



## Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation

H. Goosse,<sup>1</sup> E. Cresspin,<sup>1</sup> A. de Montety,<sup>1</sup> M. E. Mann,<sup>2</sup> H. Renssen,<sup>3</sup> and A. Timmermann<sup>4</sup>

Received 26 June 2009; revised 25 November 2009; accepted 15 December 2009; published 13 May 2010.

[1] Ensemble simulations have been performed with a climate model constrained to follow temperature histories obtained from a recent compilation of 56 well-calibrated surface temperature proxy records, using a new data assimilation technique. First, we demonstrate that the data assimilation technique provides a faithful representation in the Northern Hemisphere of the signal recorded by the climate proxies at both the regional and gridbox scales. Second, by varying the external forcing, the parameters of the data assimilation method, and the parameters controlling the equilibrium climate sensitivity of the climate model, we demonstrate that the uncertainty in model results is much lower in simulations using data assimilation than in those without it. This observation implies that the data assimilation, using a set of 56 proxies, is providing an efficient and robust constraint on the simulated climate variability over the past centuries. At the hemispheric and continental scales, the model reconstructions using data assimilation are in good agreement with both the instrumental record of the past 150 years and reconstructions of climate in past centuries derived from the application of traditional statistical approaches to networks of proxy data. This increases the confidence in both the data assimilation and traditional statistical approaches. Our data assimilation method, however, is unable to provide a reliable reconstruction over the North Atlantic Ocean, which we attribute to the paucity of proxy data in that region.

**Citation:** Goosse, H., E. Cresspin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann (2010), Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, *J. Geophys. Res.*, *115*, D09108, doi:10.1029/2009JD012737.

### 1. Introduction

[2] Due to the importance of placing recent 20th and 21st century warming in a longer-term perspective, the climate of the last millennium has received considerable attention over the past decade. This renewed focus has led to numerous efforts employing climate “proxy” records to reconstruct climate at the hemispheric [e.g., Mann *et al.*, 1999, 2008; Esper *et al.*, 2002; Mann and Jones, 2003; Moberg *et al.*, 2005; D’Arrigo *et al.*, 2006; Juckes *et al.*, 2007; Hegerl *et al.*, 2007] and regional [Luterbacher *et al.*, 2004; Guiot *et al.*, 2005] scales and to simulate climate over this time frame using climate models of various

complexity driven by both natural and anthropogenic forcings [e.g., Crowley, 2000; Bertrand *et al.*, 2002; Bauer *et al.*, 2003; Crowley *et al.*, 2003; Goosse *et al.*, 2005a; Gonzalez-Rouco *et al.*, 2006; Stendel *et al.*, 2006; Raible *et al.*, 2006; Tett *et al.*, 2007; Ammann *et al.*, 2007]. These various studies in large part are responsible for the recent assessment by the Intergovernmental Panel on Climate Change [see Jansen *et al.*, 2007] that “It is very likely that average Northern Hemisphere temperatures during the second half of the 20th century were higher than any other 50-year period in the last 500 years. It is also likely that this 50-year period was the warmest Northern Hemisphere period in the last 1.3 kyr” and that “Paleoclimate model simulations are broadly consistent with the reconstructed NH [Northern Hemisphere] temperatures over the past 1 kyr. The rise in surface temperatures since 1950 very likely cannot be reproduced without including anthropogenic greenhouse gases in the model forcings.”

[3] Despite these robust conclusions with regard to the anomalous nature of recent warming, the recent syntheses have also underlined clear differences among the various reconstructions as well as among the simulations, even at the hemispheric scale [e.g., Jansen *et al.*, 2007; Jones *et al.*, 2009]. At the regional scale, the uncertainties are typically

<sup>1</sup>Université Catholique de Louvain, Institut d’Astronomie et de Géophysique G. Lemaître, Louvain-la-Neuve, Belgium.

<sup>2</sup>Department of Meteorology and Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania, USA.

<sup>3</sup>Faculty of Earth and Life Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.

<sup>4</sup>International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii, USA.

even greater due to the paucity of reliable proxy data in some regions. Furthermore, internal variability (defined here as the variability that is purely related to the internal dynamics of the system that would be present even in the absence of any natural or anthropogenic forcing) plays an important role at this scale [e.g., *Goosse et al.*, 2005a; *Tett et al.*, 2007]. As a consequence, when comparing various simulations or model results with observations, it is often impossible to determine if a difference is related to model biases or is simply due to the different realizations of internal variability in model simulations and reality. When ensemble simulations are employed to sample the plausible range of internal variability of the system over the past millennium, one typically finds that the range of the ensemble is large enough during any particular time interval that model/paleodata comparisons do not provide especially strong constraints on the model physics [*Goosse et al.*, 2005a, 2006b]. Although the precise relative contributions of forced and internal variability are difficult to determine for any one climate event of the past millennium, comparisons of climate reconstructions for past centuries with the modeled mean response of the climate to estimated natural (solar and volcanic) external forcing have provided some potentially important clues regarding past forced responses of climate [*Haigh*, 1996; *Robock*, 2000; *Shindell et al.*, 2001, 2004; *Waple et al.*, 2002].

[4] Several studies have sought to isolate the reasons for the discrepancies among various published proxy climate reconstructions and, in particular, to disentangle the role of the selection and the treatment of the proxy records and of the statistical method used to obtain the reconstructions [e.g., *Mann et al.*, 2005, 2007; *Rutherford et al.*, 2005; *Juckes et al.*, 2007; *Lee et al.*, 2008; *Jones et al.*, 2009]. On the other hand, the differences in model simulation results have typically been attributed to differences in the applied forcing and in model characteristics such as the equilibrium climate sensitivity [e.g., *Goosse et al.*, 2005b; *Osborn et al.*, 2006] or in the model representation of the natural modes of climate variability. Such studies are essential, as they provide a valuable opportunity to reduce uncertainties in past climate reconstructions and in our understanding of the causes of past climate changes. In particular, they are expected to provide guidance on how to further improve model physics.

[5] A complementary approach to reducing uncertainty is to directly combine the information from proxy records, the estimates of past climate forcing, and the physics embodied in climate models. This methodology, often referred to as data assimilation in meteorology and climatology, has been successfully applied in many fields but is still not standard in paleoclimatology, in particular when analyzing the climate of the past millennium [e.g., *Goosse et al.*, 2008; *Widman et al.*, 2009].

[6] Data assimilation nonetheless is a potentially powerful tool for analyzing the climate changes over past centuries. It can provide alternative estimates of climate changes in a manner equivalent to how model reanalyses are used to fill in gaps in the instrumental record of the past 50–60 years [*Kalnay et al.*, 1996; *Uppala et al.*, 2005]. These physically constrained estimates can be then compared with

reconstructions using complementary, traditional statistical approaches. The data assimilation approach potentially makes use of the same underlying proxy data as the statistical approach, but it does so in a very different manner. The physics of the climate system, as represented by the model equation, is applied formally as a constraint, ensuring that the estimated climate fields are consistent with model physics. The data assimilation approach thus offers the potential advantage over purely statistical approaches of ensuring dynamical consistency in the resulting assessment of past climate change, particularly in the context of using a coupled climate model. Indeed, it is then possible to obtain dynamically consistent states of climate system components that are difficult to constrain directly on long time scales because of a lack of the requisite data (e.g., the deep ocean or the sea ice) using more easily available data, such as surface temperature proxies [e.g., *Goosse et al.*, 2009].

[7] In the present framework, we will refer to these dynamically consistent estimates as “reconstructions.” However, to highlight the difference with purely statistical approaches to proxy climate reconstructions that are based on a fundamentally different conceptual foundation, we will call our reconstructions DALE (for data assimilation using large ensembles; see section 2 for the description of the data assimilation technique) reconstructions.

[8] Simulations with data assimilation provide information on variables and at locations that are not directly available from proxy data, filling the gaps between sparse observational data. All the important variables simulated by the model can thus be analyzed. This can naturally lead to the proposition of mechanisms to explain the observed changes that are consistent among all the elements of the climate system and are in agreement with the evidence recorded in the proxies used to constrain the model results. The mechanism can then be tested using independent proxies when available or to suggest particularly interesting locations where additional proxies would be very useful. Data assimilation in this manner provides a useful means for testing specific hypotheses. It is for instance possible to force the model to maintain a positive North Atlantic Oscillation index or a strong meridional overturning circulation in the Atlantic during several years or decades in order to test if the simulated changes are compatible with the proxy evidence during a particular period. If this is the case, such circulation changes provide, then, one possible explanation for the observed anomalies during this period.

[9] Of course, data assimilation also has its own limitations. In particular, there is no guarantee that it will be technically possible to find a system state that is both compatible with model physics and faithful to the observational data if the observational data happen to contain real structures that are more complex than can be described by the simplified physics of the model. Data assimilation techniques are also quite computationally intensive, so that applying them over a long period such as the past millennium can be a daunting computational task. Furthermore, transferring the assimilation techniques employed to analyze modern instrumental climate data to a framework for analyzing proxy evidence of longer-term climate variability requires a number of adaptations, each of which must be

tested and validated. These adaptations must take into account various challenging characteristics of paleoclimate proxy data, including the spatially sparse nature of the data networks, the relatively low signal-to-noise ratios of the data, and the coarse and often variable temporal sampling provided by the data.

[10] It is generally not possible to determine the precise cause of a particular climate event in a simulation with data assimilation. In particular, we cannot in general determine whether the climate state obtained was linked to the forcing or the data constraints. If the forced response over a particular period can be determined through other means, however, then the likely relative contribution of the forced response and internal variability can be ascertained.

[11] Independently testing the quality of a climate model by investigating its ability to reproduce observed changes is more challenging with data assimilation than without it, as some information from the data has necessarily already been used to constrain model results. However, analyzing the innovation brought by data assimilation can help in finding some model biases.

[12] Finally, when using data assimilation, model biases can significantly impact the quality of the reconstructions and the validity of the mechanisms inferred to explain the data. Analyzing jointly the results of complementary approaches including (1) climate reconstructions based purely on statistical approaches applied to climate proxy networks, (2) forced model simulation results without data assimilation, and (3) “reconstructions” based on forced model simulations with data assimilation can help to diagnose those features in the data and reconstructions which are likely robust and those which may be an artifact of various sorts of biases.

[13] Because the use of data assimilation in paleoclimatology is relatively recent, the potential range of new applications to this area is wide, and one can expect considerable progress in this area in years to come. Readers interested in a general overview of applications of data assimilation to studying the climate of past centuries, including the analysis of particular periods and the testing of specific hypotheses, are referred to other recent work in this area [e.g., *van der Schrier and Barkmeijer*, 2005; *Goosse et al.*, 2006b, 2008, 2009; *Crespin et al.*, 2009; *Widmann et al.*, 2009].

[14] The present study focuses on the reconstruction of past temperature changes at the hemispheric and continental or ocean basin scales. The data assimilation technique employed is as described by *Goosse et al.* [2006a] but with several more recent innovations as described below. The goal remains to select among a large ensemble of model simulations that which is closest to “reality,” in the sense indicated by the available paleoclimate proxy data. *Goosse et al.* [2006a] selected this best model analog a posteriori from an existing ensemble; that is, the simulations were completed before any comparison with data was performed. This approach had the advantage of computational efficiency, lending itself to the generation of large ensembles at low computational cost. On the other hand, the potential lack of compatibility between best model analogs selected for different periods made the interpretation of the results challenging at times. The revised procedure used in the present studies alleviates this problem, as dynamical con-

sistency is ensured by generating a new ensemble at each step of the assimilation procedure, starting from the best model analog selected for the previous period (see section 2 for more details on the method).

[15] By contrast with *Goosse et al.* [2006a], in which an ad hoc compilation of only 12 proxy time series were used to constrain the model evolution, a more recent and more comprehensive compilation of temperature proxy data extending over the past millennium [*Mann et al.*, 2008] is employed here. We are thus able to test the data assimilation method for the first time with a large fraction of the available proxy data. The proxy records selected pass a screening against modern instrumental data for a local temperature signal and are available at least at decadal resolution (see *Mann et al.* [2008] for more details). All data are smoothed at the decadal time scale to ensure uniformity of temporal resolution in the data assimilation process and allow for direct comparison against statistical paleoclimate reconstructions using these same data.

[16] A small subset of the experiments discussed here was described previously by *Crespin et al.* [2009]. The latter study focused on specific interval of apparent Arctic warmth during the 15th century. Here, our goal is different, as we are interested in investigating additional details of the methodology and, in particular, the ability of simulations *with data assimilation* to yield useful reconstructions of past temperature changes.

[17] Specifically, we seek to address three principal questions (discussed in more detail in sections 3–5, respectively):

[18] 1. Using the data assimilation technique, is it possible to obtain model states that reasonably follow the trajectories indicated by selected proxy data over the past 600 years? In other words, can we find a set of climate states consistent with both the physics of the model and the empirical information present in individual proxy records?

[19] 2. Does the additional information contained within the presently available proxy data allow for a reduction in the potential errors arising from deficiencies in the model physics and estimated external forcing histories? While such questions have been answered affirmatively in the context of modern data assimilation [e.g., *Kalnay*, 2003], they remain open within the context of paleoclimate data assimilation.

[20] 3. Do the reconstructions obtained through this data assimilation exercise compare favorably with observed changes over the more recent, historical period (i.e., the last 150 years) and with other reconstructions based on other approaches?

[21] The first two questions are largely independent of the quality of the proxy data. The technique should be able to follow any reasonable time series, and this constraint should reduce the uncertainty compared to simulations without data assimilation. For the third question, however, the quality of the signal recorded in the proxies clearly plays a crucial role. In this study, we focus only on whether the reconstructions using the current proxy data set employed (that of *Mann et al.* [2008]) provide useful constraint on past patterns of surface temperature change. We acknowledge that this is just one element of a larger problem, which involves, in addition, the attempt to understand the detailed influence of the proxy data set on the reconstructions that are obtained. This wider subject was briefly addressed in a

previous study [Goosse *et al.*, 2006a] and is the subject of ongoing work.

## 2. Experimental Design

[22] The climate model we used is named LOVECLIM 1.1, an acronym of the acronyms of its components: LOch (Liège Ocean Carbon Heteronomous)-Vocode (VEgetation COntinuous DEscription model)-Ecbilt (model name, not acronym)-CLIO (Coupled Large-scale Ice Ocean)-agIsm (Antarctic and Greenland Ice Sheet Model); this same model was used by Goosse *et al.* [2007]. LOVECLIM [Driesschaert *et al.*, 2007] is a three-dimensional Earth system model of intermediate complexity that includes representations of the atmosphere, the ocean and sea ice, the land surface (including vegetation), the ice sheets, and the carbon cycle. In the present study, the ice sheet and carbon cycle components are not active and are thus not described. The atmospheric component is ECBILT2 [Opsteegh *et al.*, 1998], a quasi-geostrophic model with a resolution of  $5.6^\circ$  in latitude,  $5.6^\circ$  in longitude, and three levels in the vertical. To close the momentum budget near the equator, a parameterization of ageostrophic terms is included. This parameterization improves the representation of tropical dynamics, including the Hadley cell circulation, although the amplitude of variability in tropical regions and the extratropical response to tropical sea surface temperature anomalies remain artificially low in LOVECLIM [Opsteegh *et al.*, 1998]. The oceanic component is CLIO3 [Goosse and Fichefet, 1999], which is made up of an ocean general circulation model coupled to a comprehensive thermodynamic-dynamic sea ice model. Its horizontal resolution is  $3^\circ \times 3^\circ$ , and there are 20 levels in the ocean. The vegetation model VECODE simulates the dynamics of two main terrestrial plant functional types, trees and grasses, as well as desert [Brovkin *et al.*, 2002]. Its resolution is the same as in ECBILT. (More information about the model and a complete list of references are available at <http://www.astr.ucl.ac.be/index.php?page=LOVECLIM%40Description>).

[23] The initial conditions for the simulations covering the last millennium are derived from previous numerical experiments covering the entire Holocene period [Goosse *et al.*, 2007]. Long-term changes in orbital parameters follow Berger [1978], and the long-term evolution of greenhouse gas concentrations is prescribed. The influence of modern (A.D. 1850–2000) anthropogenic sulfate aerosols is represented through a modification of surface albedo [Charlson *et al.*, 1991]. Forcing by anthropogenic land-use change (including both surface albedo and surface evaporation and water storage) is applied as by Goosse *et al.* [2005a], following Ramankutty and Foley [1999]. Finally, natural external forcing due to changes in solar irradiance and explosive volcanism are prescribed following the reconstructions of Muscheler *et al.* [2007] and Crowley *et al.* [2003], respectively. In the standard run, the total solar irradiance changes have been scaled to provide an increase of  $1 \text{ W m}^{-2}$  between the Maunder minimum (late 17th century) and the late 20th century. This corresponds to roughly a threefold reduction in amplitude in comparison with previous simulations conducted with the LOVECLIM model [e.g., Goosse *et al.*, 2005b, 2006b] but is in better agreement with recent reassessments [Lean *et al.*, 2002; Foukal *et al.*, 2006].

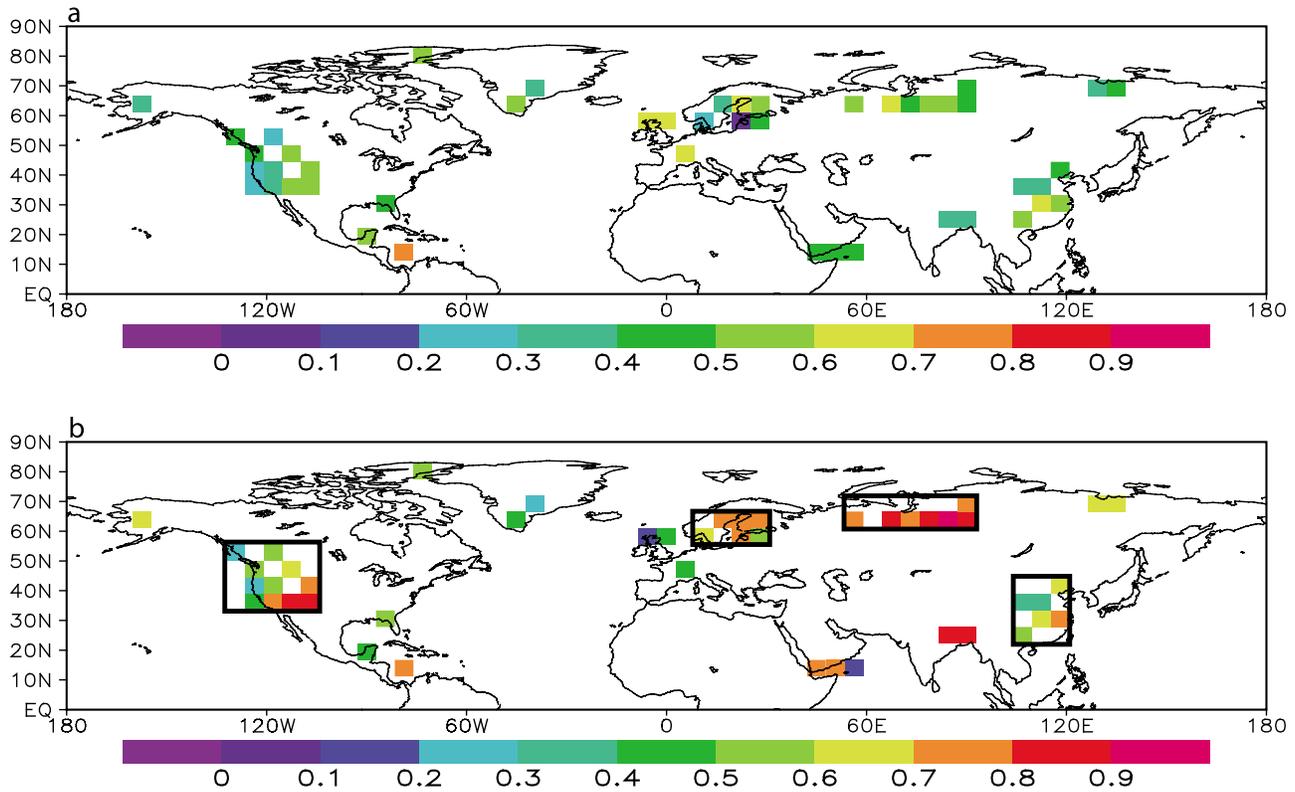
[24] In the majority of the simulations presented here, the model is forced to follow estimates of surface temperature changes derived from proxy records, using an updated version of the data assimilation technique described by Goosse *et al.* [2006a] [see also Collins, 2003]. The method is briefly summarized here as follows. For the first year (A.D. 1000), a large ensemble of initial conditions is generated by introducing small perturbations in the quasi-geostrophic potential vorticity field of the atmosphere. The perturbations are applied through a simple procedure: for each member of the ensemble, the quasi-geostrophic potential vorticity is multiplied by a positive constant within the range 0.95–1.05, roughly preserving the large-scale coherence of the field.

[25] Short simulations of duration 1–20 years are then performed starting from this set of initial conditions. Owing to the chaotic nature of the atmospheric dynamics, the differences between the various ensemble members grow quickly, leading to fundamentally different atmospheric circulation states within just a few days. After the simulations have been performed, the model results are compared with the available observations averaged over the simulation interval, using a cost function  $CF$ :

$$CF_k(t) = \sqrt{\sum_{i=1}^n w_i (F_{\text{obs}}(t) - F_{\text{mod}}^k(t))^2}, \quad (1)$$

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 observations used in the model/data comparison,  $F_{\text{obs}}$  is the observed value of the variable  $F$ , and  $F_{\text{mod}}^k$  is the value of  $F$  simulated in model experiment  $k$  at the same location.  $w_i$  is a weight factor. The member of the ensemble closest to the observations, i.e., the one that minimizes the cost function, is selected as representative for that particular period. The state obtained at the end of this short simulation is then used as the initial condition for the subsequent simulation. The procedure is repeated until the termination date of the analysis (year 1995) is reached.

[26] Our DALE reconstructions employ an ensemble of roughly 100 (96 to be precise: 3 times 32 simulations performed in parallel on 32 processors) realizations. Ideally, our ensemble should cover *all* states potentially accessible to the climate system during the period under investigation, thus ensuring that a suitable analog state to that indicated by the observations can be found among the ensemble members. However, fully exploring this state space would require a prohibitively high number of simulations [e.g., Lorenz, 1969; Van den Dool, 1994; Nicolis, 1998; Snyder *et al.*, 2008]. On the other hand, previous tests [Goosse *et al.*, 2006a] have demonstrated that an ensemble of roughly 100 realizations, still quite affordable from a computational viewpoint for a model such as LOVECLIM, provides a reasonable sampling of the range of substantially different potential large-scale atmospheric states. As a consequence, it is possible even with this modest ensemble to find realizations that reproduce the main features of the large-scale climate indicated by the available proxy records, addressing in part the first question posed in section 1. Precisely what ensemble size is truly “optimal” is a different issue, which we intend to address in future work.



**Figure 1.** (a) Correlation between the proxy data selected here [Mann *et al.*, 2008] and instrumental data compiled in the HadCRUT3 data set [Brohan *et al.*, 2006] over the period 1850–1995. (b) Correlation between the surface temperature averaged over the 11 simulations using data assimilation (see Table 1) and the proxy data time series over the period 1400–1995. The analyses have been performed using annual mean temperatures and decadal smoothed time series for both proxy data and model results. The boxes in Figure 1b represent the regions for which the averages are presented in Figure 2.

[27] As in the work of Crespin *et al.* [2009], the proxy data used to constrain model results comprise a set of 56 proxy series derived from a recent compilation of Mann *et al.* [2008], processing the raw proxy time series exactly as proposed in their work except for the additional interpolation on our model grid (see below). To avoid the logistical challenges of a time-dependent observational network, we have used a constant number of proxies for the whole period investigated, selecting thus only the ones available back to the year 1400 and which extend through 1995 (the geographical distribution is shown in Figure 1). Data assimilation is performed starting in 1000, using the proxies available at that time, but the first 400 years are used for spin-up, and only the results over the period 1400–1995 are analyzed below. Furthermore, because of data availability, we will analyze only results in the Northern Hemisphere.

[28] The proxy data set consists of tree rings, ice cores, corals, speleothems, and some lake sediments and historical documents (see Mann *et al.* [2008] and their supplementary information for further details). Through a statistical screening analysis described by Mann *et al.* [2008], only those proxies with a significant correlation with local instrumental annual mean surface temperatures during the modern calibration interval (1850–1995) are retained. Records with only decadal resolution were first interpolated to annual resolution. As described by Mann *et al.* [2008], all proxy

records were decadal smoothed (using a low-pass filter with half-power cutoff at  $f=0.1$  cycle  $\text{yr}^{-1}$ ) and centered to have zero mean over the modern calibration interval. The proxies were then averaged onto the  $5^\circ$  latitude  $\times$  longitude grid of the available HadCRUT3 instrumental gridbox temperature [Brohan *et al.*, 2006] data set and scaled to the same mean and decadal standard deviation as the corresponding instrumental surface temperature gridbox over the calibration period. The time series obtained can then be denoted as gridbox composite-plus-scale (CPS) reconstructions for the procedure used to generate them. Finally, the gridbox CPS reconstructions were spatially interpolated for consistency with our model grid, using the nearest-neighbor method. Despite the modifications that have been applied to the proxies during this procedure, we will refer here to those local temperature reconstructions as the proxies or the proxy records. However, we should keep in mind that we are not using directly the proxy time series but temperature reconstructions derived from those proxies. The gridbox CPS reconstructions can also be hemispherically averaged to yield a Northern Hemisphere mean CPS temperature reconstruction, as in the work of Mann *et al.* [2008].

[29] Because the proxy CPS gridbox temperature reconstructions are decadal smoothed, we cannot expect to constrain the simulated interannual variability. Before

**Table 1.** Simulations Covering the Last Millennium, Performed With Data Assimilation

Name of Experiment	Group	Description
MNCY1	ASSIM	All the weights are equal in the computation of the cost function; the evaluation of the cost function is computed on 1 year averages
MCY1	ASSIM	Weights in the cost function are proportional to the correlation between proxies and recent temperature observations; the evaluation of the cost function is computed on 1 year averages
	SEN-ASSIM	
	FORC-ASSIM	
MCY5	ASSIM	The same as MCY1, except that the evaluation of the cost function is computed on 5 year averages
MCY10	ASSIM	The same as MCY1, except that the evaluation of the cost function is computed on 10 year averages
MCY20	ASSIM	The same as MCY1, except that the evaluation of the cost function is computed on 20 year averages
MSOL	FORC-ASSIM	The same as MCY1, except that the solar forcing has been multiplied by a factor of 3
MVOL	FORC-ASSIM	The same as MCY1, except that the volcanic forcing has been multiplied by a factor of 0.5 in one third of the experiments of the ensemble and by 2 in another third
MCSE1	SEN-ASSIM	The same as MCY1, except that a different parameter set is used, leading to a model climate sensitivity of 1.6°C instead of 2.6°C
MCSE2	SEN-ASSIM	The same as MCY1, except that a different parameter set is used, leading to a model climate sensitivity of 2.1°C instead of 2.6°C
MCSE4	SEN-ASSIM	The same as MCY1, except that a different parameter set is used, leading to a model climate sensitivity of 3.2°C instead of 2.6°C
MCSE5	SEN-ASSIM	The same as MCY1, except that a different parameter set is used, leading to a model climate sensitivity of 3.8°C instead of 2.6°C

analyzing the results of our simulations, the model time series also therefore have to be decadal smoothed. This lack of constraint on interannual variability is, of course, not a limitation of the method but rather a limitation due to the character of the available proxy data. Indeed, with a different data set (surface temperature obtained from thermometers) but the same methodology, *Goosse et al.* [2009] successfully reproduced the year-to-year variations in temperature at high southern latitudes over the last 50 years [e.g., *Goosse et al.*, 2009, Figure 2].

[30] An important goal of the present study is to estimate the influence of the large uncertainties that are still present in the way that the data assimilation method is implemented, in the model equations and parameters, and in the external forcing that was prevailing during the past millennium. This is done here by performing 11 different experiments with data assimilation, all covering the past millennium and all consisting of 96 ensemble members. Those experiments with data assimilation can be divided into three different groups (Table 1). For the first group, the only changes between the various experiments are related to the data assimilation procedure (group ASSIM, five experiments). In the first experiment (MNCY1), the weight factors  $w_i$  are the same for all the proxy records, and the cost function is evaluated for 1-year averages. Previous tests [*Goosse et al.*, 2006a] have shown that such a short averaging period is adequate even if we are interested in decadal to multicentennial-scale variability as is the case here. In all the other simulations, the weight factors  $w_i$  are proportional to the correlation between the proxy records and the observations of temperature obtained during the instrumental period in order to give a larger weight to proxies that are judged as more reliable. The other experiments of the group differ only regarding the averaging period selected in the computation of the cost function, which is set to 1 (MCY1), 5 (MCY5), 10 (MCY10), or 20 (MCY20) years.

[31] The second group (FORC-ASSIM, three experiments) is devoted to the analysis of the role of the forcing, the data assimilation procedure being the same as for MCY1. For the solar forcing, the largest uncertainty lies in

the scaling of the total irradiance anomalies. In experiment MSOL, we have multiplied them by a factor 3 in order to have an amplitude similar to the one applied in previous experiments with the model. While a single factor is justifiable for the solar forcing, the error in the volcanic forcing can be different for each volcanic event. In MVOL, we have thus divided the 96-member ensemble into three sub-ensembles each of 32 simulations; the first one has the volcanic forcing multiplied by a factor 0.5 compared to MCY1, the second one has it multiplied by a factor 2, and the last group uses the standard forcing.

[32] The goal of the last group of simulations (SEN-ASSIM, five experiments) is to test the influence of model uncertainties. To do so, we have performed simulations equivalent to MCY1 but using different values for some model parameters such as the ocean vertical diffusivity, surface albedo, certain parameters that govern the longwave radiative scheme, and so on. All the parameter sets selected provide a reasonable fit to the present-day climate but clearly different model responses to perturbations [*Goosse et al.*, 2007]. In particular, the selection of these parameter sets leads to different climate sensitivities: MCSE1, MCSE2, MCSE4, MCSE5, and MCY1, corresponding to experiments E1, E2, E4, E5, and E3 of *Goosse et al.* [2007], have climate sensitivity of 1.6°C, 2.1°C, 3.2°C, 3.8°C, and 2.6°C, respectively. Ideally, one might use completely different climate models to more thoroughly test the influence of the fundamental uncertainties in our current physical representations of the climate system (often referred to as structural uncertainties), but such a highly ambitious undertaking is well beyond the scope of our current analysis.

[33] Our objective here is to analyze, within the framework of the LOVECLIM model, these three different types of simulations to determine whether uncertainties in (1) the assimilation methodology, (2) the specified external forcing, or (3) physical parameters of the model most significantly impact the quality of our reconstruction, thus guiding where the greatest attention should be given in future work seeking to improve the skill in paleoclimate data assimilation. In our case, the number of simulations used to explore the sensi-

**Table 2.** Simulations Covering the Last Millennium, Performed Without Data Assimilation

Group	Description
INT-NA	Ensemble of five simulations with the same parameter sets and forcings as in experiment MCY1 (see Table 1), which corresponds to a climate sensitivity of 2.6°C
SENSF-NA	Ensemble of simulations using parameters sets corresponding to climate sensitivities of 1.6°C, 2.1°C, 2.6°C, 3.2°C, and 3.8°C driven by the standard forcings as well as a simulation using the version with a climate sensitivity of 2.6°C and a solar forcing multiplied by a factor 3 (as in MSOL; see Table 1)

tivity in each of these three areas is relatively modest but is sufficient to address a number of interesting sensitivity issues (see section 4). Nonetheless, a considerably larger number of simulations would be required to estimate with confidence such issues as the precise optimal parameter choices. In our analysis of the full set of individual simulations performed, we were not, for example, able to isolate a simulation or small set of simulations that were clearly optimal in their ability to reproduce variations for all regions investigated and for all periods of the past 600 years.

[34] Simulations *without* data assimilation (denoted NA) were also performed for comparison (Table 2). A comparison of simulations both with and without data assimilation allows us to quantify the improvements introduced by the data assimilation in the present framework. All those experiments (with and without data assimilation) represent the equivalent of more than one million years of simulation. Performing such tests at present is thus possible only with relatively fast Earth system models such as LOVECLIM.

### 3. Constraining Model Results to Follow Proxy Records

[35] Proxy data include both a climatic (e.g., “temperature”) signal and residual variance that is considered here to be “noise.” Furthermore, individual proxies can be influenced by a variety of local climate influences other than the larger-scale (i.e., gridbox-scale) temperature variations of interest. As a consequence, the correlation between the proxy series and associated instrumental gridbox temperatures is generally substantially less than unity. For the proxies that successfully pass the *Mann et al.* [2008] screening procedure, correlation is typically between 0.3 and 0.8 (Figure 1a).

[36] The correlation between proxy records and model results with data assimilation is generally in the same range. Figure 1b illustrates this for the mean over the 11 simulations with data assimilation described in Table 1, but results are similar for individual simulations. The values in the latter case are lower, however, as the averaging procedure smoothes the quasi-random internal variability of the model that is not well constrained by the proxies (see section 6).

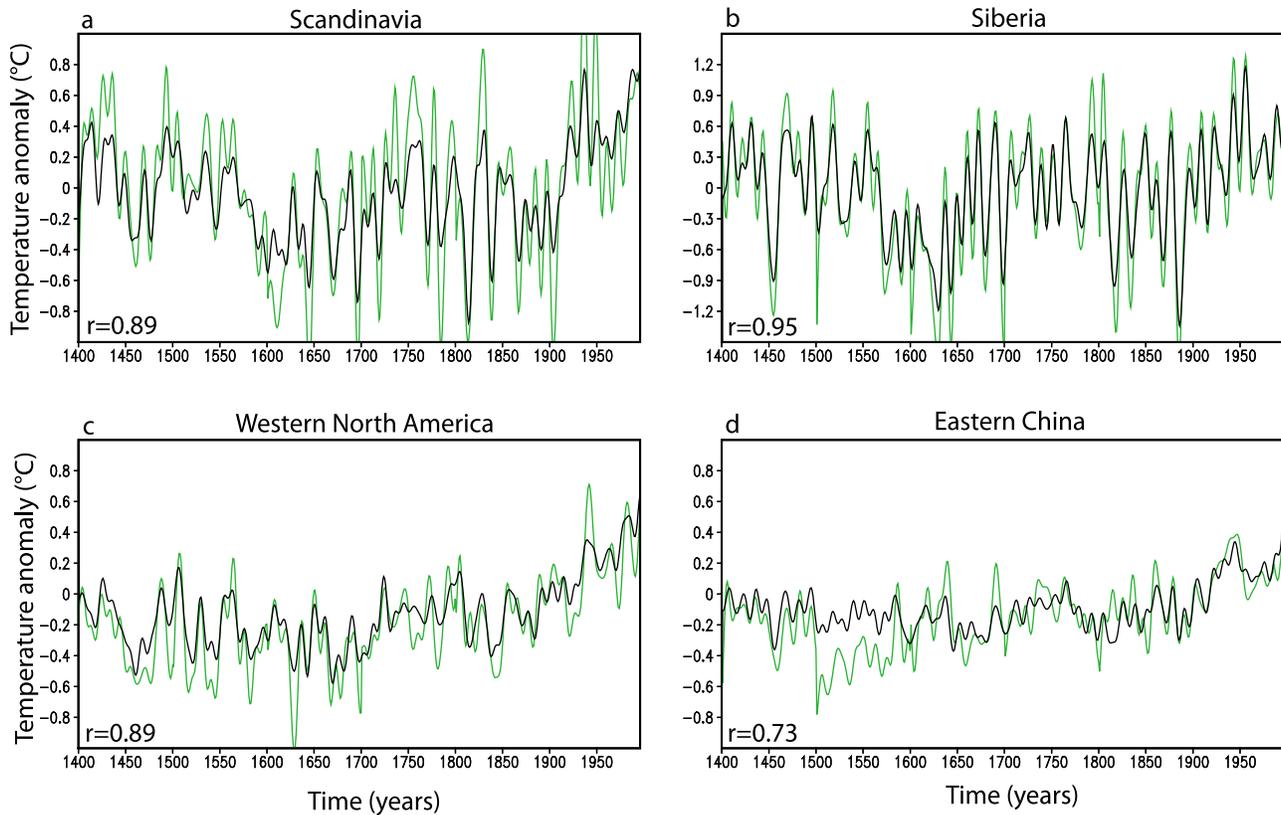
[37] The high correlations indicate that the data assimilation adequately forces the model to remain reasonably close to the proxy records, even at the grid scale. The correlation is particularly high in northern Siberia, reaching more than 0.9. As a consequence, in this region, the model reproduces both the climatic signal and the nonclimatic noise recorded in the proxies. The correlation is much lower in some other regions, such as Scotland or in the southeastern part of the Arabian peninsula.

[38] This comparison at the grid scale is informative, given that it is the gridbox scale at which the data assimilation procedure is implemented (equation (1)). The model results, however, are not best interpreted at this scale, given that the larger-scale features are the ones best captured by the data assimilation procedure. Furthermore, when the climate variation indicated by neighboring proxy records is not coherent (potentially because of data quality problems in one or more of the proxy records), the model will in general be unable to reproduce the resulting high-frequency apparent spatial structure. It is thus advisable to assess the compatibility of the DALE reconstructions with proxy records at regional or continental scales (i.e., the scales at which the coarse dynamics of a model such as LOVECLIM is most likely to be faithful). Accordingly, we performed comparisons between the proxy data and model assimilation results for four particular distinct regions where the proxy data density is high and the model constraint therefore likely best: Scandinavia, Siberia, eastern China, and western United States (see Figure 1b). In each of these cases, the model results were averaged only over the grid points where proxies are available.

[39] For all four regions, the level of agreement between the DALE reconstruction and the proxy temperature records themselves is quite good, both at decadal and longer time scales, although the amplitude of decadal variability appears to be somewhat underestimated in the average over all simulations (Figure 2). For Scandinavia, Siberia, and western United States, the correlation between the DALE reconstruction and proxy records at the regional scale is close to or even exceeds 0.9. The correlation is somewhat lower over China, primarily due to the mismatch over the interval A.D. 1500–1600, where the proxies indicate a substantial cooling not reproduced by the simulations. We note that the cooling in the regional proxy composite results from a large excursion in just one of the proxy records and could plausibly therefore be dismissed as a data quality problem.

### 4. Reducing the Influence of Uncertainties in the Model Physics and in the Forcing

[40] Before assessing the degree of constraint provided by data assimilation, it is first necessary to evaluate the sensitivity to other factors (model parameter choices and external forcing estimates) alone, through simulations *without* data assimilation. To do so, an ensemble of five simulations has been performed using parameters corresponding to experiment MCY1 to allow us to measure the magnitude of the model internal variability (hereafter referred to as INT-NA; Table 2). All those simulations are driven by the same forcing and differ only in their initial conditions, leading to the selection of a different sample of the model internal variability in each experiment. In the second group of simu-



**Figure 2.** Time series of the proxies (green) and the surface temperature averaged over the 11 simulations (see Table 1) using data assimilation (black) for the four boxes displayed in Figure 1 (in °C): (a) Scandinavia, (b) Siberia, (c) western North America, and (d) eastern China. The correlation between the time series over the years 1400–1995 is given in the lower left corner of each panel.

lations (hereafter referred to as SENSF-NA) different parameter sets and forcings are selected (the same as in the corresponding simulations performed *with* data assimilation; see Table 1). In evaluating the large-scale performance, we averaged the model results over the whole Northern Hemisphere (NHM), Europe (EUR, 0°E–40°E, 36°N–64°N), Asia (ASI, 45°E–135°E, 30°N–63°N), North America (AME, 230°E–300°E, 30°N–63°N), the Arctic (ARC, north of 69°N), and the North Atlantic Ocean (ATL, 300°E–360°E, 30°N–63°N). All of these regions, except NHM, are located in the extratropics, where our model is expected to perform best, and where the majority of the available proxy data used as constraints in the simulations with assimilation are located (i.e., northward of 30°N).

[41] The spread within any given group of simulations was estimated by computing the root-mean-square error (RMSE; Figure 3) and the correlation between individual members of a group (Figure 4). The corresponding values are computed for all the possible pairs in a group before determining their mean and standard deviation.

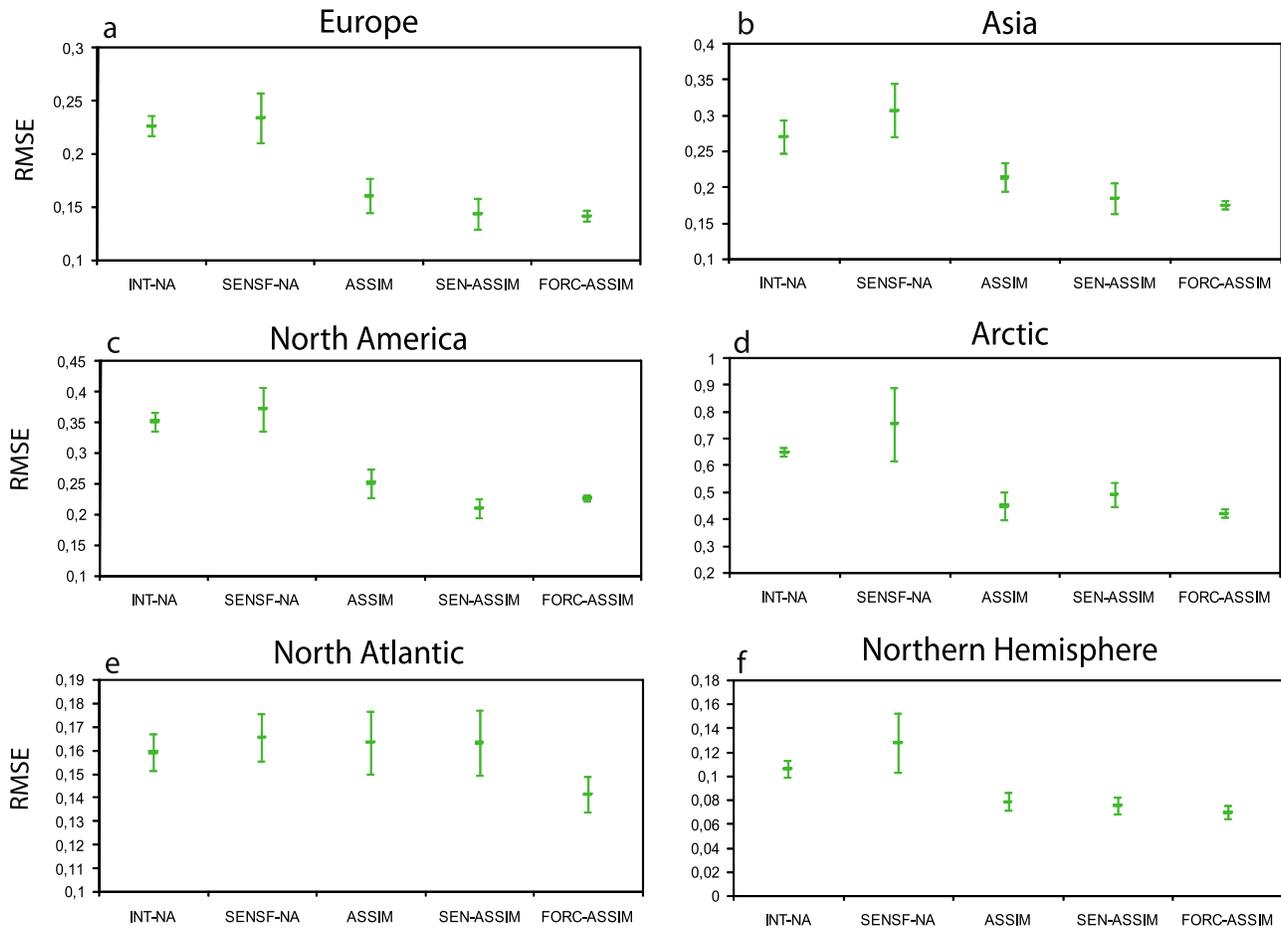
[42] In INT-NA, the RMSE has roughly the same magnitude for Europe, Asia, and America (between 0.2°C and 0.3°C). It is higher for the Arctic because of the large amplitude of the simulated variations in this region, while it is smaller over the Atlantic because of the small amplitude of the simulated changes in that region. The RMSE is also small for the Northern Hemisphere mean, and the correlation

between the different experiments is high due to the smaller influence of internal variability at the hemispheric versus continental scale [Goosse *et al.*, 2005a; Tett *et al.*, 2007].

[43] As expected, the RMSE is higher (between 10% and 20%, except for the Atlantic) in the simulations with varying model parameters and forcing (SENSF-NA) than in the experiments in which the only cause of difference between the simulations is the internal variability (INT-NA). Furthermore, the correlation generally decreases between INT-NA and SENSF-NA, although this is not clearly significant for all regions. Nevertheless, the changes between the two groups are not dramatic, confirming that internal variability alone contributes a substantial component of the observed spread in model results, even at the continental scale [Goosse *et al.*, 2005a; Tett *et al.*, 2007].

[44] For Europe, Asia, America, and the Arctic, the RMSE strongly decreases and correlation increases in the groups of simulations with data assimilation (Figures 3 and 4). A reduction of nearly 50% for the RMSE (Figure 3) and improvement of 0.15–0.4 in correlations is observed (Figure 4) in comparison with experiments without data assimilation. The data assimilation technique is thus clearly seen to provide improved agreement of model results with observations in those regions.

[45] The improvement is modestly less for the Northern Hemisphere mean temperature (decrease in RMSE of 0.02°C–0.06°C, increase in correlation of 0.06–0.17), but



**Figure 3.** Mean and standard deviation (in  $^{\circ}\text{C}$ ) of the root-mean-square error between the surface temperature simulated in the different experiments of five groups of simulations for (a) Europe, (b) Asia, (c) North America, (d) the Arctic, (e) the North Atlantic, and (f) the Northern Hemisphere. The five groups of simulations consists of an ensemble of five simulations with the same forcing and same model parameters without data assimilation (INT-NA), six simulations with different parameter sets and forcing without data assimilation (SENSF-NA), five simulations with data assimilation using different parameters in the data assimilation method (ASSIM), five simulations with data assimilation using different model parameters (SEN-ASSIM), and three simulations with data assimilation using different solar and volcanic forcing (FORC-ASSIM). The analyses have been performed using annual mean temperatures and decadal smoothed time series.

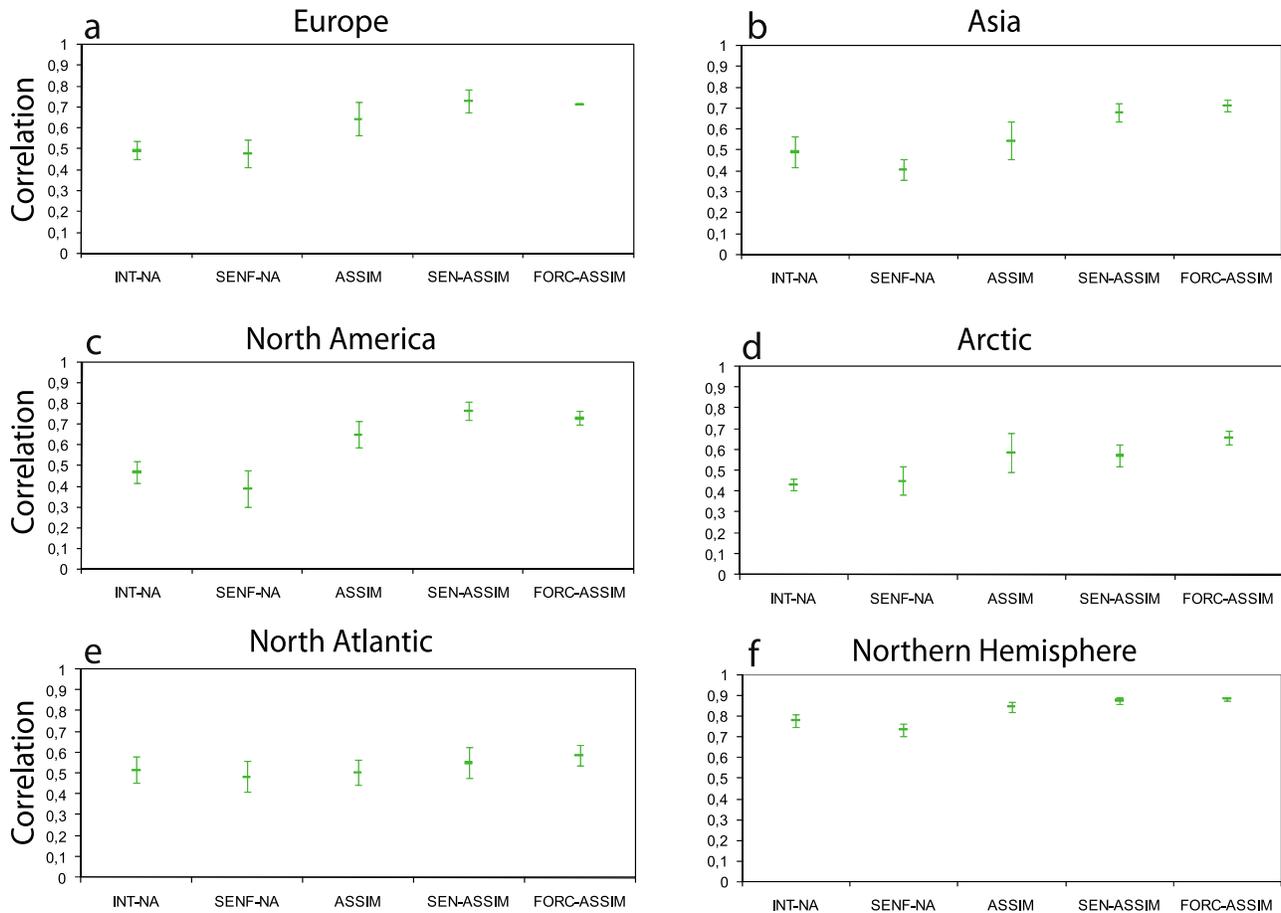
the uncertainties are smaller too in the experiments without data assimilation. By contrast, the data assimilation fails to reduce the model uncertainties for the North Atlantic Ocean. The performance of the data assimilation procedure in this regard is consistent with the relative abundance (nine or more) of proxy records available as constraints in all regions other than the North Atlantic Ocean, for which no proxy records are available (Figure 1).

[46] For all regions, both the RMSE (Figure 3) and correlation (Figure 4) are similar for the three groups of simulations with data assimilation. As a consequence, none of the three elements investigated (i.e., the parameters in the data assimilation method, the model parameters used, and the external forcing estimates) appears solely critical to the success of the method. The simulations testing the role of the parameters in the data assimilation method (ASSIM) have the highest mean RMSE and lowest correlation for

nearly all the regions, but the differences in the skill measures between the three groups were found to be small compared to the corresponding within-group standard deviations.

## 5. Reconstructing Past Temperature Changes

[47] Our DALE reconstructions are not directly constrained by instrumental data such as the HadCRUT3 data set [Brohan *et al.*, 2006]. However, our results are not strictly independent of those observations since HadCRUT3 has been used to select the proxies included in the data assimilation procedure and to calibrate them locally. Furthermore, during the development phase of a model, a more or less explicit criterion for some choices in model parameterizations or in parameters is a good fit between model results and observations. Although no systematic calibration of the model has been performed here, we cannot rule out the possibility that such factors might have played some role



**Figure 4.** Same as Figure 3, for the correlation between different simulations in a group.

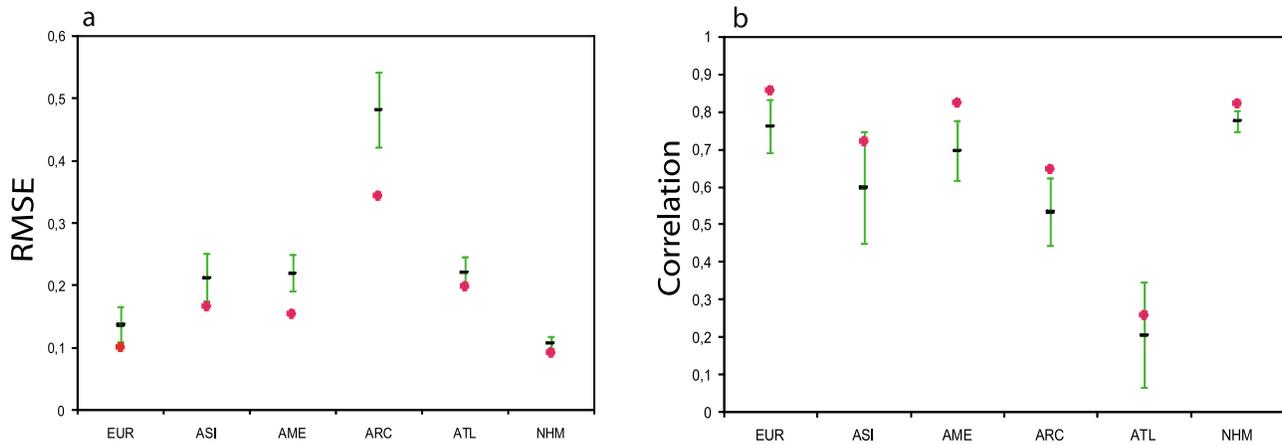
during the model development phase and hence may have produced some artificial hindcast skill.

[48] Nevertheless, instrumental temperature observations at continental or oceanic basin scale can be considered as sufficiently independent of our model to be useful for diagnosing the quality of our DALE reconstructions. In our methodology, there is no calibration or validation phase in the conventional sense, as temperature observations at continental scale are never used in the simulations with data assimilation. The whole period for which we have direct instrumental observations can thus be considered as the validation phase. As a consequence, we used the RMSE and the correlation with HadCRUT3 gridbox temperatures over the period 1850–1995 to diagnose the skill of the different simulations (Figure 5).

[49] In all the experiments, the RMSE is low and the correlation with HadCRUT3 is high for EUR, ASI, AME, and NHM (Figure 5). The agreement between the DALE reconstructions is particularly good on long time scales, while the quality of the reconstruction is lower at decadal timescales (Figure 6). In the Arctic, the RMSE is higher and the correlation with HadCRUT3 is lower than in the other regions analyzed. Nevertheless, the model is still able to reproduce the main features, such as the substantial warming trend between 1910 and 1940 followed by a cooling and a warming again after 1980 (Figure 6d). The simulation with data assimilation fails to reproduce the multidecadal varia-

tions observed in the North Atlantic (Figure 6e), leading to a very poor correlation between model results and observations in the region for all experiments.

[50] The average over the 11 simulations with data assimilation (shown in red on Figure 5) is in general better than any of the individual experiments (range shown in green on Figure 5), with, for instance, correlation with HadCRUT3 data higher than 0.8 for Europe, America, and Northern Hemisphere mean. As mentioned in section 3, each of the 11 simulations with data assimilation still contains a significant residual noise that is not directly constrained by the proxy data. In this aspect, the method can be considered as nonoptimal, as it is desirable that this noise is as small as possible. The problem here could be related to the methodology itself, the regional distribution of the proxies, or a too small ensemble size in each of the simulations with data assimilation. However, if we make the hypothesis that the noise can be considered as uncorrelated between the 11 experiments with data assimilation, averaging over a sizable (e.g., in our case, 11) number of samples would significantly reduce the magnitude of this noise. It would thus bring the mean of the reconstructions into better agreement with the observations, as seen in Figure 5. This reasoning based on uncorrelated noise among the 11 experiments with data assimilation is valid only if there is no systematic bias in those experiments. This appears not to be the case for the North Atlantic since the average does not



**Figure 5.** (a) Root-mean-square error (RMSE, in °C) and (b) correlation between the surface temperature obtained in simulations using data assimilation and HadCRUT3 observations [Brohan *et al.*, 2006] over the period 1850-1995 for averages over Europe, Asia, America, the Arctic, the North Atlantic, and the Northern Hemisphere. For each region, the mean and standard deviation of RMSE and correlation for all the simulations with data assimilation are displayed (green) as well as the RMSE and correlation between the mean over the 11 simulations with data assimilation and HadCRUT3 observations (red).

improve the results. For this region, the differences between the simulations with data assimilations and observations are thus not due to some noise not properly constrained by the method but to a more fundamental problem, as discussed in the next section.

[51] The uncertainty of our reconstructions can be derived from the computation of the RMSE discussed above. As expected, the observations generally fall within the mean of all the simulations with data assimilation plus and minus one RMSE of this average compared to observations (Figure 6). An alternative option is to consider that the uncertainty of the DALE reconstructions equals the standard deviation of the ensemble. The advantage of this procedure is that this estimate is obtained using only the information provided by the ensemble itself. The uncertainty can thus be estimated in this way even if no independent data are available [e.g., Goosse *et al.*, 2009]. Furthermore, this measure can be obtained for any particular time and can thus help determine if the uncertainty is higher or lower during some periods. Finally, if the number of proxies changes with time, this reduces the strength of the constraint and thus directly affects the uncertainty estimated by the standard deviation of the ensemble. On the other hand, the RMSE computed from independent data over the last 150 years provides a bulk estimate, valid only for the number of proxies in this validation period. It cannot be used in a reasonable way for earlier times if the number of proxies is changing. However, the standard deviation of the ensemble is based only on simulation results and can thus be biased, as it basically represents the level of internally generated climate variability.

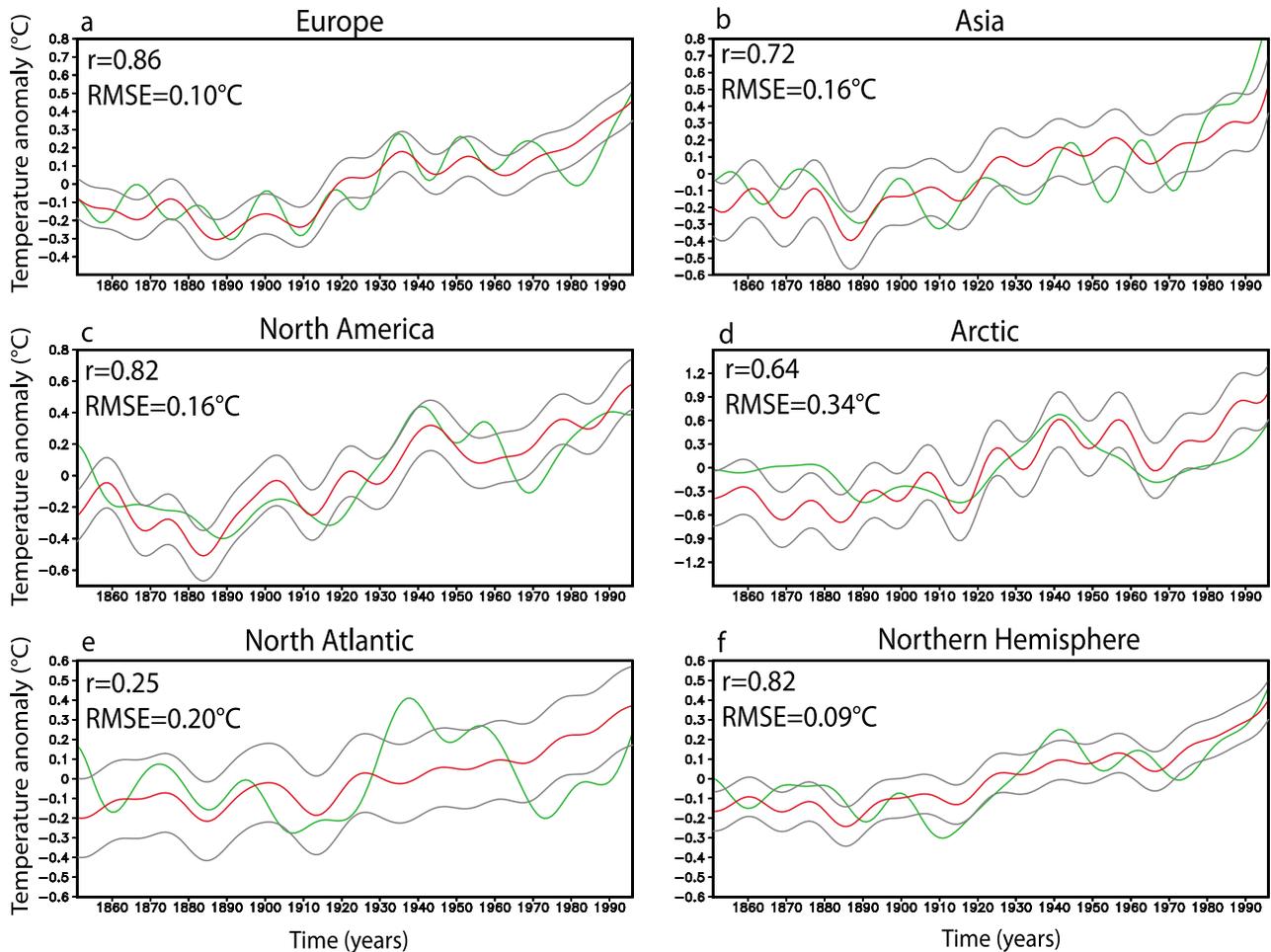
[52] Over land areas (Europe, Asia, America) and for the Arctic, both methods provide uncertainties that are close to each other (Figure 7). In those regions, the scatter of the simulations could thus be used to extend the estimate of uncertainty back to the beginning of the simulations in year 1000. Besides, for the Atlantic and for the Northern Hemisphere mean, the standard deviation of the simulations with

data assimilation is much lower than the RMSE of their mean compared to instrumental observations (blue line in Figure 7). As a consequence, for those regions, estimating the uncertainty from the scatter of our simulation would greatly underestimate it. This result also has important implications for the design of the data assimilation procedure, as discussed in the next section.

[53] Finally, our reconstructions can be compared over the last 600 years with previously published reconstructions based on statistical approaches (Figure 8). For the Northern Hemisphere mean, the agreement with the CPS reconstruction of Mann *et al.* [2008], which uses the same set of proxies, is very good. The correlation is equal to 0.85, and the RMSE equals 0.08°C. This is very similar to the values obtained when comparing our results with observations. Mann *et al.* [2008] demonstrate favorable comparisons between their CPS reconstruction and a number of other proxy-based reconstructions of Northern Hemisphere mean temperature, all of which largely agree within estimated uncertainties over the past 600 years.

[54] To our knowledge, no reconstruction is available for America, Asia, and the North Atlantic. For the first two regions, our results provide a basis for future analysis. The DALE reconstruction for the North Atlantic is also shown for completeness (Figure 8), but the large error bars there, due to the aforementioned poor results in the region, preclude any reliable conclusions for this region.

[55] For the Arctic, Overpeck *et al.* [1997] developed a reconstruction based on 29 proxies. Their original time series provide only relative changes, so it has been scaled here using instrumental data. Some of the proxies selected by Overpeck *et al.* [1997] are also included here. Their reconstruction is thus not independent of ours, but their network is larger than the one used here [see Figure 1 in Overpeck *et al.*, 1997]. Despite this difference, the correlation between the Overpeck *et al.* [1997] reconstruction and ours for the Arctic is very high (0.81), and the RMSE (0.25°C) is smaller



**Figure 6.** Time series of HadCRUT3 observations (green) and the surface temperature (in °C) averaged over the 11 simulations using data assimilation (red) for (a) Europe, (b) Asia, (c) North America, (d) the Arctic, (e) North Atlantic, and (f) the Northern Hemisphere. The RMSE and the correlation between the time series are given at the upper left corner of each panel. The gray lines represent the mean of the 11 simulations using data assimilation plus and minus one RMSE compared to the HadCRUT3 observations.

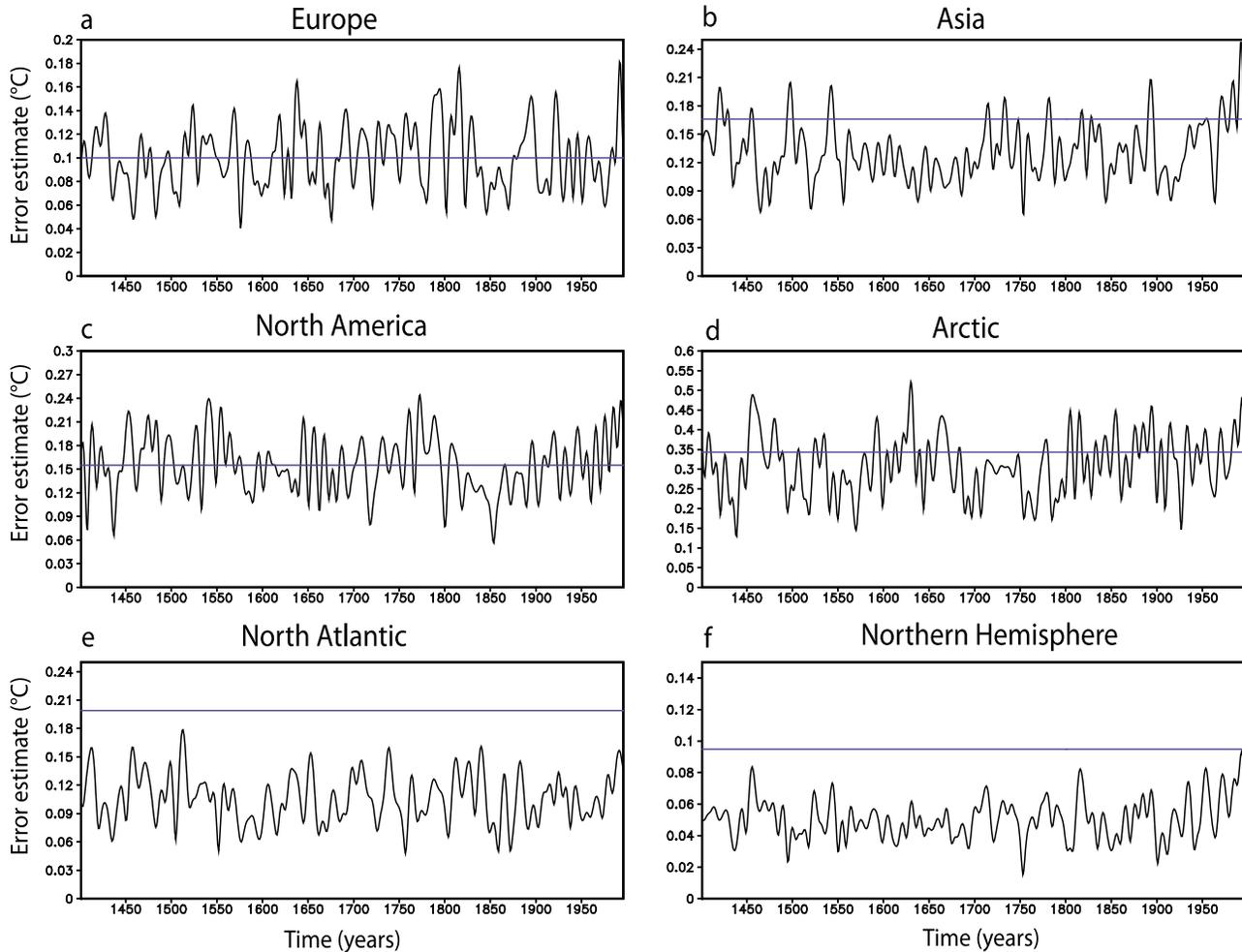
than the scatter of the various simulations with data assimilation. Our results are also in good agreement with early instrumental data collected in the Arctic [Przybylak *et al.*, 2009] that show a warming of 0.8°C between the years 1801–1920 and the years 1961–1990 (difference of 0.9°C over the same periods in the mean of the DALE reconstructions).

[56] For land area in Europe, the DALE reconstruction is very similar to that of Luterbacher *et al.* [2004] over the last 150 years. This finding is consistent with Figure 6a, as the data set of Luterbacher *et al.* [2004] is mainly based on instrumental observations over this period. For the years 1500–1850, the long-term trends are also similar in the two reconstructions with a cooling trend until 1700, a general warming trend peaking in the beginning of the 19th century, and a cooling trend during the 19th century. Even at the decadal time scale, the two reconstructions agree relatively well, with, for instance, a large cooling at the end of the 16th and 17th centuries being captured in both of them. Again, the two reconstructions cannot be considered as strictly independent, as they share some data. Despite this general agreement, the reconstruction of Luterbacher *et al.* [2004]

appears shifted by about 0.15°C over the period 1500–1850 compared to our simulations with data assimilation. The DALE reconstruction thus displays long-term variations of a higher magnitude. Because of this shift, the RMSE between the two reconstructions reaches 0.17°C and the correlation is only 0.68. It is interesting to note that Küttel *et al.* [2007] independently tested the method of Luterbacher *et al.* [2004] in a surrogate climate derived from the results of two general circulation models. They suggested that the Luterbacher *et al.* [2004] reconstruction may underestimate the magnitude of the changes before 1820 [see Küttel *et al.*, 2007, Figure 3], mainly because of the paucity of the data in the earlier period. Such an underestimation appears consistent with our results.

## 6. Discussion and Conclusions

[57] From the previous sections, it is fair to state that the answers to the three questions posed in section 1 are all yes: the data assimilation procedure performs satisfactorily with the proxy data constraints used in this study:



**Figure 7.** Time series of the standard deviation of the surface temperature (in  $^{\circ}\text{C}$ ) computed for the 11 simulations with data assimilations (black) compared to the RMSE over the period 1850–1995 (blue) for (a) Europe, (b) Asia, (c) North America, (d) Arctic, (e) North Atlantic, and (f) the Northern Hemisphere.

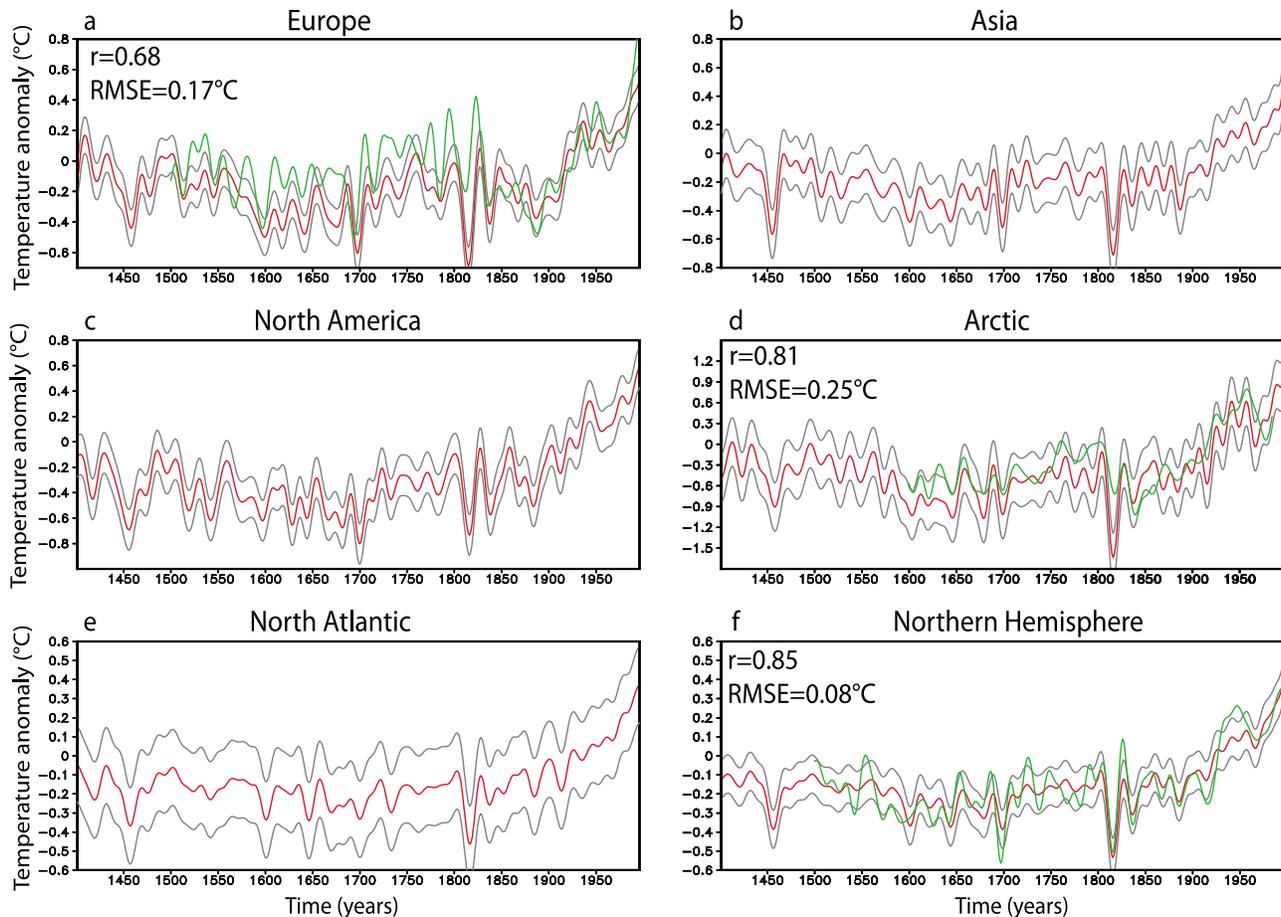
[58] 1. The simulations with data assimilation follow the signal recorded in the majority of the proxy records, as required by the methodology. The agreement between proxy data and the DALE reconstruction is best at the regional scale in areas where the data coverage is the highest, with correlations generally higher than 0.85. This is an important finding, as this regional scale is the smallest at which coarse resolution models such as LOVECLIM can be expected to provide useful information with regard to climate system dynamics.

[59] 2. The uncertainty in model results associated with internal climate variability, the selection of model parameters, and the estimates of external radiative forcing, each of which plays an important role in the differences observed between previous simulations of the past millennium performed with different climate models, is seen to decrease substantially when data assimilation is applied. At the continental scale, the additional constraint brought by the proxies reduces RMSE up to a factor of 2 in comparison with simulations without data assimilation.

[60] 3. The agreement between our DALE reconstruction and instrumental data is good at the hemispheric scale and at continental scale but to a lesser extent for the Arctic. The

correlation between the average of all the simulations with data assimilation and the HadCRUT3 surface temperature data is between 0.72 and 0.86 for Europe, Asia, North America, and the Northern Hemisphere, while it reaches 0.64 for the Arctic. Our DALE reconstructions are also very similar to previously published statistical proxy reconstructions for the Arctic, the Northern Hemisphere mean, and, with certain reservations (as discussed below), Europe. This comparison across reconstruction techniques indicates internal consistency between different approaches: the data assimilation methodology at the very least is seen not to introduce spurious information into regional-scale reconstructions, as the features in the DALE reconstructions are similar to those of the regional proxy composites themselves. At the same time, the previous statistical-based temperature reconstructions appear to be dynamically consistent, as it is possible to reproduce similar time histories using our model assimilation framework, which respects the physical constraints on the climate system.

[61] However, we must recall that satisfactory results are only demonstrated here for one set of proxy data and one model. While this finding is encouraging, there is no a priori guarantee that it would hold for other proxy sets with dif-



**Figure 8.** Time series of the temperature (in °C) averaged over the 11 simulations using data assimilation (red) for (a) Europe, (b) Asia, (c) North America, (d) the Arctic, (e) the North Atlantic, and (f) the Northern Hemisphere. The gray lines represent the mean of the 11 simulations using data assimilation plus and minus one RMSE compared to the HadCRUT3 observations. For Europe, the Arctic, and the Northern Hemisphere, reconstructions of *Luterbacher et al.* [2004], *Overpeck et al.* [1997], and *Mann et al.* [2008], respectively, are displayed in green. The RMSE and the correlation between those time series and the mean of the 11 simulations using data assimilation are given at the top left corner. In contrast with previous figures, the mean for Europe is based on land-only temperatures, and the domain has been modified for a better comparison with *Luterbacher et al.* [2004].

ferent characteristics or other models. The procedure described above, and the attempts to address the three questions we posed in section 1, should thus be repeated each time a new group of simulations with data assimilation is performed using different data in order to test the validity of the procedure. On the basis of our experiments alone, we would conclude, for example, that there is a lower bound on the number of reliable proxy records required to obtain suitable results. For the Northern Hemisphere continents, our analyses suggest that a set of roughly 50 quality proxy records is sufficient to obtain skill in reconstructed annual mean temperature at the continental scale.

[62] Despite the good agreement between simulations with data assimilation and proxy records at local and regional scale, our DALE reconstructions do not blindly follow all the proxy records. In addition, we observed a shift between our results and the *Luterbacher et al.* [2004] European temperature reconstruction prior to A.D. 1850. A number of factors could be responsible for the observed

discrepancy. These factors include possible biases in model physics, a too-small ensemble size precluding adequate sampling of the available climate state space, or misspecification of the relevant external forcing, for example, relatively local forcings such as regional land-use changes not adequately represented in the specified forcing of the model. The proxy series themselves can suffer from biases in their ability to reproduce both high- and low-frequency climate variations [see, e.g., *Jones et al.*, 2009]. Our DALE reconstructions do not draw an intrinsic distinction between calibration and validation periods, which might be an advantage compared to purely statistical methods when analyzing long-term trends or when there is a potential for non-stationarity in the system that might be difficult to capture using the relatively short available modern calibration intervals.

[63] Simulations with data assimilation may help identifying inconsistencies between the signals recorded by different proxies. Indeed, when such inconsistencies occur, the

model tends to follow the proxy time series that are the most coherent with the model large-scale dynamics, while the remaining proxies have a weaker influence on the model evolution. Identifying precisely the cause of those inconsistencies, as well as the reasons for differences between our continental-scale reconstruction and previously published statistical reconstructions, would require additional experiments using different proxy data sets to constrain the model evolution. This would allow for an analysis of the role of each proxy (or groups of proxies), controlling for the role of the model itself. Such an analysis is outside of the scope of the present paper but would be fertile ground for future investigation.

[64] In contrast to the other regions tested here, the simulations with data assimilation do not reduce the uncertainty in model results for the North Atlantic, and our reconstruction in this region is not strongly correlated with instrumental data in the region. A likely explanation for this deficiency is the absence of proxy data in the Atlantic to constrain model results there. To test this hypothesis, we have directly assimilated HadCRUT3 data in an additional set of simulations (as done by *Goosse et al.* [2009] for the Southern Ocean). In this case, the agreement between model results and observations is much higher, with RMSE close to  $0.1^{\circ}\text{C}$  and correlation higher than 0.75.

[65] However, the absence of proxy data in the North Atlantic is likely not the only source of bias in this case. Even without data constraint, the model should nonetheless be able to reasonably sample the range of variability of the system, finding a realization wherein the behavior over the North Atlantic region at least approximates the gross features of past temperature changes in the region. However, we are unable to reproduce the amplitude of the observed multidecadal variability over the last 150 years in the North Atlantic in any of our simulations with data assimilation. This is probably related to the inability of LOVECLIM without data assimilation to simulate large multidecadal variations in this region.

[66] The data assimilation technique itself also plays a role in the observed discrepancy. The initial states for all the members of the ensembles in each simulation with data assimilation differ only in their quasi-geostrophic potential vorticity field. This approach to generating alternative ensembles is quite efficient in generating rapid, random perturbations in atmospheric dynamics, but it may be inadequate for exciting multidecadal perturbations in the dynamics of the ocean [e.g., *Zanna and Tziperman*, 2008]. When adequate data are available, as in the aforementioned experiments driven by modern HadCRUT3 surface temperature data, this shortcoming can be compensated by the very strong data constraints provided by the instrumental record. In the absence of such strong data constraints, however, that is, when only relatively sparse proxy records are available, this deficiency is more pronounced. In ongoing work, we are investigating alternative approaches to ensemble generation that might more faithfully incorporate lower-frequency ocean dynamical variability, leading to improvements in reproducing multidecadal variability in regions such as the North Atlantic. This could be achieved by using methods based on the ones already applied in weather forecasting to generate initial perturbations, such as the breeding method, or by using more sophisticated tech-

niques for which generating the new ensemble after each assimilation step is an intrinsic element of the method [e.g., *Evensen*, 2003; *van Leeuwen*, 2009; *Yang et al.*, 2009].

[67] The main goal of the current study is to test the methodology in operational conditions, that is, using a realistic set of proxies, and to guide future developments of data assimilation methods adapted to the study of the climate of the past millennium. In this framework, it is interesting to note that our results with data assimilation are not particularly sensitive to the forcing applied or to the model climate sensitivity, in contrast to simulations without data assimilation. This is in a certain sense an auspicious finding since the uncertainty in our knowledge of climate sensitivity is unlikely to be substantially narrowed in the near future. That having been said, these conclusions are valid only for the relatively narrow range of parameter choices and forcing estimates considered here, which is somewhat modest in comparison with that adopted in some other studies [e.g., *Annan and Hargreaves*, 2006; *Hegerl et al.*, 2006]. Adopting a more extreme choice for (e.g.) the model climate sensitivity would almost certainly impact our simulated temperature histories, despite the tendency of the data assimilation to constrain the model results. In addition, the temperature changes are relatively modest during the last millennium and conclusions regarding the role of the forcing may be different for some periods presenting large and rapid climate changes. For instance, an adequate representation of the forcing variations is likely much more important for periods like the 8.2 ka event, during which the timing and magnitude of the freshwater discharge in the North Atlantic appears to have a very strong impact on simulated results [e.g., *Wiersma et al.*, 2006].

[68] The choice of parameters in the data assimilation method has a clear influence on our results. Varying those parameters has been considered here as a simple way to estimate the uncertainties associated with the data assimilation technique. However, the applied method is fairly elementary. As a consequence, rather than try to optimize those parameters, it appears more justified to use a more sophisticated data assimilation scheme.

[69] Two aspects of our analysis could be refined in future work, with likely perspectives for improved results. First, we have deliberately considered the proxy as “truth.” This appears to be a reasonable first approximation, as our DALE reconstructions compare favorably with instrumental estimates over the last 150 years in the majority of the investigated regions. However, this assumption clearly represents an oversimplification, and many techniques exist now in data assimilation to take into account the uncertainties in the observational data. Such techniques should and will, in future work, be adapted in the context of our paleoclimate data assimilation efforts. Second, the information from the ensemble could be employed in a more sophisticated manner. Our current assimilation method can be considered as a crude “particle filter” approach [see, e.g., *Widmann et al.*, 2009]. The “crudeness” of the approach comes from the fact that at each analysis step only one ensemble member is retained based on the comparison between model results and proxy data, and all other ensemble members are simply discarded. A more efficient use of the information within the ensemble would be to form an optimal reconstruction through an appropriately weighted combination of individual ensemble

members, and to use the spread within the ensemble to estimate uncertainties and to generate initial conditions for the next step, as is done routinely in modern atmospheric data assimilation [e.g., Kalnay, 2003]. While some of these shortcomings are compensated by our use of a set of 11 different ensembles of simulations, more efficient techniques are certainly available (see, e.g., van Leeuwen [2009] for a recent review). While such techniques are more challenging to implement, the preliminary successes demonstrated in the current study would appear to justify the investigation of more sophisticated approaches to the problem.

[70] In addition to addressing certain methodological questions, the current study also provides a new set of reconstructions of surface temperature changes at the continental scale spanning the past 600 years. Despite clear potential for improvements, the quality of those reconstructions is sufficient for meaningful comparisons to be drawn with other reconstructions or with newly developed proxy climate records. Furthermore, the model results should provide a useful platform for formulating hypotheses regarding the mechanisms responsible for the reconstructed climate changes [e.g., Crespin et al., 2009].

[71] **Acknowledgments.** The authors would like to thank the scientists that made available all their data sets. H.G. is Research Associate with the Fonds National de la Recherche Scientifique (Belgium) and is supported by the Belgian Federal Science Policy Office. H.R. is sponsored by the Netherlands Organization for Scientific Research (N.W.O). A.T. is supported by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) through its sponsorship of the International Pacific Research Center. M.E.M. acknowledges support from the ATM program of the National Science Foundation (grant ATM-0542356).

## References

- Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR climate system model, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 3713–3718, doi:10.1073/pnas.0605064103.
- Annan, J. D., and J. C. Hargreaves (2006), Using multiple observationally based constraints to estimate climate sensitivity, *Geophys. Res. Lett.*, *33*, L06704, doi:10.1029/2005GL025259.
- Bauer, E., M. Claussen, V. Brovkin, and A. Huenerbein (2003), Assessing climate forcings of the Earth system for the past millennium, *Geophys. Res. Lett.* *30*(6), 1276, doi:10.1029/2002GL016639.
- Berger, A. L. (1978), Long-term variations of daily insolation and Quaternary climatic changes, *J. Atmos. Sci.*, *35*, 2363–2367, doi:10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2.
- Bertrand, C., M.-F. Loutre, M. Crucifix, and A. Berger (2002), Climate of the last millennium: A sensitivity study, *Tellus*, *54A*, 221–244.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Brovkin, V., J. Bendtsen, M. Claussen, A. Ganopolski, C. Kubatzki, V. Petoukhov, and A. Andreev (2002), Carbon cycle, vegetation and climate dynamics in the Holocene: Experiments with the CLIMBER-2 model, *Global Biogeochem. Cycles*, *16*(4), 1139, doi:10.1029/2001GB001662.
- Charlson, R.J., J. Langner, H. Rodhe, C.B. Leovy and S.G. Warren (1991), Perturbation of the Northern Hemisphere radiative balance by backscattering from anthropogenic sulfate aerosols, *Tellus*, *43A*, 152–163.
- Collins, M. (2003), Quantitative use of paleo-proxy data in global circulation models, *Geophys. Res. Abstr.*, *5*, 0114.
- Crespin, E., H. Goosse, T. Fichefet, and M. E. Mann (2009), The 15th century Arctic warming using coupled model simulations with data assimilation, *Clim. Past*, *5*, 389–401.
- Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*, 270–277, doi:10.1126/science.289.5477.270.
- Crowley, T. J., S. K. Baum, K. Y. Kim, G. C. Hegerl, and W. T. Hyde (2003), Modeling ocean heat content changes during the last millennium, *Geophys. Res. Lett.*, *30*(18), 1932, doi:10.1029/2003GL017801.
- D'Arrigo, R., R. Wilson, and G. Jacoby (2006), On the long-term context for late twentieth century warming, *J. Geophys. Res.*, *111*, D03103, doi:10.1029/2005JD006352.
- Driesschaert, E., T. Fichefet, H. Goosse, P. Huybrechts, I. Janssens, A. Mouchet, G. Munhoven, V. Brovkin, and N. Weber (2007), Modeling the influence of Greenland ice sheet melting on the Atlantic meridional overturning circulation during the next millennia, *Geophys. Res. Lett.*, *34*, L10707, doi:10.1029/2007GL029516.
- Esper, J., E. R. Cook, and F. H. Schweingruber (2002), Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, *Science*, *295*, 2250–2253, doi:10.1126/science.1066208.
- Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.
- Foukal, P., C. Frölich, H. Spruit, and T. M. L. Wigley (2006), Variations in solar luminosity and their effect on the Earth's climate, *Nature*, *443*, 161–166, doi:10.1038/nature05072.
- Gonzalez-Rouco, J. F., H. Beltrami, E. Zorita, and H. von Storch (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., and T. Fichefet (1999), Importance of ice-ocean interactions for the global ocean circulation: A model study, *J. Geophys. Res.*, *104*, 23,337–23,355, doi:10.1029/1999JC900215.
- Goosse, H., H. Renssen, A. Timmermann, and R. S. Bradley (2005a), Internal and forced climate variability during the last millennium: A model-data comparison using ensemble simulations, *Quat. Sci. Rev.*, *24*, 1345–1360, doi:10.1016/j.quascirev.2004.12.009.
- Goosse, H., T. Crowley, E. Zorita, C. Ammann, H. Renssen, and E. Driesschaert (2005b), Modelling the climate of the last millennium: What causes the differences between simulations?, *Geophys. Res. Lett.*, *32*, L06710, doi:10.1029/2005GL022368.
- Goosse, H., H. Renssen, A. Timmermann, R. S. Bradley, and M. E. Mann (2006a), Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium, *Clim. Dyn.*, *27*, 165–184, doi:10.1007/s00382-006-0128-6.
- Goosse, H., O. Arzel, J. Luterbacher, M. E. Mann, H. Renssen, N. Riedwyl, A. Timmermann, E. Xoplaki, and H. Wanner (2006b), The origin of the European “Medieval Warm Period,” *Clim. Past*, *2*, 99–113.
- Goosse, H., E. Driesschaert, T. Fichefet, and M.-F. Loutre (2007), Information on the early Holocene climate constrains the summer sea ice projections for the 21st century, *Clim. Past*, *3*, 683–692.
- Goosse, H., M. E. Mann, and H. Renssen (2008), What we can learn from combining paleoclimate proxy data and climate model simulations of past centuries? In *Natural Climate Variability and Global Warming: A Holocene Perspective*, edited by R. W. Battarbee and H. A. Binney, pp. 163–188, Blackwell, Oxford, U. K.
- Goosse, H., W. Lefebvre, A. de Montety, E. Crespin, and A. Orsi (2009), Consistent past half-century trends in the atmosphere, the sea ice and the ocean at high southern latitudes, *Clim. Dyn.*, *33*, 999–1016, doi:10.1007/s00382-008-0500-9.
- Guiot, J., A. Nicault, C. Rathgeber, J. L. Edouard, F. Guibal, G. Pichard, and C. Tili (2005), Last-millennium summer-temperature variations in Western Europe based on proxy-data, *Holocene*, *15*, 489–500, doi:10.1191/0959683605hl819p.
- Haigh, J. D. (1996), The impact of solar variability on climate, *Science*, *272*, 981–984, doi:10.1126/science.272.5264.981.
- Hegerl, G. C., T. C. Crowley, W. T. Hyde, and D. J. Frame (2006), Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, *440*, 1029–1032, doi:10.1038/nature04679.
- Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita (2007), Detection of human influence on a new, validated 1500-year temperature reconstruction, *J. Clim.*, *20*, 650–666, doi:10.1175/JCLI4011.1.
- Jansen, E., et al. (2007), Palaeoclimate, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 433–497, Cambridge Univ. Press, New York.
- Jones, P. D., et al. (2009), High-resolution paleoclimatology of the last millennium: A review of the current status and future prospects, *Holocene*, *19*, 3–49, doi:10.1177/0959683608098952.
- Juckes, M. N., M. R. Allen, K. R. Briffa, J. Esper, G. C. Hegerl, A. Moberg, T. J. Osborn, and S. L. Weber (2007), Millennial temperature reconstruction intercomparison and evaluation, *Clim. Past*, *3*, 591–609.
- Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, 339 pp., Cambridge Univ. Press, Cambridge, U. K.

- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Küttel, M., J. Luterbacher, E. Zorita, E. Xoplaki, N. Riedwyl, and H. Wanner (2007), Testing a European winter surface temperature reconstruction in a surrogate climate, *Geophys. Res. Lett.*, **34**, L07710, doi:10.1029/2006GL027907.
- Lean, J. L., Y.-M. Wang, and N. R. Sheeley (2002), The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate, *Geophys. Res. Lett.*, **29**(24), 2224, doi:10.1029/2002GL015880.
- Lee, T. C. K., F. Zwiers, and M. Tsao (2008), Evaluation of proxy-based millennial reconstruction methods, *Clim. Dyn.*, **32**, 263–281, doi:10.1007/s00382-007-0351-9.
- Lorenz, E. N. (1969), Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, **26**, 636–646, doi:10.1175/1520-0469(1969)26<636:APARB>2.0.CO;2.
- Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004), European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, **303**, 1499–1503, doi:10.1126/science.1093877.
- Mann, M. E., and P. D. Jones (2003), Global surface temperatures over the past two millennia, *Geophys. Res. Lett.*, **30**(15), 1820, doi:10.1029/2003GL017814.
- Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, **26**, 759–762, doi:10.1029/1999GL900070.
- Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann (2005), Testing the fidelity of methods used in proxy-based reconstructions of past climate, *J. Clim.*, **18**, 4097–4410, doi:10.1175/JCLI3564.1.
- Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann (2007), Robustness of proxy-based climate field reconstruction methods, *J. Geophys. Res.*, **112**, D12109, doi:10.1029/2006JD008272.
- Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. Miller, S. Rutherford, and F. Ni (2008), Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 13,252–13,257, doi:10.1073/pnas.0805721105.
- Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlén (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, **433**, 613–617, doi:10.1038/nature03265.
- Muscheler, R., F. Joos, J. Beer, S. A. Müller, M. Vonmoos, and I. Snowball (2007), Solar activity during the last 1000 yr inferred from radionuclide records, *Quat. Sci. Rev.*, **26**, 82–97, doi:10.1016/j.quascirev.2006.07.012.
- Nicolis, C. (1998), Atmospheric analogs and recurrence time statistics: Toward a dynamical formulation, *J. Atmos. Sci.*, **55**, 465–475, doi:10.1175/1520-0469(1998)055<0465:AAARTS>2.0.CO;2.
- Opsteegh, J. D., R. J. Haarsma, F. M. Selten, and A. Kattenberg (1998), ECBILT: A dynamic alternative to mixed boundary conditions in ocean models, *Tellus*, **50A**, 348–367.
- Osborn, T. J., S. C. B. Raper, and K. R. Briffa (2006), Simulated climate change during the last 1,000 years: Comparing the ECHO-G general circulation model with the MAGICC simple climate model, *Clim. Dyn.*, **27**, 185–197, doi:10.1007/s00382-006-0129-5.
- Overpeck, J., et al. (1997), Arctic environmental change of the last four centuries, *Science*, **278**, 1251–1256.
- Przybylak, R., Z. Vizi, and P. Wyszynski (2009), Air temperature changes in the Arctic from 1801 to 1920, *Int. J. Climatol.*, doi:10.1002/joc1918.
- Raible, C. C., et al. (2006), Climate variability: Observations, reconstructions, and model simulations for the Atlantic-European and alpine region from 1500–2100 AD, *Clim. Change*, **79**, 9–29, doi:10.1007/s10584-006-9061-2.
- Ramankutty, N., and J. A. Foley (1999), Estimating historical changes in global land cover: Croplands from 1700 to 1992, *Global Biogeochem. Cycles*, **13**(4), 997–1027, doi:10.1029/1999GB900046.
- Robock, A. (2000), Volcanic eruptions and climate, *Rev. Geophys.*, **38**, 191–219, doi:10.1029/1998RG000054.
- Rutherford, S., M. E. Mann, T. J. Osborn, R. S. Bradley, K. R. Briffa, M. K. Hughes, and P. D. Jones (2005), Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain, *J. Clim.*, **18**, 2308–2329, doi:10.1175/JCLI3351.1.
- Shindell, D. T., G. A. Schmidt, M. E. Mann, D. Rind, and A. Waple (2001), Solar forcing of regional climate change during the Maunder minimum, *Science*, **294**, 2149–2152, doi:10.1126/science.1064363.
- Shindell, D. T., G. A. Schmidt, M. E. Mann, and G. Faluvegi (2004), Dynamic winter climate response to large tropical volcanic eruptions since 1600, *J. Geophys. Res.*, **109**, D05104, doi:10.1029/2003JD004151.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, **135**, 4629–4640, doi:10.1175/2008MWR2529.1.
- Stendel, M., I. A. Mogensén, and J. H. Christensen (2006), Influence of various forcings on global climate in historical times using a coupled atmosphere-ocean general circulation model, *Clim. Dyn.*, **26**, 1–15, doi:10.1007/s00382-005-0041-4.
- Tett, S. F. B., R. Betts, T. J. Crowley, J. Gregory, T. C. Johns, A. Jones, T. J. Osborn, E. Ostrom, D. L. Roberts, and M. J. Woodage (2007), The impact of natural and anthropogenic forcings on climate and hydrology since 1550, *Clim. Dyn.*, **28**(1), 3–34, doi:10.1007/s00382-006-0165-1.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.
- Van den Dool, H. M. (1994), Searching for analogues, how long must we wait? *Tellus*, **46A**, 314–324.
- van der Schrier, G., and J. Barkmeijer (2005), Bjerknes' hypothesis on the coldness during AD 1790–1820 revisited, *Clim. Dyn.*, **24**, 335–371, doi:10.1007/s00382-004-0506-x.
- van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Mon. Weather Rev.*, **137**, 4089–4114.
- Waple, A. M., M. E. Mann, and R. S. Bradley (2002), Long-term patterns of solar irradiance forcing in model experiments and proxy based surface temperature reconstructions, *Clim. Dyn.*, **18**, 563–578.
- Widmann, M., H. Goosse, G. van der Schrier, R. Schnur, and J. Barkmeijer (2009), Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, *Clim. Past Discuss.*, **5**, 2115–2156.
- Wiersma, A. P., H. Renssen, H. Goosse, and T. Fichefet (2006), Evaluation of different freshwater forcing scenarios for the 8.2 ka BP event in a coupled climate model, *Clim. Dyn.*, **27**, 831–849, doi:10.1007/s00382-006-0166-0.
- Yang, S. C., C. Keppenne, M. Rienecker, and E. Kalnay (2009), Application of the coupled bred vectors to seasonal-to-interannual forecasting and ocean data assimilation, *J. Clim.*, **22**, 2850–2870, doi:10.1175/2008JCLI2427.1.
- Zanna, L., and E. Tziperman (2008), Optimal surface excitation of the thermohaline circulation, *J. Phys. Oceanogr.*, **38**(8), 1820–1830, doi:10.1175/2008JPO3752.1.

E. Cressin, A. de Montety, and H. Goosse, Université Catholique de Louvain, Institut d'Astronomie et de Géophysique G. Lemaître, 1348 Louvain-la-Neuve, Belgium. (hugues.goosse@uclouvain.be)

M. E. Mann, Department of Meteorology and Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA.

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

A. Timmermann, International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI 96822, USA.