean

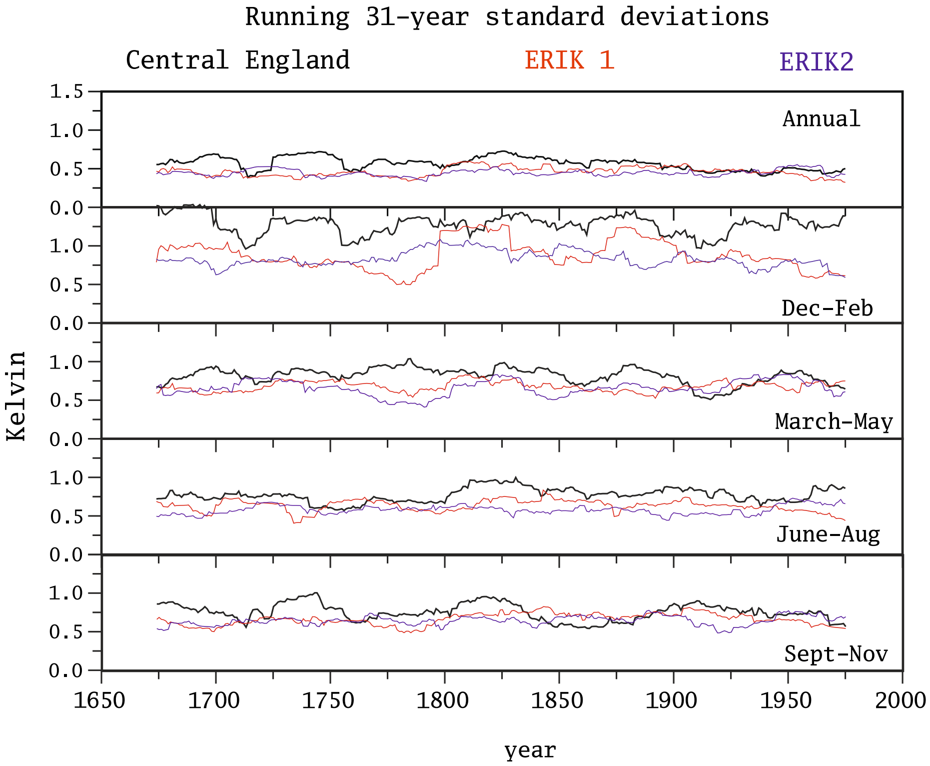

Fig. 7 Running inter-annual standard deviation of the annual and seasonal temperature derived from the Central England temperature records and from output of two simulations with the global climate model ECHO-G in the corresponding grid-cells

estimate a correction for the summer temperature bias for this region in the range of 0.3–0.4 K (see their Figure 13). A similar correction for the CET, if applicable, would put the CET summer value closer to the model results, but still, the simulations and the CET would not fully agree in summer. The magnitude of the appropriate correction for CET would probably be different to that for the Alps, as the summer insolation in Britain and the typical location of the thermometers could be also different. An analysis of the possible presence of warm biased summer temperatures at the stations used to derive the CET record, in the period before the improvement of thermometer screens, seems worthwhile.

The inter-annual variance of the observed and simulated records (Fig. 7) is notably different for winter temperatures, when the observations display a systematically larger variance than the simulations. There is also an overall larger variance in the instrumental series than the simulated annual mean temperatures before around AD 1900, and particularly before AD 1800. However, none of the seasonal records of inter-annual variance show any clear trends, perhaps with the exception of a slight downward trend in the observed annual CET record since AD 1800. Moreover, the inter-annual variance in the two simulations does not generally evolve in phase, possibly reflecting limited influence of the external forcing on the inter-annual variability.

6 Discussion and conclusions

The comparisons of reconstructed/instrumental and simulated temperatures for three European locations over the last 350 to 500 years display encouraging agreements—especially considering the global model's coarse resolution—but also disagreements. Good agreement is seen between the multi-centennial trends for the Stockholm and CET records. Both forced simulations also agree at multi-centennial scales. The control simulation fails to produce comparable trends, and we interpret these results as an indication that the long-term trends are probably externally forced. However, in all cases the simulated trends are slightly larger than in the reconstructions, although for the Stockholm case they agree within their uncertainty bounds. In the CET record the instrumental observations might be affected by an unrealistic trend in summer. Although this point has to be investigated in much more detail, the absence of a clear long-term summer temperature warming trend in the CET record, which agrees with the Central European Temperature reconstruction in this respect, is an example of disagreement between simulations and reconstructions in the European region in summer. Other dendroclimatological evidence does show clear centennial trends in summer temperature in the Alpine region over the last four centuries (Büntgen et al. 2006).

The disagreement between the model and reconstruction series presented here should be considered with the coarse spatial resolution of global models in mind. Coastlines and topographic features that in reality may be important for regional climate are only approximately represented in these models. It should also be remembered that the estimation of external forcings for the past centuries is burdened with uncertainties. Further modelling work would be required by analyzing a whole ensemble of simulations with the global and regional climate models combined. The present analysis, however, indicates, that these discrepancies could be of significance with regard to discussions about the amplitude of past changes in solar irradiance and also raises questions regarding uncertainties in the long-term instrumental records. One possible explanation for the disagreement between the simulated and reconstructed summer temperature trends is that the long-term trend in the external forcing may have been overestimated in the simulations. More recent estimations of the changes in solar irradiance between the Late Maunder Minimum and present are of the order of 0.17% (Wang et al. 2005), or even less than 0.1% (Foukal et al. 2004), instead of the 0.3% in the ECHO-G simulations. Different choices of the amplitude of solar irradiance would affect the summer temperatures more strongly than winter. A too strong irradiance amplitude in the ECHO-G simulations could thus explain why the model exhibits long-term warming trends in summer in Central Europe and Central England, at odds with the absence of summer trends in the reconstructions for these locations. On the other hand, such small amplitude changes in solar irradiance of less than 0.1% would hardly be compatible with the general agreement found between the simulated temperatures in the other seasons and the CET or the Stockholm series. The climate sensitivity of ECHO-G (2.5 K caused by a doubling of atmospheric CO_2 concentrations) lies well within the range of the IPCC climate models (Meeh et al. 2007), so that an anomalously high sensitivity of this particular model to external perturbations is not a likely possibility.

A striking difference between the reconstructed Central European Temperature and the simulations is the long-term temperature trend for all seasons (i.e. not just for summer as in the case of CET), which are much smaller in magnitude than those simulated by the climate model. Although a model shortcoming cannot be ruled out, the fact that the long-term trends of the Stockholm January–April temperature and the seasonal Central England Temperatures (except in summer) are much more in agreement with the model results, may indicate that the reconstructed Central European record underestimates the multi-centennial variations. There are plausible reasons for this behaviour: in contrast to the Stockholm temperature reconstruction, which is derived from physically-anchored documentary information, the Central European record is composed of more varied documentary records, some of which could suffer from a historical based variant of the so called ‘segment-length-curse’ discussed in the tree ring community (Cook et al. 1995). As mentioned by Dobrovolný et al. (2009a, b), the documentary sources contain descriptions of weather recorded according to the contemporary authors’ own perceptions of what constituted ‘normal’ conditions. These perceptions were framed by the period in which the authors were living. Therefore, the index series derived from documentary evidence expresses primarily deviations from the respective normal conditions. Slow changes in these normal conditions can thus likely not be fully captured by the reconstructed indices. This shortcoming can possibly be partly addressed if other proxies are used that are more directly anchored in the prevailing environmental conditions, such as series of phenophases, freezing of rivers and lakes, agricultural harvest dates, etc. Certainly, further combined analyses and comparisons with other proxy-based records are necessary to ascertain whether or not this shortcoming is relevant for a particular record.

At shorter timescales both forced simulations (ERIK1 and ERIK2) present different temperature trajectories for the locations analyzed here. These regional variations in the model are probably internally generated and indicate the larger internal climate variability at these small spatial scales. This is in contrast with the temperature variability at hemispheric scales (Zorita et al. 2007), where different simulations do display clear similarities between them and with proxy-based reconstructions (Jansen et al. 2007). One important difference between models and reconstructions is the amplitude and persistence of the multi-decadal warm episode peaking around the AD 1730s that is displayed by the Stockholm reconstruction and also by the Central England record. In CET, it is present not only in the winter season, but also in spring and autumn and hence is reflected also in the annual mean temperatures. The simulations do not show similar events superimposed on the long-term centennial trend. The origin of the warm episode may lie in the wintertime North Atlantic circulation, either atmospheric or oceanic. In summer, the season where the short-wave radiation forcing at the surface is most powerful, the decadal deviations from the long-term trend are not particularly marked.

In contrast to our results, previous comparative studies of model simulations and reconstructions at the European scale have concluded that control simulations can produce internal decadal variations large enough to explain the variations found in the reconstructions of European mean annual temperature (Bengtsson et al. 2007). The results obtained here for the Stockholm and CET temperature series indicate, however, that this is not the case for multi-decadal trends. Our findings suggest that the simulated internal variability is too small to explain the observed multi-decadal variations. A possible alternative interpretation would be that external forcing is required to explain the reconstructed multi-decadal variations.

A tentative explanation for the mild decades around AD 1730 could be related to a prolonged positive phase of the North Atlantic Oscillation (Luterbacher et al. 1999). Increased solar irradiance at the end of the 17th century and the first half of the 18th century might have induced a shift to a positive NAO mode. Similar findings were suggested for European winters and springs (Luterbacher et al. 2004; Xoplaki et al. 2005). Thus, the solar irradiance changes might be a major trigger to explain the increasing trend in European winter and spring temperature at decadal time scale. However, reconstructions of the NAO index do not agree (Schmutz et al. 2000; Cook et al. 2002). For instance, the NAO reconstruction by Cook et al. (2002) and the more recent by Trouet et al. (2009) do not show such anomalous decades as the Luterbacher et al. (1999) series. From a modeling point of view, simulations indicate that increased solar irradiance at the end of the Maunder Minimum (Shindell et al. 2001) might have indeed induced a shift towards positive AO/NAO during November–April, leading to continental warming. It is known that climate models are not able to replicate the observed low-frequency variations of the atmospheric circulation in the North Atlantic in the 20th century (Osborn 2004), and even future trends simulated by different models under much stronger external forcing, according to future emissions scenarios, do not agree in magnitude (Miller et al. 2006). It is also possible that natural, multi-decadal excursions of the heat transport by the atmosphere are not well replicated by models.

It should to be noted that our conclusions regarding internal variability are based on one control simulation and only two forced simulations. A better estimation would require a much larger ensemble of forced simulations over the whole period. For instance, the difference in Stockholm temperature between ERIK1 and ERIK2 around AD 1630 (Fig. 1) is larger than our estimation of internal variability based on a control simulation, and indeed the NAO indices in ERIK1 and ERIK2 show the largest differences of the whole millennium around AD 1630. This will be explored in the future with a larger ensemble of simulations.

With regard to the CET record, the notably warm temperatures in summer before AD 1800 compared to the twentieth century, which are also warmer than the simulated summer temperatures before the late twentieth century, may perhaps indicate that the early observations suffer from a positive bias before the introduction of modern-type thermometer screens, as has been found in early instrumental records from the Great Alpine Region (Frank et al. 2007a, b; Böhm et al. 2009) and also suggested for Stockholm (Moberg et al. 2003). This points to the importance of further investigating the possible early warm instrumental temperature bias, not only in the CET, but potentially in all temperature records extending back before the introduction of modern-type thermometer screens (typically before the mid-to-late-19th century). The absence of a long-term warming trend in the reconstructed Central European summer temperatures analyzed here cannot be easily explained by any warm bias in the early instrumental observations as these were adjusted for this problem (Böhm et al. 2009).

In summary, the comparison between modeled and reconstructed records points to particular lines of research for all teams involved. It seems important to confirm the limitation of the climate model to simulate the large multi-decadal regional temperature excursions as found for the Stockholm and Central England temperature, as it could point to possible future regional deviations from the global trend. Two simulations are too few and this should be more accurately ascertained with larger ensemble simulations with regional climate models. Further investigations with larger ensemble simulations using regional models would allow us also to investigate the physical origins and dynamics of these excursions. In the end, the analysis should yield a consistent picture of model and reconstructions, thus further constraining the possible range of models and identifying possible biases in climate reconstruction methods.

Acknowledgements This study is part of the European project MILLENNIUM (contract no 017008-2 (GOCE)). A. Moberg, L. Leijonhufvud and J. Söderberg acknowledge the Swedish Research Council. L. Leijonhufvud expresses her gratitude to FORMAS. F. González-Rouco was supported by the Spanish 'Ministerio de Educación y Ciencia' through grant REN-2002-04584-CLI. E. Zorita acknowledges the support from the German Science Foundation project SINCOS. The Central England temperature data were kindly provided by the Hadley Centre for Climate Prediction and Research www.metoffice.gov.uk/hadobs.

References

Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones PD, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin J-M, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovic Z, Nieplova E (2007) HISTALP—historical instrumental climatological surface time series of the greater Alpine region 1760–2003. Int J Climatol 27:17–46

Bader DC, Covey C, Gutowski WJ, Held IM, Kunkel K, Miller RL, Tokmakian RT, Zhang MH (2008) Climate models: an assessment of strengths and limitations. Synthesis and assessment products 3.1 report by the U.S. In: Climate change science program and the subcommittee on global change research, Washington DC, 133 pp

Bengtsson L, Hodges KI, Roeckner E, Brokopf R (2007) On the natural variability of the pre-industrial European climate. Clim Dyn 27:743–760

Böhm R, Jones PD, Hiebl J, Frank D, Brunetti M, Maugeri M (2009) The early instrumental warm-bias: a solution for long central European temperature series 1760–2007. Clim Change. doi:10.1007/s10584-009-9649-4

Brázdil R, Pfister C, Wanner H, von Storch H, Luterbacher J (2005) Historical climatology in Europe—the state of the art. Clim Change 70:363–430

Büntgen U, Frank DC, Nievergelt D, Esper J (2006) Summer temperature variations in the European Alps: AD 755–2004. J Climate 19:5606–5623

Casty C, Handorf D, Raible CC, González-Rouco JF, Weisheimer A, Xoplaki E, Luterbacher J, Dethloff K, Wanner H (2005) Recurrent climate winter regimes in reconstructed and modelled 500 hPa geopotential height fields over the North Atlantic/European sector 1659–1990. Clim Dyn 24:809–822

Casty C, Raible CC, Stocker TF, Wanner H, Luterbacher J (2007) A European pattern climatology 1766–2000. Clim Dyn 29:791–805

Chuine I, Yiou P, Viovy N, Seguin B, Daux V, Le Roy Ladurie E (2004) Grape ripening as a past climate indicator. Nature 432:289–290

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

Cook E, D'Arrigo R, Mann ME (2002) A well verified multiproxy reconstruction of the winter North Atlantic Oscillation index since A.D. 1400. J Climate 15:1754–1764

Dobrovolný P, Brázdil R, Valášek H, Kotyza O, Macková J, Halíčková M (2009a) A standard paleoclimatological approach to temperature reconstruction in historical climatology: an example from the Czech Republic, AD 1718–2007. Int J Climatol 29:1478–1492

Dobrovolný P, Moberg A, Brázdil R, Pfister C, Glaser R, Wilson RJS, van Engelen A, Limanowka D, Kiss A, Halíčková M, Macková J, Riemann D, Luterbacher J, Böhm R (2009b) Monthly and seasonal temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500. Clim Change. doi:10.1007/s10584-009-9724-x

Esper J, Wilson RJS, Frank DC, Moberg A, Wanner H, Luterbacher J (2005) Climate: past ranges and future changes. Quat Sci Rev 24:2164–2166

Fischer EM, Luterbacher J, Zorita E, Tett SFB, Casty C, Wanner H (2007) European climate response to tropical volcanic eruptions over the past half millennium. Geophys Res Lett 34:L05707. doi:10.1029/2006GL027992

Foukal P, North G, Wigley T (2004) A stellar view on solar variations and climate. Science 306:68–69

Frank D, Büntgen U, Böhm R, Maugeri M, Esper J (2007a) Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quat Sci Rev 26: 3298–3310

Frank D, Esper J, Cook ER (2007b) Adjustment for proxy number and coherence in a large-scale temperature reconstruction. Geophys Res Lett 34:L16709. doi:10.1029/2007GL030571

Gagen M, McCarrol D, Hicks S (2006) The Millennium project: European climate of the last millennium. PAGES News 14:4–4

González-Rouco JF, Beltrami H, Zorita E, von Storch H (2006) Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. Geophys Res Lett 33:L01703. doi:10.1029/2005GL024693

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

Goosse H, Arzel O, Luterbacher J, Mann ME, Renssen H, Riedwyl N, Timmermann A, Xoplaki E, Wanner H (2006) The origin of the European medieval warm period. Clim Past 2:99–113

Gouirand I, Moberg A, Zorita E (2007) Climate variability in Scandinavia for the past millennium simulated by an atmosphere–ocean general circulation model. Tellus 59A:30–49

Guiot J, Nicault A, Rathgeber C, Edouard J, Guibal F, Pichard G, Till C (2005) Last-millennium summer-temperature variations in western Europe based on proxy data. Holocene 15:489–500

Jansen E et al (2007) Paleoclimate. In: Solomon S et al (eds) Climate change 2007: the physical basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 433–497

Jones PD (1999) Classics in physical geography revisited. Manley, G. 1974: Central England temperatures: monthly means 1659–1973. Prog Phys Geogr 23:425–428

Jones PD, Briffa KR (2006) Unusual climate in northwestern Europe during the period 1730 to 1745 based on instrumental and documentary data. Clim Change 79:361–379

Jones PD, Briffa KR, Osborn TJ (2003) Changes in the Northern Hemisphere annual cycle: implications for paleoclimatology. J Geophys Res 108:4588. doi:10.1029/2003JD003695

Jones PD et al (2009) High-resolution paleoclimatology of the last millennium: a review of current status and future prospects. Holocene 19:3–49

Jones PD, Mann ME (2004) Climate over past millennia. Rev Geophys 42:RG2002

Kjellström E, Bärring L, Gollvik S, Hansson U, Jones C, Samuelsson P, Rummukainen M, Ullerstig A, Willén U, Wyser K (2005) A 140-year simulation of European climate with the new version of the Rossby Centre regional atmospheric climate model (RCA3). Reports Meteorology and Climatology 108, SMHI, Norrköping

Legutke S, Voss R (1999) The Hamburg atmosphere–ocean coupled model circulation model ECHO-G'. Technical Report No. 18. German Climate Computing Centre (DKRZ), Hamburg, 62 pp

Leijonhufvud L, Wilson R, Moberg A (2008) Documentary data provide evidence of Stockholm average winter to spring temperatures in the eighteenth and nineteenth centuries. Holocene 18:333–343. doi:10.1177/0959683608086770

Leijonhufvud L, Wilson R, Moberg A, Söderberg J, Retsö D, Söderlind U (2009) Five centuries of winter/spring temperatures in Stockholm reconstructed from documentary evidence and instrumental observations. Clim Change. doi:10.1007/s10584-009-9650-y

Luterbacher J et al (2006) Mediterranean climate variability over the last centuries: a review. In: Lionello P, Malanotte-Rizzoli P, Boscolo R (eds) The Mediterranean climate: an overview of the main characteristics and issues. Elsevier, Amsterdam, pp 27–148

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

Luterbacher J, Liniger MA, Menzel A, Estrella N, Della-Marta PM, Pfister C, Rutishauser T, Xoplaki E (2007) The exceptional European warmth of Autumn 2006 and Winter 2007: historical context, the underlying dynamics and its phenological impacts. Geophys Res Lett 34:L12704

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

Luterbacher et al (2009) Circulation dynamics and influence on European and Mediterranean winter climate over the past half millennium: results and insights from instrumental data, documentary proxy evidence and coupled climate models. Clim Change. doi:10.1007/s10584-009-9782-0

Luterbacher J, Xoplaki E, Küttel M, Zorita E, González-Rouco FJ, Jones PD, Stössel M, Rutishauser T, Wanner H, Wibig J, Przybylak R (2010) Climate change in Poland in the past centuries and its relationship to European climate: evidence from reconstructions and coupled climate models. In: Przybylak R et al (eds) The Polish climate in the European context: an historical overview, pp 3–39. Springer, Berlin. doi:10.1007/978-90-481-3167-9_1

Manley G (1953) The mean temperature of Central England, 1698 to 1952. Q J R Meteorol Soc 79:242–261

Manley G (1974) Central England temperatures: monthly means 1659 to 1973. Q J R Meteorol Soc 100:389–405

Meeh GA, Stocker TF , Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG , Weaver A, Zhao Z-C (2007) Global Climate Projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 747–845

Meier N, Rutishauser T, Pfister C, Wanner H, Luterbacher J (2007) Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD 1480. Geophys Res Lett 34:L20705. doi:10.1029/2007GL031381

Miller RL, Schmidt GA, Shindell DT (2006) Forced annular variations in the 20th century Intergovernmental Panel on Climate Change fourth assessment report models. J Geophys Res 111:D181101. doi:10.1029/2005JD006323

Min SK, Legutke S, Hense A, Kwon W (2005) Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. I. Near surface temperature, precipitation, and mean sea-level pressure. Tellus 57A:605–621

Mironov DV (2008) Parameterization of lakes in numerical weather prediction. Description of a lake model. COSMO Technical Report, No. 11, Deutscher Wetterdienst, Offenbach am Main, Germany, 41 pp

Moberg A, Bergström H, Ruiz Krigsman J, Svanered O (2002) Daily air temperature and pressure series for Stockholm (1756–1998). Clim Change 53:171–212

Moberg A, Alexandersson H, Bergström H, Jones PD (2003) Were Southern Swedish summer temperatures before 1860 as warm as measured? Int J Climatol 23:1495–1521

Moberg A, Tuomenvirta H, Nordli Ø (2005) Recent climatic trends. In: Seppälä M (ed) Physical geography of Fennoscandia. Oxford Regional Environments Series. Oxford University Press, Oxford, pp 113–133

Moberg A, Gouirand I, Schoning K, Wohlfarth B, Kjellström E, Rummukainen M, de Jong R, Linderholm H, Zorita E (2006) Climate in Sweden during the past millennium—evidence from proxy data, instrumental data and model simulations. SKB technical report TR-06-35. Swedish Nuclear Fuel and Waste Management Co., Stockholm, 87 pp

Osborn TJ (2004) Simulating the winter North Atlantic oscillation: the roles of internal variability and greenhouse gas forcing. Clim Dyn 22:605–623

Osborn TJ, Briffa KB, Jones PD (1997) Adjusting variance for sample size in tree-ring chronologies and other regional mean timeseries. Dendrochronologia 15:89–99

Parker DE, Horton EB (2005) Uncertainties in Central England temperature 1878–2003 and some improvements to the maximum and minimum series. Int J Climatol 25:1173–1188

Parker DE, Legg TP, Folland CK (1992) A new daily Central England temperature series, 1772–1991. Int J Climatol 12:317–342

Pauling A, Luterbacher J, Casty C, Wanner H (2006) Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. Clim Dyn 26:387–405

Raible CC, Casty C, Luterbacher J, Pauling A, Esper J, Frank DC, Büntgen U, Roesch AC, Tschuck P, Wild M, Vidale P-L, Schär C, Wanner H (2006) Climate variability—observations, reconstructions and model simulations for the Atlantic-European and Alpine region from 1500–2100 AD. Clim Change 79:9–29

Ramstein G, Kageyama M, Guiot J, Wu H, Hély C, Krinner G, Brewer S (2007) How cold was Europe at the Last Glacial Maximum? A synthesis of the progress acheived since the first PMIP model-data comparison. Clim Past 3:331–339

Rutishauser T, Luterbacher J, Defila C, Frank D, Wanner H (2008) Swiss spring plant phenology 2007: extremes, a multi-century perspective and changes in temperature sensitivity. Geophys Res Lett 35:L05703. doi:10.1029/2007GL032545

Schmutz C, Luterbacher J, Gyalistras D, Xoplaki E, Wanner H (2000) Can we trust proxy-based NAO index reconstructions? Geophys Res Lett 27:1135–1138

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

Trouet V, Esper J, Graham N, Baker A, Scourse JD, Frank D (2009) Persistent positive North Atlantic Oscillation mode dominated the medieval climate anomaly. Science 324:78–80

Wohlfahrt J, Harrison SP, Braconnot P, Hewitt CD, Kitoh A, Mikolajewicz U, Otto-Bliesner BL, Weber S (2008) Evaluation of coupled ocean–atmosphere simulations of the mid-Holocene using palaeovegetation data from the northern hemisphere extratropics. Clim Dyn 31:871–890. doi:10.1007/s00382-008-0415-5

Wagner S, Zorita E (2005) The influence of volcanic, solar and CO_2 forcing on the temperatures in the Dalton minimum (1790–1830): a model study. Clim Dyn 25:205–218

Wang Y-M, Lean JL, Sheeley NR (2005) Modeling the Sun's magnetic field and irradiance since 1713. Astrophys J 625:522–538

Wanner H, Beer J, Bütikofer J, Crowley TJ, Cubasch U, Flückiger J, Goosse H, Grosjean M, Joos F, Kaplan JO, Küttel M, Müller SA, Prentice C, Solomina O, Stocker TF, Tarasov P, Wagner M, Widmann M (2008) Mid- to Late Holocene climate change: an overview. Quat Sci Rev 27:1791–1828

Wilson R, Tudhope A, Brohan P, Briffa K, Osborn, Tett SFB (2006) Two-hundred-fifty years of reconstructed and modeled tropical temperature. J Geophys Res 111:C10007. doi:10.1029/2005JCC003188

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

Yoshimori M, Stocker TF, Raible CC, Renold M (2005) Externally forced and internal variability in ensemble climate simulations of the Maunder Minimum. J Climate 18:4253–4270

Zorita E, González-Rouco F, Legutke S (2003) Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model. J Climate 16:1378–1390

Zorita E, González-Rouco F, von Storch H, Montavez JP, Valero F (2005) Natural and anthropogenic modes of surface temperature variations in the last thousand years. Geophys Res Lett 32:L08707

Zorita E, González-Rouco F, von Storch H (2007) Comments on 'Testing the fidelity of methods to used in proxy-based reconstructions of past climate'. J Climate 20:3693–3698