Manuscript prepared for J. Name
with version 2014/05/30 6.91 Copernicus papers of the LaTeX class copernicus.cls.
Date: 3 November 2014

On-line and off-line data assimilation in palaeoclimatology: a case study

Anastasios Matsikaris[1,2], **Martin Widmann**[1], **and Johann Jungclaus**[2]

[1]University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[2]Max-Planck-Institute for Meteorology, Hamburg, Germany

Correspondence to: ANASTASIOS MATSIKARIS (axm368@bham.ac.uk)

Abstract. Different ensemble-based data assimilation (DA) approaches for palaeoclimate reconstructions have been recently followed, but no systematic comparison among them has been attempted. We compare an off-line and an on-line ensemble-based method, with the testing period being the 17th century, which led into the Maunder Minimum. We use a low-resolution version of Max Planck Institute for Meteorology's model MPI-ESM, to assimilate the PAGES 2K continental temperature reconstructions. In the off-line approach the ensemble for the entire simulation period is generated first and then the ensemble is used in combination with the empirical information to produce the analysis. In contrast, in the on-line approach the ensembles are generated sequentially for sub-periods based on the analysis of previous sub-periods. Both schemes perform better than the simulations without DA. The on-line method would be expected to perform better if the assimilation led to states of the slow components of the climate system that are close to reality and the system had sufficient memory to propagate this information forward in time. In our comparison, which is based on analysing correlations and differences between the analysis and the proxy-based reconstructions, we find similar skill for both methods on the continental and hemispheric scales. This indicates either a lack of control of the slow components in our setup or a lack of skill in the information propagation on decadal timescales. Although the performance of the two schemes is similar and the on-line method is more difficult to implement, the temporal consistency of the analysis in the on-line method makes it in general preferable.

1 Introduction

Reconstructing the climate of the past is crucial for quantifying and understanding natural climatic change, which in turn is essential for detecting anthropogenic climate change, as well as for the validation of climate models that are used to provide future climate projections. As the instrumental meteorological records are too short to estimate low-frequency variability, reconstructions based on climate proxy data or numerical simulations are used for this purpose. However, both approaches are associated with substantial uncertainties. In principle, the best state estimates can be expected by employing data assimilation (DA) techniques, which systematically combine the empirical information from proxy data with the representation of the processes that govern the climate system given by climate models. Although DA is a very mature field in numerical weather prediction, the specific problem in palaeoclimatology is different and the methods cannot be directly transferred (e.g. Widmann et al., 2010; Hakim et al., 2013). DA is an emerging research area and can be considered as one of the key challenges in palaeoclimatology.

There are two types of proxy-based reconstructions, those for large-scale, e.g. continental or hemispheric averages (e.g. Crowley and Lowery, 2000; Moberg et al., 2005; Mann et al., 2008; Ljungqvist, 2010; PAGES 2K Consortium, 2013) and spatial field reconstructions (e.g. Briffa et al., 1994; Luterbacher et al., 2004; Jones and Mann, 2004; Xoplaki et al., 2005; Mann et al., 2009). Proxy-based estimates of climate variability contain considerable errors: different proxies usually represent different seasons, different statistical methods used in the reconstructions lead to different results, and non-climatic factors influence the proxies (e.g. Jansen et al., 2007; Jones and Mann, 2004). Moreover, the poor spatial coverage of the climate proxies leads to errors in hemispheric or continental means and even larger errors in full-field reconstructions. The climate states provided by standard model simulations are spatially complete and provide an independent estimate which can be checked for consistency with the proxies, on both large and regional scales. However, the simulations also have errors, e.g. systematic model biases and errors in the climate forcings or in the response to them. Additionally, interannual to decadal temperature variations have a large random, non-forced component and thus agreement of simulations and observations is very unlikely on these timescales. The forcings do not precisely determine the temporal evolution of the climate, in particular on regional scales. Ensemble simulations are indispensable in order to better assess the internal variability for periods within the last millennium (Jungclaus et al., 2010).

Data assimilation combines the two previous methods to find estimates that are both consistent with the empirical knowledge and with the dynamical understanding of the climate system, providing complete spatial fields. It uses the empirical data after the construction of the model to either estimate, correct or select the system state (e.g. Hakim et al., 2013; Bronnimann et al., 2013), or to systematically improve some model parameters (e.g. Annan et al., 2013). Here, we consider the case of state estimation, where DA aims to capture the real-world random, non-forced variability in a simulation and to provide information for variables for which no empirical estimates exist.

Attempts to assimilate proxy data into models include different approaches, such as the selection of ensemble members, forcing singular vectors, and pattern nudging (e.g. Widmann et al., 2010). Ensemble member selection techniques, like the one implemented here, are based on the selection of simulations from an ensemble that are closest to the empirical evidence on climate. A general advantage of these techniques is that they are easy and straightforward to implement, and they are the most frequently used methods by the community. Goosse et al. (2006) were the first to use this method for palaeoclimate research, employing a simplified global 3-D climate model. An updated version was employed by Goosse et al. (2010), using a more advanced 3-D Earth-System Model of Intermediate Complexity(EMIC), along with a set of 56 proxy series derived from a comprehensive compilation of Mann et al. (2008). In the first case, the best model analog was selected by comparing the simulations with proxy-based temperature reconstructions after the completion of the simulations, an approach called off-line DA. In the second case a new ensemble was generated at each step of the assimilation procedure, starting from the best simulation selected for the previous period, an approach called on-line DA. The revised method offered dynamical consistency between best model analogs of different periods, while the former benefited from its computational simplicity. Both methods showed positive reconstruction skill, particularly at the regional scale in areas with high data coverage. The on-line method was also employed by Crespin et al. (2009) to analyse the fifteenth century Arctic warming. The novelty of the current manuscript is the focus on the comparison of the on-line and off-line approaches.

In addition to the above methods, where a single simulation having the best fit to the data is chosen during the assimilation (“degenerate particle filter”), another approach employs weights for each member of the ensemble, calculated after the comparison with the proxies and generating a probabilistic posterior distribution (“particle filter”). The technique was applied by Annan and Hargreaves (2012), who performed off-line assimilation based on a simple likelihood weighting algorithm, implementing thus all the DA after the completion of the ensemble integration. In the “particle filter” methods (both in the on-line and off-line techniques), more than one member proceeds to the next assimilation step after the first filtering. The most unlikely ensemble members (particles) are being discarded and the highly likely particles are being copied proportionally to their likelihood. The same “probabilistic posterior distributions” technique was used by Goosse et al. (2012). The outcomes of the approach led to distributions with larger overlaps with the proxy-based reconstruction. The method has also been used by Mairesse et al. (2013) to reconstruct the climate of the mid-Holocene (6 kyr BP).

Other ensemble-based DA approaches include the use of the Kalman filter and the explicit treatment of time-averaged observations. The off-line approach of DA was advanced by Bhend et al. (2012), through the assimilation of proxy data into a high-resolution general circulation model (GCM). The ensemble square root filter (EnSRF), a variant of the ensemble Kalman filter, was used to update the ensembles with climate proxy information. The use of an atmosphere-only GCM rather than a coupled atmosphere-ocean GCM left no possibility for information propagation on long timescales, therefore the DA was performed off-line. In other words, an on-line DA scheme would not have benefited the reconstruction skill, apart from leading to temporal consistency of the analysis. Dirren and Hakim (2005) examined the case where only time-averaged observations are available. Their algorithm constitutes a natural extension of the ensemble Kalman filter, and reduces to the ensemble Kalman filter in the limit of zero time averaging (Dirren and Hakim, 2005). Huntley and Hakim (2010) applied the new algorithm to test the method in a simple atmospheric model. Similarly, Pendergrass et al. (2012) tested two idealized models, which captured adequate climate variability related to the palaeoproxies. In order to identify initial conditions, an ensemble Kalman filter technique was applied to the two models. Another computationally inexpensive DA method, adapted for past climates, was presented by Steiger et al. (2014), requiring only a static ensemble of climatologically plausible states.

An advantage of the on-line compared to the off-line ensemble-based DA methods is the temporal consistency of the simulated states. The off-line approach on the other hand is computationally less complicated and can also be computationally cheaper if one uses simulations that already exist. The question we address in this paper is whether the on-line reconstruction is closer to the proxy-based reconstructions compared to the off-line version. This depends on the memory of the slow components of the climate system, such as the ocean. If these propagate the information contained in the assimilated proxy data forward in time on decadal timescales, and this information is correct, the on-line approach is expected to perform better. If, on the other hand, the chaotic nature of the system dominates and the predictability of the system is limited, or the simulated ocean states are unrealistic, the computationally easier off-line method would be sufficient. The experiment design with decadal assimilation is motivated by a number of reasons. Firstly, since we aimed for a complete Northern Hemisphere reconstruction, the 10-year resolution of the North American proxy reconstructions did not allow us to use annually resolved proxy data for the assimilation. Additionally, the annually resolved proxies include substantial noise, which is cancelled out with the decadal averaging. Finally, in a climate change context, the yearly changes are in general of less interest compared to the decadal variability. GCMs exhibit up to decadal predictability in the North Atlantic (e.g. Branstator et al., 2012; Hawkins and Sutton, 2009a) and the ocean predictability can in turn lead to atmospheric predictability. The extent of decadal predictability and the relevant mechanism behind are not yet clear and many studies have recently been performed on these topics (e.g. Hawkins and Sutton, 2009a, b; Keenlyside and Ba, 2010).

In this paper, we compare two ensemble-based DA approaches, an off-line and an on-line method, to reconstruct the climate for the period 1600-1700 AD. This is a period for which many proxy studies and model simulations exist, and which is interesting due to the large temperature variations exhibited in the transition to the prolonged cold period of the Maunder Minimum (about 1645 AD to 1715 AD). We employ ensemble simulations with the Max Planck Institute for Meteorology's General Circulation Model MPI-ESM, and specifically a low-resolution version of the MPI CMIP5 model. The proxy temperature reconstructions of the PAGES 2K project are used in our assimilation (PAGES 2K Consortium, 2013). The structure of the paper is as follows: In section 2, we review the model characteristics and the proxy datasets used, and give the details of our methodology. Section 3 gives the results of the validation of the off-line and the on-line DA approaches and a comparison of them, discusses their limitations and includes a significance test of the results. Finally, in section 4, we summarize, draw conclusions and discuss the benefits of each approach.

2 Experimental Design

2.1 Model Simulations

We used the Max Planck Institute for Meteorology Earth System Model (MPI-ESM), comprising of the general circulation models ECHAM6 (Stevens et al., 2013) for the atmosphere and MPIOM (Marsland et al., 2003) for the ocean. ECHAM6 was run at T31 horizontal resolution (3.75° × 3.75°), with 31 vertical levels, resolving the atmosphere up to 10 hPa. MPIOM was run at a horizontal resolution of 3.0° (GR30) and 40 vertical levels. The OASIS3 coupler was used to couple the ocean and the atmosphere daily without flux corrections. The land surface model was JSBACH (Raddatz et al., 2007) and no ocean biogeochemistry model was employed. The model is a low-resolution version of the model used for the Coupled Model Intercomparison Project Phase 5 (CMIP5) simulations.

The simulations described here are based on a simulation covering the last millennium (850-1849 AD) following the "past1000" protocol of the Paleo Model Intercomparison Project Phase 3 (Schmidt et al., 2011). Prescribed external forcing factors are reconstructed variations of total solar irradiance (Vieira et al., 2011), volcanic aerosols (Crowley and Unterman, 2012), concentrations of the most important greenhouse gases (Schmidt et al., 2011), and anthropogenic land-cover changes (Pongratz et al., 2008). The past1000 simulation has been started after a 700-year long spin-up with constant 850 AD boundary conditions.

The high computational cost restricted us to running 10 ensemble members for each experiment. This choice is consistent with Bhend et al. (2012) who found that ensembles of size 10 or more can be successful in finding a simulation moderately close to the proxies, and that considerable skill in regions close to the assimilated data can be found for ensembles of 15 members or more, while larger sizes are needed for areas further away. The ensemble members have been generated by slightly varying values of an atmospheric diffusion parameter. The method leads to a fast divergence of the different simulations and an adequate ensemble spread, not only in surface variables like the 2m or sea-surface temperature, but also in deeper ocean variables, such as the AMOC - Atlantic meridional overturning circulation. Figure 1 shows the AMOC time-series of the ensemble spread at 26.5°, for the first 100 days after the initialisation of the ensemble in year 1600 AD, illustrating the fast growth of the ensemble spread in ocean variables. The selected ensemble generation method does not directly introduce any disturbance in the ocean, which may limit the capability of the assimilation scheme. For this reason, a different way of generating ensembles was also tested, namely the lagged-ocean initialization method, generating the ensemble members by using different ocean initial conditions, based on different dates close to the original starting date of the generation. The similarity in the output of the two methods however, and the fact that the lagged-ocean initialization is more complicated, led us to choose the atmosphere-only disturbance.

2.2 Proxy Datasets

For our assimilation procedure, we used the "2k Network" of the IGBP Past Global Changes (PAGES) proxy datasets. The PAGES project used a global set of proxy records and produced temperature reconstructions for seven continental-scale regions (PAGES 2K Consortium, 2013). The dataset covers different periods during the last millennium for each continent, and specifically the years 167-2005 AD for Antarctica, 1-2000 AD for the Arctic, 800-1989 AD for Asia, 1001-2001 AD for Australasia, 1-2003 AD for Europe, 480-1974 AD for North America and 857-1995 AD for South America. It has been produced by nine regional working groups, who identified the best proxy climate records for the temperature reconstruction within their region, using criteria they had established a priori.

Here, we assimilate the reconstructions for the period 1600 AD to 1700 AD, which led into the Maunder Minimum. The Maunder minimum (1645 AD to 1715 AD) was characterized by a large reduction in the number of sunspots and hence a reduction in solar radiation, and corresponds to the middle part of the Little Ice Age. Volcanic forcing likely had a role in this cooling as well. The PAGES 2K reconstructions exhibit a cooling in all the continents except Antarctica for this period, being in agreement with previous studies.

The techniques followed by the majority of the groups were either the "composite plus scale" (CPS) approach for the adjustment of the mean and variance of a predictor composite to an instrumental target (e.g. Mann et al., 2008, 2009), or regression-based techniques for the predictors, including principal component pre-filters or distance weighting (PAGES 2K Consortium, 2013). The dataset of individual proxies consists of 511 time series that include ice cores, tree rings, pollen, speleothems, corals, lake and marine sediments as well as historical documents of changes in biological or physical processes. The reconstructions have annual resolution, apart from North America, which is resolved in ten- and thirty-year periods.

2.3 Selection of the best ensemble members

We simulated the period 1600-1700 A.D using the standard forcings for this period. The initial conditions were taken as the last day of the year 1599 AD from a transient forced simulation starting in 850 AD. We performed ensemble experiments of 100-year duration. In the off-line experiment, in the first year (1600 AD), the ten ensemble members used slightly different values of an atmospheric diffusion parameter. For each member, the simulation period was divided into 10-year intervals, and the decadal means of the 2m temperature were calculated for each of the Northern Hemisphere continents. Using a root mean square (RMS) error-based cost function, the model outputs were compared to the proxy-based continental temperature reconstructions, averaged over the respective 10-year periods. The ensemble member that minimized the cost function in each decade was selected as the best simulation for that period. The same process was followed for all the decades within the analysis period, so that in the end we obtained the analysis, by merging the best members of each decade.

The selection of the "optimal" simulation of the ensemble for each decade of the simulation period was done after the calculation of the following cost function:

$$CF(t) = \sqrt{\sum_{i=1}^{k} \left(T^i_{mod}(t) - T^i_{prx}(t)\right)^2} \tag{1}$$

where i are the Northern Hemisphere continents, namely the Arctic, Asia, Europe and North America, $T^i_{mod}(t)$ is the standardized modelled decadal mean of the temperatures in each Northern Hemisphere continent and $T^i_{prx}(t)$ is the standardized proxy-based reconstruction for the decadal mean of the temperatures in each Northern Hemisphere continent. The algorithm filters out the ensemble members that are considered poor representations of the actual state, by throwing away the ones that are less consistent with the proxies and promoting the best fitting member. We include only the data of the Northern Hemisphere in the cost function, in an effort to reduce the degrees of freedom of the system and make it easier to find good analogues with our small ensemble size. Moreover, the Southern Hemisphere is affected by bigger uncertainties and is reconstructed by less dense proxy networks.

The reason for basing the cost function on standardized simulated and proxy-based temperatures is to remove systematic biases in means and variances between the model and the proxy-based reconstructions, and to ensure that continental temperatures with differing variance contribute equally to the analysis. The standardized model and proxy time-series were calculated by subtracting the 850-1850 AD means of the model output and the proxies from the 1600-1700 A.D raw model output and proxies respectively, and dividing by the respective standard deviations, based on the decadal averages for the 850-1850 AD period. The datasets were not weighted according to the size of the different regions, as we consider all continents to be equally important. We also decided against weighting on the base of the errors of the proxy datasets, as the different methods followed by each of the PAGES 2K groups make the errors not directly comparable. Moreover, the errors of the continental reconstructions are of similar order and thus error weighting would only have a small effect.

In the on-line experiment, a ten-member ensemble was generated for the first year of the analysis period, by introducing small perturbations in the atmospheric diffusion field. Simulations with 10-year duration were run. Using the same cost function as the one used in the off-line experiment, the temperature decadal means of the model outputs were compared to the PAGES 2K continental proxy reconstructions. In contrast to the off-line method, the selected member for that period, i.e. the one that minimized the cost function, was used as the initial condition for the subsequent simulation. A new ensemble consisting of 10 members was performed for the second decade, starting from the previous best member's final conditions and having slightly varying values of the atmospheric diffusivity parameter in the different members. The same procedure was repeated until the year 1700 AD.

The comparison of the two experiments is based on the proximity to the proxy-based reconstructions. We note however that it is not the aim of DA to exactly reproduce the assimilated empirical information, since these have errors. Ideally, a validation of different DA methods would be based on a comparison with the true and spatially complete temperature field, but as this is not available, a validation based on proximity to the assimilated information is a useful first step to investigate whether the on-line and off-line approaches perform differently.

Having a good chance to find a close analogue of an atmospheric state requires a large number of ensemble members, if the state space has a high dimension. Van Den Dool (1994) showed that to find an accurate analogue for daily data over a large area, such as the Northern Hemisphere, one needs daily data from a period of about 10^{30} years. According to Van Den Dool (1994), using a shorter library, like the current libraries of only 10-100 years of data, analogues can be found only in just 2 or 3 degrees of freedom (e.g. Bretherton et al., 1999). In our case, by using only the continental averages of the Northern Hemisphere as targets for the assimilation process, we have a low number of degrees of freedom for our cost function (less than 3). This makes the detection of a good analogue much more likely with our small ensemble size of 10 members.

3 Results

The performance of the two schemes was assessed by computing the correlation and the root-mean-square (RMS) error for each Northern Hemisphere (NH) continent between the simulated and the proxy-based reconstructions of the 2m air temperatures. We also investigated whether there exists information propagation on decadal timescales in the model, by comparing the standard deviation of the ensembles during the sub-periods in the on-line and off-line cases. An additional significance test to evaluate the role of the sampling effects that may affect many of the aspects discussed in the study was also conducted.

3.1 Comparison of the two DA schemes

Despite the fact that the cost function for the selection of the best members was based on standardized data, we demonstrate the performance of the two schemes using the non-standardized, but unbiased model output (absolute anomalies). This is because the latter represents the actual assimilated temperatures that come out of the model, which can be compared with other studies. Starting with the off-line DA scheme, the validation shows a clear improvement of the simulated reconstruction for the period under consideration, presenting higher correlations between model and proxies for all the continents of the Northern Hemisphere and lower root mean square errors for the analysis compared to the individual members. The on-line DA scheme was also successful, improving the skill of the analysis time-series compared to the individual members. However, the scheme presented very similar correlations between the DA analysis and the proxy-based reconstructions with the ones found with the off-line approach, and no major improvements to the RMS errors, both on the continental and hemispheric scales.

Figure 2 shows the Northern Hemisphere continents' decadal mean temperature anomalies w.r.t. the 850-1850 AD mean for the 17th century, for the on-line and off-line ensemble members, the on-line and off-line DA analysis and the proxy-based reconstructions. The figure displays the ensemble spreads as shadings, but a more detailed investigation shows that the DA analysis for all the NH continents is closer to the proxies than any of the individual ensemble members, in both schemes. This result is not trivial, as the cost function only minimizes the RMS error with respect to all NH continents. Even better agreement is exhibited by the direct average of the four Northern Hemisphere continents and the Northern Hemisphere mean for both DA schemes, as illustrated in Figure 3. The direct average of the four NH continental temperatures in the simulations makes use of the same sea-land masks and seasonal representativity as the ones employed by the proxy reconstructions. Hence, it is directly comparable to the proxy datasets, which are only available as continental means. The NH mean on the other hand is the true spatial average temperature of the whole Northern Hemisphere. We show this time-series as it is the usual mean temperature given in most climate studies, despite the fact that in our comparison it not the direct equivalent of the proxy-based reconstructions (the proxy time-series in the two cases are the same).

The correlations in the off-line experiment between the analysis and the proxies are relatively high for all the NH continents (0.56 for the Arctic, 0.78 for Asia, 0.79 for Europe and 0.89 for North America). Since the cost function includes all the NH continents, the correlation is maximum for the Northern Hemisphere direct average (0.94), while the correlation for the Northern Hemisphere mean is also high (0.92). These values are much higher than the correlations of the individual members with the proxies, and also higher than the correlation of the ensemble mean with the proxies (0.73 for the NH direct average). The ensemble mean has a higher ratio of forced to random variability and thus a higher correlation with the proxy-based reconstructions than the individual members, but because of the fact that the random components of the individual members partly cancel each other, the total variance of the ensemble mean is much lower than the individual members. Similarly, the validation of the absolute anomalies in the on-line experiment reveal high correlations between analysis and proxies for all the NH continents (0.79 for the Arctic, 0.76 for Asia, 0.79 for Europe and 0.81 for North America). The correlation is again the maximum for the Northern Hemisphere direct average (0.93), and the Northern Hemisphere mean (0.92). The above values are again higher than the correlations of any individual member with the proxies, as well as higher than the correlation of the ensemble mean with the proxies (0.67).

The RMS error of the simulated time-series for each continent provides a quantification of the local agreement between the model and the proxy-based reconstructions. It is calculated based on the decadal mean differences of the model and the proxy time-series for each continent. Figure 4 shows the RMS errors for the individual members, the ensemble mean and the analysis of the four Northern Hemisphere continents in the two DA schemes. In both experiments, the RMS errors are either minimal or among the lowest for the analysis compared to all other members. The result is even more evident when considering the RMS errors for the direct average and the mean of the Northern Hemisphere (Figure 5). The fact that the RMS error of the ensemble mean is lower than the error of most of the individual members in the two experiments, might either indicate the influence of forcings, or can be simply due to the lower variance of the ensemble mean compared to the individual members, which might bring it closer to the proxies. However, a better estimate can be obtained from the DA analysis, which indicates that some of the internal variability has been successfully captured by the assimilation schemes. The RMS errors between the analysis and the proxies in the on-line DA scheme are 0.18 for the Arctic, 0.21 for Asia, 0.16 for Europe and 0.18 for North America. The RMS error for the direct average of the four Northern Hemisphere continents is 0.12, insignificantly different to the off-line one (0.11).

The assessment of the performance of the two DA schemes using the standardized data produced very similar correlations and RMS errors to the ones found when using the absolute anomalies as presented above. For the Southern Hemisphere, it is more meaningful to assess the performance of the method using the standardized data, as the RMS error only has a meaning with this approach. Not using the standardized outputs in this case would result in non-comparable scales because of the different standard deviations between model and proxies. In contrast to the good skill of the two schemes in the Northern Hemisphere, the agreement between the analysis for the Southern Hemisphere (SH) and the proxy-based reconstructions is not good, as expected from the fact that SH data are not included in the cost function.

The construction of our cost function on the basis of decadal mean temperatures of the NH, means that the analysis is not expected to be more skilful than the individual members when considering the hundred-year average. The average 17th century temperature differences between model and reconstructions, for the on-line and off-line ensemble members, the on-line and off-line ensemble means and the two analyses are presented in Figure 6, and indeed do not exhibit the best agreement between the analysis and the proxy-based reconstructions in all the regions, although this is the case in some continents.

3.2 Random sampling effects

Sampling effects may affect many of the aspects discussed in the study due to the limited ensemble size and the relatively short time period analysed. Therefore, sampling uncertainty should be more thoroughly addressed where possible. We applied a resampling method to illustrate the distribution of the skill metrics (correlation and RMS error) when randomly sampling a best model in the off-line method.

Initially, we calculated the correlations between model and proxy-based reconstructions for the NH direct average for 100 random analyses in the off-line experiment, after randomly selecting one member as the best for each of the 10 decades. The mean correlation of the randomly sampled distribution with the proxies was 0.48 (with a standard deviation of 0.21), ranging between negative values and 0.8. These correlations are very low compared to the value of 0.94 from the off-line DA analysis. For the NH mean, the mean correlation of the randomly sampled analyses was 0.63 (with a standard deviation of 0.15). It is noteworthy that the correlations from the random analyses are not centred around zero, due to the presence of the forcings.

The same resampling experiment was performed for the RMS error of the NH direct average. The mean RMS error was 0.62 (with a standard deviation of 0.13), ranging between 0.3 and 0.95. On the other hand, the RMS error found for the off-line DA analysis was only 0.11, falling well outside the above range. Similarly, for the NH mean, the mean RMS error of the random analyses was 0.51 (with a standard deviation of 0.10). The above results reveal that the DA analysis performs much better and is clearly outside the range of the randomly sampled distribution. The skill of the DA analysis is significantly different from the skill obtained from the random sampling.

3.3 Discussion

As previously noted, both DA schemes perform better than the simulations without DA, but there is not much difference in performance between them. In seven out of the 10 decades of the testing period, a lower cost function for the best member and the ensemble mean is found when using the on-line method, but the differences to the off-line approach are very small (Table 1). The respective ensemble mean (EM) cost functions are also shown in the table and are substantially larger than in the DA cases. Tables 2 and 3 summarise the Northern Hemisphere correlations and RMS errors respectively, between simulations and proxy-based reconstructions for the analysis and the ensemble mean of the two data assimilation schemes. The correlations and the RMS errors, on the continental scale and the hemispheric averages of the NH, are very close to each other. None of the two analyses can be deemed as better in following the proxy-based reconstruction. The similarity of the two analyses can also be seen in Figure 7, which shows the 2m mean temperature for the two analyses (anomalies w.r.t. the 1961-90 AD mean) and the 500hPa geopotential height (anomalies w.r.t. the 1961-90 AD mean) for the decade 1640-49 AD. Similar patterns can be seen, e.g. cool Barents Sea and warm NW Atlantic.

There are three potential reasons for the fact that the on-line method does not perform better than the off-line method: i) there might be no information propagation on decadal timescales in the model, ii) the simulated information propagation might be not skilful, i.e. different from reality, or iii) the ocean initial conditions used at the start of each decade in the on-line DA might be not sufficiently close to reality. A possible insufficient control of the ocean state would affect only the on-line method, as the off-line method is an a posteriori selection for which the ocean state is irrelevant.

While it is difficult and beyond the scope of this study to test whether the second and the third factor contribute to the similarity of skill of the two DA methods, we have assessed in a simple way whether there is any information propagation during the decadal sub-periods used in our DA. In the on-line assimilation all ensemble members are initialized with the same ocean state at the beginning of each decade. Therefore, if there is information propagation, one would expect less spread in the on-line ensemble than in the off-line one. We tested this by calculating the standard deviation of the ensemble spreads for the on-line and off-line methods for the different continents. The results are shown in table 4. For the NH direct average, we computed the standard deviation of the ensemble spreads for the whole period (for every year of the simulation period), as well as for the last year of each decade, and then computed the mean of these standard deviations. The standard deviations were 0.25 for the on-line compared to 0.30 for the off-line ensemble in the yearly test, and 0.28 compared to 0.31 respectively for the final year test. For the NH mean, the differences were a bit smaller. The standard deviations were 0.19 for the on-line compared to 0.23 for the off-line in the yearly test, and 0.22 compared to 0.23 respectively for the final year test. All the results show that the members are slightly closer together in the on-line experiment, a fact which is also in agreement with Figures 2 and 3. There is no significant difference between the all-year ensemble spreads and the spreads of the last year of each decade. The different spreads in the two DA approaches is evidence for the influence of the initialisation during the entire decadal assimilation time-step. The smaller spread in the on-line ensemble compared to the off-line one, which starts from different ocean initial states, is a hint for information propagation. However, we note that it is not clear from this analysis whether the information propagation is strong enough to lead to substantially higher skill of the on-line DA method.

As mentioned above the question whether the information propagation in the coupled GCM used here is realistic is difficult to answer and is linked to the question whether such models have skill in decadal predictions. The question whether the ocean state at the beginning of each assimilation decade is close enough to reality to be useful for bringing the ensemble members during the decadal assimilation cycle closer to reality can also not be answered here. Reasons why the ocean state might be unrealistic include a too small ensemble size, errors in the assimilated, proxy-based temperature reconstructions, and lack of control over the ocean states by assimilation of atmospheric variables.

Due to the specific choices of the approach and due to the wide range of alternative choices, the study is only a first step in the characterization of the interest of the on-line versus off-line approach. The differences between the two approaches may be specific to the target selected for the evaluation of the performance, the period investigated, the variable assimilated, the number of members in the ensemble, the frequency of assimilation, the assimilation method, and many other factors. A different setup could produce different conclusions that could prove the on-line DA scheme more skilful than the off-line one. Reasons for which the on-line DA is not better than the off-line DA in following the proxy-based reconstructions in our setup but could be more skilful in a different setup could be various. Firstly, the insufficient control of the ocean state could be due to the small ensemble size. If the ensemble size is too small to find a member that is close to the true climatic state, there will be no added skill by propagating this misleading information forward in time. A second reason for the initial state of the ocean not being accurately enough determined throughout the on-line assimilation could be that the selection of the best member was based on the atmospheric temperature state. A correct atmospheric state cannot guarantee that the ocean state is also determined correctly. A differently defined cost function, considering for example the global or direct average of the PAGES 2K continental reconstructions or different timescales could also change the performance of the two schemes. Another aspect that could have influenced our approaches is the proxy datasets. The use of proxies with the minimum possible noise would give a better chance to the on-line approach to capture the true climatic state, as they would represent the true climate better and the correct information would be propagated when applying the on-line approach, whereas the off-line one would not be benefited to the same extent, as it is an a posteriori selection. Finally, the use of a full particle filter rather than a degenerate one might produce a bigger ensemble spread for the ocean, giving again a better possibility to the on-line DA scheme to capture the true ocean state more closely.

4 Conclusions

Two main approaches have so far been employed to reconstruct the past climate: empirical and dynamical methods. Direct assimilation of proxy-based reconstructions into climate model simulations addresses some of the weaknesses of the two methods. Here, we have compared two ensemble-based DA schemes, an off-line and an on-line one, with the test case corresponding to the climate of the period leading into the Maunder Minimum, i.e. 1600-1700 AD.

The two DA schemes outperform the simulations without DA. The correlations between simulations and proxy-based reconstructions for the analyses of the DA schemes were higher than the correlations of the individual members, whilst the RMS errors were lower. The RMS errors of the ensemble means were lower than the errors of most of the individual members either due to the influence of forcings, or simply due to the lower variance of the ensemble mean compared to the individual members, but the DA analyses perform better, implying that some of the internal variability has been successfully captured by the DA. No big difference was found between the two approaches. The majority of the cost functions for the best member and the ensemble mean of the on-line DA method were found to be slightly lower than the ones of the off-line DA method, but the correlations and the RMS errors, at both the continental and the hemispheric level were very close to each other. The results suggest that there is either no skilful information propagation on the decadal timescales, i.e. no substantial predictability that could give the on-line DA an advantage over the off-line DA, or that the ocean states that are used at the beginning of each decade for generating the on-line ensembles are not sufficiently close to reality, and thus even if there was skilful predictability in the real and in the model world, the on-line DA could not benefit from it.

These results raise the question of which approach should be preferred in the future. In some cases, since the reconstruction skill of the on-line approach is not improved compared to the off-line equivalent, it would appear natural to use the less complicated off-line approach to DA, especially when computationally less expensive alternatives of off-line DA schemes can be used, for example when employing simulations that already exist. The temporal consistency of the simulation is eliminated in these cases though, which does not happen in the on-line approach. In the majority of the cases, and especially in the cases where the computational cost of the two methods is equal, the on-line approach should be preferred, as a result of the temporally consistent states that it provides.

Yet, we cannot be sure through these experiments whether a different setup could produce a better agreement for the on-line DA. Validation is only done with respect to the proximity to the proxy-based reconstructions, which is only a first step. We do not validate against the unknown true climate, as this would require pseudoproxy studies, which are beyond the scope of this paper. A differently defined cost function or different performance measures could also alter the comparison. Special care must be taken to make sure that the initial state of the ocean is being captured correctly throughout the on-line assimilation. A future direction for our work would be to test different setups, by employing the full rather than the degenerate particle filter, or by defining the cost function based on one- or thirty-year means instead of decadal means, in order to check whether ocean memory on those timescales leads to different results and maybe improvements to the on-line approach. More tests could be carried out by enhancing the ensemble size for both approaches or by using different proxy datasets.

Acknowledgements. A.M. is supported by a NERC studentship, the University of Birmingham and the Max Planck Institute for Meteorology in Hamburg. We would like to thank Helmuth Haak and Davide Zanchettin from MPI Hamburg for their support and guidelines on running the model and for useful discussions during the implementation of this work.

References

Annan, J. D. and Hargreaves, J. C.: Identification of climatic state with limited proxy data, Climate of the Past, 8, 1141–1151, 2012.

Annan, J. D., Crucifix, M., Edwards, T. L., and Paul, A.: Parameter estimation using paleodata assimilation, PAGES news, 21, 78–79, 2013.

Bhend, J., Franke, J., Folini, D., Wild, M., and Broennimann, S.: An ensemble-based approach to climate reconstructions, Climate of the Past, 8, 963–976, 2012.

Branstator, G., Teng, H. Y., Meehl, G. A., Kimoto, M., Knight, J. R., Latif, M., and Rosati, A.: Systematic Estimates of Initial-Value Decadal Predictability for Six AOGCMs, Journal of Climate, 25, 1827–1846, doi:10.1175/jcli-d-11-00227.1, 2012.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Blade, I.: The effective number of spatial degrees of freedom of a time-varying field, Journal of Climate, 12, 1990–2009, 1999.

Briffa, K. R., Jones, P. D., and Schweingruber, F. H.: Summer temperatures across Northern North-America - Regional reconstructions from 1760 using tree-ring densities, Journal of Geophysical Research-Atmospheres, 99, 25 835–25 844, doi:10.1029/94jd02007, 1994.

Bronnimann, S., Franke, J., Breitenmoser, P., Hakim, G., Goosse, H., Widmann, M., Crucifix, M., Gebbie, G., Annan, J., and van der Schrier, G.: Transient state estimation in paleoclimatology using data assimilation, PAGES news, 21, 74–75, 2013.

Crespin, E., Goosse, H., Fichefet, T., and Mann, M. E.: The 15th century Arctic warming in coupled model simulations with data assimilation, Climate of the Past, 5, 389–401, crespin, E. Goosse, H. Fichefet, T. Mann, M. E., 2009.

Crowley, T. J. and Lowery, T. S.: How warm was the medieval warm period?, Ambio, 29, 51–54, doi:10.1579/0044-7447-29.1.51, 2000.

Crowley, T. J. and Unterman, M. B.: Technical details concerning development of a 1200-yr proxy index for global volcanism, Earth System Science Data Discussions, 5, 1–28, doi:10.5194/essdd-5-1-2012, 2012.

Dirren, S. and Hakim, G. J.: Toward the assimilation of time-averaged observations, Geophysical Research Letters, 32, L04 804, doi:10.1029/2004GL021444, 2005.

Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., and Mann, M. E.: Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium, Climate Dynamics, 27, 165–184, 2006.

Goosse, H., Crespin, E., de Montety, A., Mann, M. E., Renssen, H., and Timmermann, A.: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, Journal of Geophysical Research-Atmospheres, 115, D09 108, doi:10.1029/2009jd012737, 2010.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M. F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly", Climate Dynamics, 39, 2847–2866, 2012.

Hakim, G. J., Annan, J., Brönnimann, S., Crucifix, M., Edwards, T., Goosse, H., Paul, A., van der Schrier, G., and Widmann, M.: Overview of data assimilation methods, PAGES news, 21, 72–73, 2013.

Hawkins, E. and Sutton, R.: Decadal Predictability of the Atlantic Ocean in a Coupled GCM: Forecast Skill and Optimal Perturbations Using Linear Inverse Modeling, Journal of Climate, 22, 3960–3978, doi:10.1175/2009jcli2720.1, 2009a.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1107, doi:10.1175/2009bams2607.1, 2009b.

Huntley, H. S. and Hakim, G. J.: Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model, Climate Dynamics, 35, 995–1009, 2010.

Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. amd Rahmstorf, S. a. R. R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D.: Palaeoclimate, in: Climate change 2007: the physical science basis. Contribution of working group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, pp. 433–497, 2007.

Jones, P. D. and Mann, M. E.: Climate over past millennia, Reviews of Geophysics, 42, RG2002, doi:10.1029/2003RG000143, 2004.

Jungclaus, J. H., Lorenz, S. J., Timmreck, C., Reick, C. H., Brovkin, V., Six, K., Segschneider, J., Giorgetta, M. A., Crowley, T. J., Pongratz, J., Krivova, N. A., Vieira, L. E., Solanki, S. K., Klocke, D., Botzet, M., Esch, M., Gayler, V., Haak, H., Raddatz, T. J., Roeckner, E., Schnur, R., Widmann, H., Claussen, M., Stevens, B., and Marotzke, J.: Climate and carbon-cycle variability over the last millennium, Climate of the Past, 6, 723–737, doi:10.5194/cp-6-723-2010, 2010.

Keenlyside, N. S. and Ba, J.: Prospects for decadal climate prediction, Wiley Interdisciplinary Reviews-Climate Change, 1, 627–635, doi:10.1002/wcc.69, 2010.

Ljungqvist, F. C.: A new reconstruction of temperature variability in the extra-tropical Northern hemisphere during the last two millennia, Geografiska Annaler Series a-Physical Geography, 92A, 339–351, 2010.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, Science, 303, 1499–1503, doi:10.1126/science.1093877, luterbacher, J Dietrich, D Xoplaki, E Grosjean, M Wanner, H, 2004.

Mairesse, A., Goosse, H., Mathiot, P., Wanner, H., and Dubinkina, S.: Investigating the consistency between proxy-based reconstructions and climate models using data assimilation: a mid-Holocene case study, Climate of the Past, 9, 2741–2757, doi:10.5194/cp-9-2741-2013, 2013.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, Proceedings of the National Academy of Sciences of the United States of America, 105, 13 252–13 257, 2008.

Mann, M. E., Zhang, Z. H., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F. B.: Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, Science, 326, 1256–1260, iSI Document Delivery No.: 524BD, 2009.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Roske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, Ocean Modelling, 5, 91–127, doi:10.1016/s1463-5003(02)00015-x, 2003.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., Karlen, W., and Lauritzen, S. E.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data (vol 433, pg 613, 2005), Nature, 439, 1014–1014, doi:10.1038/nature04575, 2005.

PAGES 2K Consortium: Continental-scale temperature variability during the past two millennia, Nature Geoscience, 6, 339–346, 2013.

Pendergrass, A. G., Hakim, G. J., Battisti, D. S., and Roe, G.: Coupled Air-Mixed Layer Temperature Predictability for Climate Reconstruction, Journal of Climate, 25, 459–472, 2012.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochemical Cycles, 22, Gb3018, doi:10.1029/2007gb003153, 2008.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K. G., Wetzel, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, Climate Dynamics, 29, 565–574, doi:10.1007/s00382-007-0247-8, 2007.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development, 4, 33–45, doi:10.5194/gmd-4-33-2011, 2011.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction, Journal of Climate, 27, 426–441, doi:10.1175/jcli-d-12-00693.1, 2014.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the Earth System Model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146–172, doi:10.1002/jame.20015, 2013.

Van Den Dool, H. M.: Searching for analogues, how long must we wait, Tellus A, 46, 314–324, 1994.

Vieira, L. E. A., Solanki, S. K., Krivova, N. A., and Usoskin, I.: Evolution of the solar irradiance during the Holocene, Astronomy and Astrophysics, 531, A6, doi:10.1051/0004-6361/201015843, 2011.

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J.: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, Climate of the Past, 6, 627–644, 2010.

Xoplaki, E., Luterbacher, J., Paeth, H., Dietrich, D., Steiner, N., Grosjean, M., and Wanner, H.: European spring and autumn temperature variability and change of extremes over the last half millennium, Geophysical Research Letters, 32, doi:10.1029/2005gl023424, 2005.

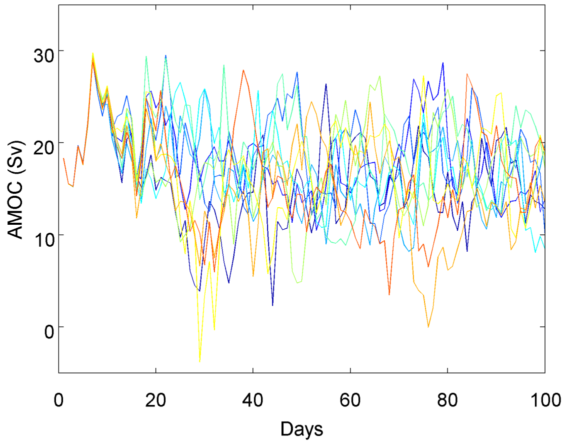

Figure 1. 26.5° AMOC time-series of the ensemble spread for the first 100 days after the initialisation of the ensemble in year 1600 AD, measured in Sverdrups (1 Sv = 10^6 m^3s^{-1}).

Table 1. Best cost functions for the off-line and the on-line DA schemes, for the decades 1 (1600-1609) to 10 (1690-1699). The respective ensemble mean (EM) cost functions are also shown.

Decade	Off-line Best	On-line Best	Off-line EM	On-line EM
1	1.47	1.47	2.53	2.53
2	1.44	1.55	1.96	2.00
3	0.51	0.45	1.31	0.86
4	1.71	2.10	2.39	2.49
5	0.72	0.60	1.57	1.14
6	1.04	0.50	1.65	0.95
7	0.53	0.49	1.22	0.97
8	0.62	0.38	1.66	1.62
9	1.72	0.66	2.28	2.21
10	1.46	1.45	1.97	1.93

Table 2. Northern Hemisphere correlations between simulations and proxy-based reconstructions for the analysis and the ensemble mean of the two data assimilation schemes.

	Arctic	Asia	Europe	N. America	NH dir. aver.
Off-line DA analysis	0.56	0.78	0.79	0.89	0.94
On-line DA analysis	0.79	0.76	0.79	0.81	0.93
Off-line DA EM	0.32	0.55	0.58	0.66	0.73
On-line DA EM	0.07	0.67	0.38	0.64	0.67

Table 3. Northern Hemisphere RMS errors between simulations and proxy-based reconstructions for the analysis and the ensemble mean of the two data assimilation schemes.

	Arctic	Asia	Europe	N. America	NH dir. aver.
Off-line DA analysis	0.24	0.19	0.18	0.13	0.11
On-line DA analysis	0.18	0.21	0.16	0.18	0.12
Off-line DA EM	0.23	0.19	0.24	0.18	0.12
On-line DA EM	0.22	0.21	0.24	0.19	0.11

Table 4. Standard deviations of the ensemble spreads for the Northern Hemisphere of the two data assimilation schemes, calculated for all the years and for the last year of each decade.

	Arctic	Asia	Europe	N. America	NH dir. aver.
Off-line all years	0.48	0.28	0.50	0.41	0.30
On-line all years	0.42	0.28	0.46	0.37	0.25
Off-line last year	0.49	0.32	0.49	0.40	0.31
On-line last year	0.47	0.30	0.49	0.38	0.28

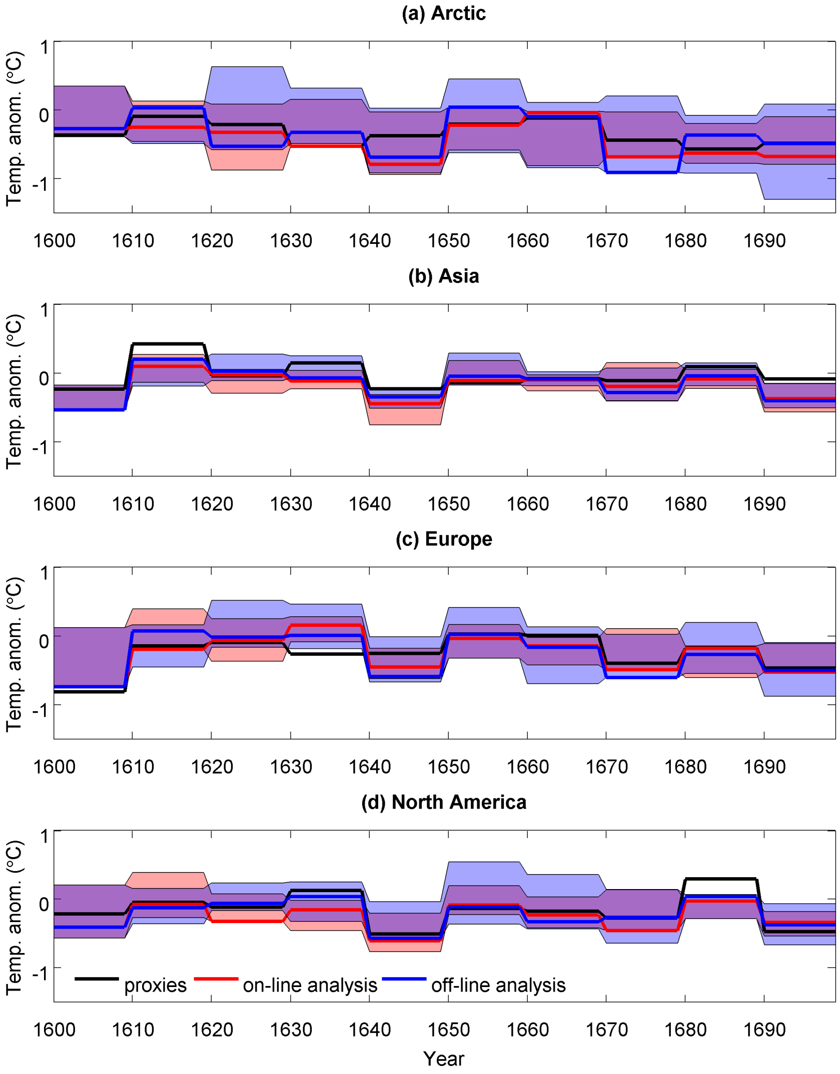

Figure 2. Continental decadal mean temperature anomalies w.r.t. the 850-1850 AD mean in the Northern Hemisphere for the 17th century, for the on-line (red shading) and off-line (blue shading) ensemble members, the on-line (red line) and off-line DA analysis (blue line) and the proxy-based reconstructions (black line).

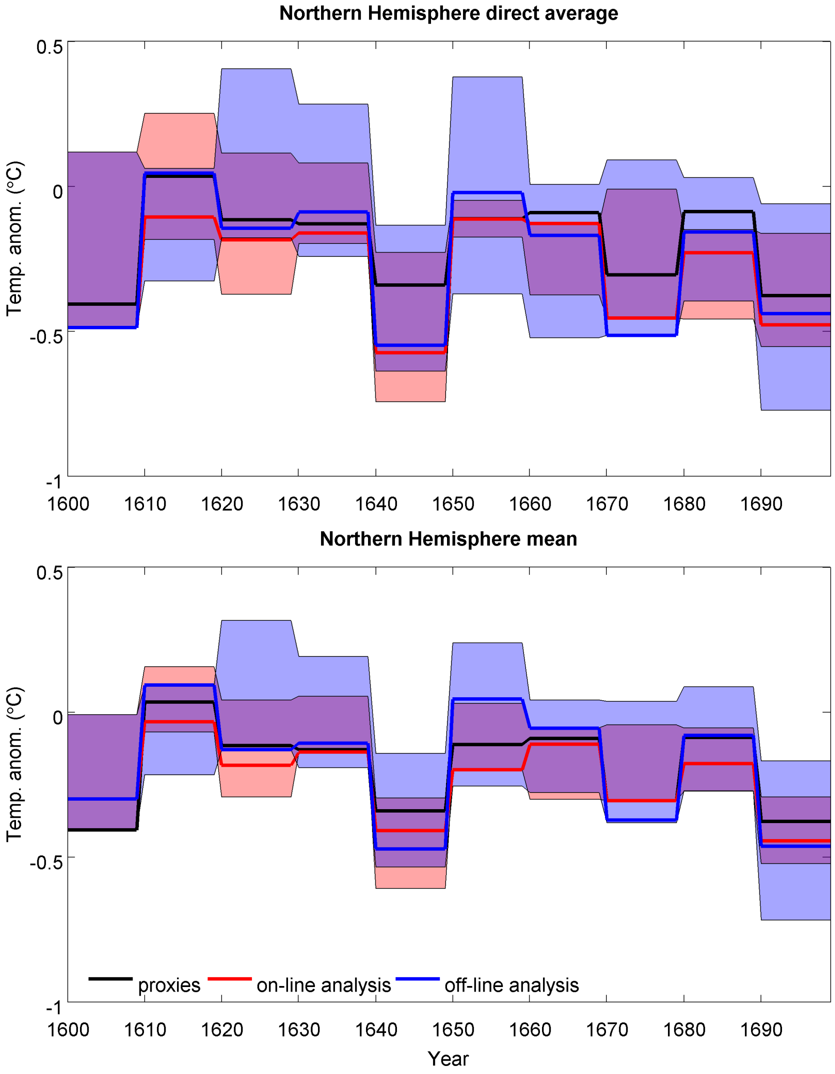

Figure 3. Direct average of the four Northern Hemisphere continental temperatures (anomalies w.r.t. the 850-1850 AD mean) and NH mean for the 17th century, for the on-line (red shading) and off-line (blue shading) ensemble members, the on-line (red line) and off-line DA analysis (blue line) and the proxy-based reconstructions (black line).

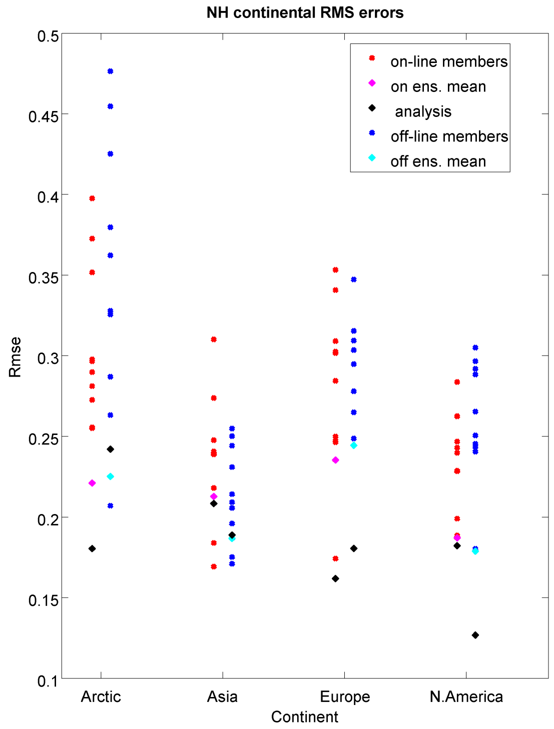

Figure 4. RMS errors for the four Northern Hemisphere continents for the 17th century, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (magenta dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

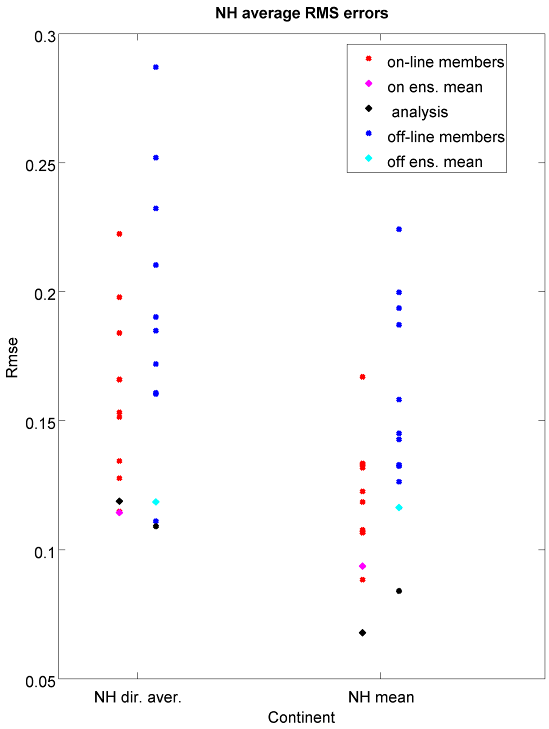

Figure 5. RMS errors for the direct average and the mean of the Northern Hemisphere for the 17th century, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (magenta dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

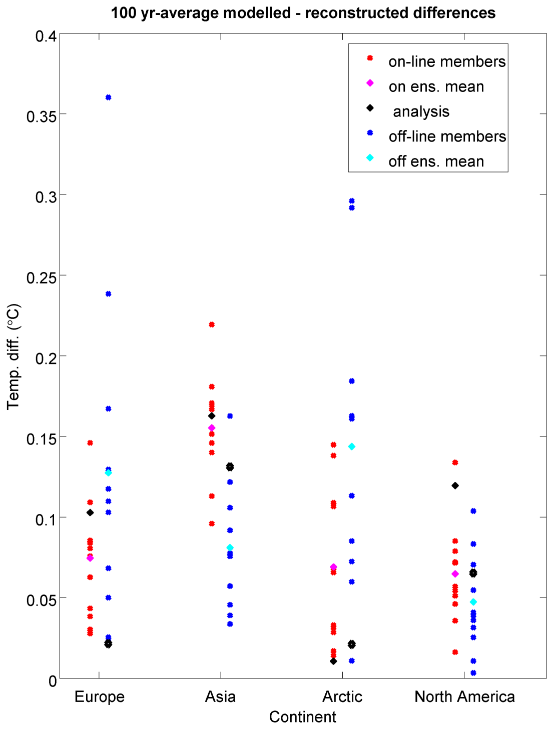

Figure 6. Average 17th century temperature differences between model and reconstructions, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (magenta dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

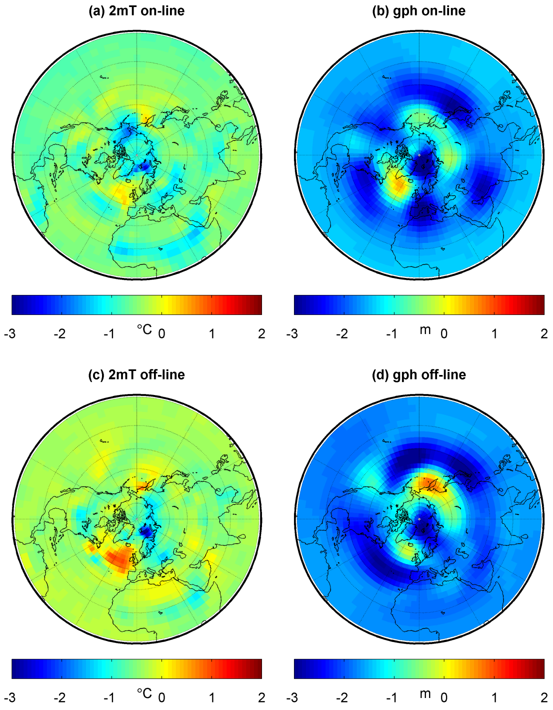

Figure 7. Analyses of the on-line and off-line DA methods for the 2m mean temperature (anomalies w.r.t. the 1961-90 AD mean) and 500hPa geopotential height (anomalies w.r.t. the 1961-90 AD mean) of the decade 1640-49 AD.