Ensemble-based Data Assimilation for the Climate of the Past Millennium

Anastasios Matsikaris

A thesis submitted to the University of Birmingham for the degree
of DOCTOR OF PHILOSOPHY.

School of Geography, Earth and Environmental Sciences
College of Life and Environmental Sciences
University of Birmingham
March 2016

UNIVERSITYOF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Data assimilation (DA) is an emerging research area in palaeoclimatology and one of the key challenges in this field. Different ensemble-based DA approaches have recently been undertaken and shown that DA has the potential to provide improved reconstructions for the climate of the past, and increase our understanding of climate variability. However, no systematic comparison among them has so far been attempted, and large knowledge gaps remain regarding the ideal characteristics for providing the best climate state estimates. In this thesis, ensemble-based DA schemes are implemented and evaluated for the reconstruction of the climate of some of the key periods from the past millennium. The study is among the first to employ a General Circulation Model for palaeoclimate DA.

An off-line and an on-line DA method are first compared, assimilating continental proxy-based temperature reconstructions and using the 17th century as testing period. Both schemes provide simulations that follow the assimilated targets on large scales better than without DA. The on-line method, with which the ensembles are generated sequentially for sub-periods based on the analysis of previous sub-periods, would be expected to perform better than the off-line one if the system had sufficient memory to propagate realistic information forward in time, however, similar skill was found for the two methods. Nevertheless, the on-line scheme has the advantage of temporal consistency of the analysis, and is subsequently used to reconstruct the climate for 1750-1850 AD. The assimilation performs well on large-scale temperatures, but there is no agreement between the DA analysis and reconstructions for regional temperature patterns. Evidence is presented to suggest that this lack of information propagation to smaller spatial scales is likely due to the fact that the Northern Hemisphere continental mean temperatures are not the best predictors for large-scale circulation anomalies, or that the assimilated reconstructions include noise. The lack of regional skill is again found when instrumental data for 1850-1949 AD are assimilated. Based on these results, it is argued that a potential way of improving the performance of DA is the assimilation of temperature reconstructions with higher spatial resolution.

Στους γονείς μου

Acknowledgements

Contents

List of Figures

List of Tables

List of Acronyms

DA	Data Assimilation
SH	Southern Hemisphere
NH	Northern Hemisphere
EBM	Energy Balance Model
EMIC	Earth System Model of Intermediate Complexity
GCM	General Circulation Model
MCA	Maximum Covariance Analysis
SLP	Sea Level Pressure
MPI-ESM	Max Planck Institute for Meteorology Earth System Model
MWP	Medieval Warm Period
LIA	Little Ice Age
PAGES	Past Global Changes
CMIP	Coupled Model Intercomparison Project
IPCC	Intergovernmental Panel on Climate Change
ENSO	El Niño Southern Oscillation
NAO	North Atlantic Oscillation
AO	Arctic Oscillation
PMIP	Paleoclimate Modelling Intercomparison Project
IGBP	International Geosphere–Biosphere Programme
NAM	Northern Annular Mode
MOC	Meridional Overturning Circulation
RMS	Root Mean Square
BEST	Berkeley Earth Surface Temperature
SVD	Singular Value Decomposition
TEC	Time Expansion Coefficient
EOF	Empirical Orthogonal Function
w.r.t.	with respect to

Chapter 1

Introduction

Reconstructing the climate of the past is crucial for quantifying and understanding natural climate variability on long timescales, which in turn is essential for detecting anthropogenic climate change. Furthermore, examining the response of the climate system to large climate forcing changes of the past is a valuable tool for assessing its potential response to future changes. For instance, the warmer-than-today climate of the Mid-Pliocene (about 3.3 to 3.0 Ma BP), which was associated with higher CO_2 levels, might be instructive for predictions of the climate response to the increased greenhouse gas levels of the future (Jansen et al., 2007). In addition, palaeoclimate reconstructions are important for the validation of climate models that are used to provide future climate projections, especially as not all model aspects can be tested against instrumental data (e.g. sensitivity to CO_2).

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. Both approaches are associated with substantial uncertainties. One of the emerging research areas in palaeoclimate modelling is the systematic combination of empirical information from proxy data with the representation of the processes that govern the climate system given by

climate models. The process of incorporating observations of a system into the numerical model state of that system is known as "data assimilation" (DA), and is applied in various fields of geoscience. DA is used widely in meteorology for weather forecasting, but adapting it to palaeoclimatic applications is not straightforward. Therefore, DA can be considered as one of the key challenges in palaeoclimatology.

In this thesis, ensemble-based DA is implemented for reconstructing the climate of the past, with the attention being focused on some key periods of the past millennium. The availability of both climate model simulations and climate reconstructions from proxy data for this period is constantly increasing (Sundberg et al., 2012). This allows a systematic evaluation of the models against the reconstructions, and provides the potential to advance our understanding on climate variability. 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). Different DA approaches for climate reconstructions of the last millennium or earlier periods have recently been undertaken, however, large knowledge gaps regarding the ideal scheme that can provide the best state estimates remain.

The present chapter serves as a brief overview of the scientific background needed for the research presented in the following chapters of the thesis. Section 1.1 presents the main types of proxies and proxy-based climate reconstructions, along with the principal uncertainties associated with them. The fundamentals of the climate models and their primary strengths and weaknesses are described in Section 1.2. Section 1.3 provides the basic concepts of DA and the main differences between DA applications in meteorology and palaeoclimatology, followed by a brief literature review of the palaeoclimatic DA approaches applied recently. Finally, the main objectives and the structure of the thesis are presented in Section 1.4.

1.1   Reconstructions from proxy data

As widespread instrumental climate records do not extend to the period before the mid-19th century, the use of natural climate archives becomes necessary for climate reconstructions of earlier periods. Indirect indicators or "proxy measures" of climate variability are provided by natural archives present in our environment, and include tree-rings, ice cores, corals, glacier ice, pollen, speleothems, boreholes, stalagmites, lake and marine sediments (Mann et al., 2008). The chemical, physical or biological parameters of these proxies are altered due to environmental changes, with the climatic phenomena being reflected in different temporal resolutions (Jansen et al., 2007); for example, sediment cores resolve centennial or longer timescales, while growth and density measurements from tree rings are seasonally or annually resolved. Using statistical models, the reconstructions from proxy data are calibrated against instrumental data during common periods of overlap. In addition, historical documentary records for the climate of the past centuries are also used as a source of information.

Two main types of proxy-based reconstructions have been employed by palaeo-climate studies; those for large-scale mean temperatures, 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 those for spatial pattern temperature reconstructions (e.g. Briffa et al., 1994; Luterbacher et al., 2004; Jones and Mann, 2004; Xoplaki et al., 2005; Mann et al., 2009). Moreover, other studies have focused on proxy-based reconstructions for different climate variables or indices, such as hydroclimatic or regional circulation reconstructions (for a review see Jones and Mann, 2004). The reconstructions are usually based on a relationship between the proxy and a particular climate variable over the calibration period, which is assumed to be stationary; however this is often not

the case, introducing an additional caveat to the reconstructions (Jones and Mann, 2004).

Various different statistical methods are followed to convert the raw proxy series into climate reconstructions. Large-scale temperature reconstructions usually follow either regression-based techniques, or the "composite plus scale" (CPS) methodology. In the latter method, proxy records sensitive to surface temperature variations are standardized and composited (averaged) to form a regional or hemispheric series. This series is then scaled against the target instrumental series, so that its mean and variance match those of the observed data over some overlap period (Mann et al., 2008). Climate field reconstructions (CFR) employ multiple linear regression techniques, often using principal components for the predictors and the predictands, or other advanced techniques (Mann et al., 2009). Proxy indicators are regressed on instrumental data during the overlap period, and the developed regression models are applied to the earlier proxy records for reconstructing the past climate. Intense discussion has however been roused regarding the loss of centennial and multi-decadal variability from which many regression methods suffer (e.g. Von Storch et al., 2004).

Proxy-based estimates of climate variability are associated with large uncertainties, caused by various factors and not always well understood. For example, different proxies usually represent different seasons, different statistical methods used in the reconstructions lead to different results, while non-climatic factors often influence the proxies (e.g. Jones and Mann, 2004; Jansen et al., 2007). Non-climatic noise can mask inter-regional relationships of the true climate or conceal the response to the forcings. Furthermore, only a limited set of variables can be reconstructed. The inadequate understanding of relationships between proxies and environmental variations remains another weakness of proxy-based reconstructions. Lastly, poor spatial coverage of the climate proxies leads to errors in hemispheric or continental

means and even larger errors in full-field reconstructions. For instance, the paucity of proxy data in the Southern Hemisphere (SH) and the tropics makes their respective reconstructions very weak (e.g. Mann and Jones, 2003; Neukom et al., 2014). For these reasons, computer models are considered by many to be a necessary tool in the interpretation of the proxies and filling of the gaps.

1.2 Modelling past climate changes

Climate models are indispensable for the development of a quantitative understanding of the mechanisms behind past climate changes (Bothe et al., 2015). 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. Models use reconstructed external forcings to produce simulations of the past climate. Such forcings are either natural or anthropogenic, and include variations in the Earth's orbital parameters, solar irradiance, volcanic activity, greenhouse gas concentrations, atmospheric aerosols, land cover and land use. As many processes, e.g. cloud formation, boundary-layer conditions, hydrological processes, and radiation, occur on scales below the model resolution, they need to be "parameterised". Real-world observations are used to define the parameterisation schemes used in the simulation of these sub-grid scale processes (e.g. O'Neill et al., 2015).

Climate models can be distinguished in different types, according to their resolution and their physical description of the climate system, ranging from simple conceptual Energy Balance Models (EBMs) to the most complex General Circulation Models (GCMs). Of an intermediate type are the Earth system Models of Intermediate Complexity (EMICs). EBMs include a very simplified representation of the climate system, with, in principle, no dynamics. Variability in the atmosphere

is not accounted for, therefore the simulations are largely a response to the external forcings. EMICs are usually quasi-geostrophic models, based on simplified atmospheric and ocean dynamics. They are thus computationally less expensive and faster than GCMs, and have a coarser resolution. This advantage makes them ideal for the performance of a large number of simulations at a lower cost. GCMs are highly complex dynamical models, which include a much more realistic representation of the processes involved with the Earth's atmosphere, oceans and land surface, as well as the interactions between them. The climate system in most GCMs is simulated based on the primitive equations, which include the hydrostatic approximation, while some high resolution GCMs are non-hydrostatic.

Advanced climate models provide reasonably skilful simulations of multi-decadal to centennial climate variability on hemispheric and global spatial scales (Jansen et al., 2007). This skill largely owes to the dominant role of forcings on the large scales. Agreement of simulations and observations is very unlikely on smaller spatial scales and shorter timescales, where the role of internal variability cannot be neglected (Hawkins and Sutton, 2009a). In principle, interannual to decadal temperature variations have a large random, non-forced component, therefore even a perfect model that includes all the forcing mechanisms would yield only one of the many possible climate realisations consistent with the prescribed forcings (PAGES 2K Consortium, 2013). The internally generated climate fluctuations are superimposed over any externally forced climate response. Ensemble simulations are indispensable in better quantifying the internal variability for periods within the last millennium (e.g. Jungclaus et al., 2010). Two aspects of internal variability can be distinguished: the spatial structure of variability, which is potentially captured by good GCMs; and the temporal characteristics of variability, which are random and which DA attempts to capture. EMICs include contributions from internal climate variability, but the structure of spatial variability in the GCMs is more realistic, owing to the more realistic representation of climate dynamics. This includes the realistic

simulation of important properties of internal variability, such as teleconnection patterns between regions, as well as features of global climate, such as the jet streams, atmospheric circulation and location of intertropical convergence zones (Zorita and von Storch, 1999).

Further to the role of internal climate variability, the skill of model simulations is limited by uncertainties mainly due to model errors (e.g. Bothe et al., 2015; Maraun et al., 2015). The model error consists of systematic model biases and errors in the climate forcings or in the response to them; the forcings are reconstructed following methods similar to the proxy-based reconstructions, and do not precisely determine the temporal evolution of the climate, in particular on regional scales. Additionally, the accuracy of the parameterisations used in the models also limits the simulation skill, as do the physics and dynamics contained (Jones and Mann, 2004).

1.3   Data assimilation approaches

The general aim of DA is to incorporate empirical information about a system state into a dynamical model of the system, in order to obtain estimates of the current and future state variables that are closer to reality. DA is applied in many different fields, such as meteorology, oceanography, hydrology, extraterrestrial planetary sciences, atmospheric chemistry, petroleum reservoir engineering, while applications expand to cases like the trajectory estimation for the Apollo program. In numerical weather forecasting, where DA is highly developed, the assimilation attempts to find the best estimate ("analysis") for the current state of the atmosphere, to be used as a starting point for the forecast. This problem is solved by combining a previous forecast with observations. A huge number of daily in situ and remote sensing observations of the state of the atmosphere and the ocean are assimilated for this purpose. In palaeoclimatology, however, the application of DA needs to overcome

many limitations and the methods need to be adapted.

1.3.1 Basic framework of data assimilation

Two main types of DA schemes are applied in meteorology; "four-dimensional DA", where all available observations over a given time window are used simultaneously to minimise a cost function for obtaining an optimal analysis for all the states in that window; and "sequential methods", where observations are assimilated one at a time and the estimated states are updated whenever an observation is incorporated (Law and Stuart, 2012). In the former category, the four-dimensional variational DA (4D-Var) is the most broadly used scheme in numerical weather forecasting. The most common sequential approach is filtering, e.g. the particle filter, a version of which is implemented in this thesis, and the Kalman filter. The filters update the posterior distribution of the system state sequentially at every observation time (Law and Stuart, 2012). A method similar to filtering is smoothing, but instead of going backwards in time to incorporate data, it moves back and forth in time. Another sequential method is the three-dimensional variational DA (3D-Var), which finds directly approximate solutions by minimising a cost function.

In the general case of sequential DA, the aim is to find an optimal analysis Ψ^a, given a set of model variables contained in a vector Ψ. The computation of the analysis can be represented by the equation:

$$\Psi^a = \Psi^b + G(H(\Psi^b) - \Theta) \tag{1.1}$$

where Ψ^b is the previous forecast of the vector Ψ, also known as the background field, and Θ is a set of observed variables. H is the observation operator, which performs a transformation of the simulated state Ψ at a given time into the observations Θ . This transformation is necessary as the simulated states and the observations

are fundamentally different quantities, not usually obtained at the same locations (Widmann et al., 2010). It can include simple spatial interpolations from the model grid to the observation locations, as well as more complicated relations between the system state and the observations. G is the so-called gain matrix, accounting for uncertainties in model results and observations. The gain matrix is calculated from the error covariance matrices of simulations and observations.

Kalman filtering is one of the principal sequential approaches to DA and is today used in a wide range of applications. As mentioned above, it estimates the state of a dynamical system from a series of incomplete measurements through linear functional relationships, applying the equation 1.1 (e.g. Bhend et al., 2012). The filter uses the system dynamics and their effect on Gaussian white noise, the observational errors, as well as the relationships between observations and state variables, and produces an optimal mean square estimate of the system state. It also provides an estimate of the error.

The actual implementation in the particle filter is different, as it does not attempt to apply the equation 1.1, but follows a simpler approach. An arbitrary initial compilation of particles (the possible states in which the model output could be in) needs to exist. The particles are given weights after being compared with measurements or proxy data. The weights need to be normalized, namely the sum of all the weights has to be equal to one. The particle filter algorithm filters out the particles that are considered poor representations of the actual state, whereas highly consistent particles are promoted. In a full particle filter, a new set of particles is generated after the sampling of the weighted set. If a degenerate particle filter is used, only the best fitting particle gets through the filtering.

1.3.2   Data assimilation in palaeoclimatology

The motivation for using DA in palaeoclimatology arises from various factors. As mentioned previously, both the empirical and dynamical methods that have been employed to reconstruct the past climate have different drawbacks and are associated with important uncertainties. DA combines the two approaches aiming to find better estimates for past climate states, which are consistent with the empirical knowledge and the dynamical understanding of the climate system. The assimilation can potentially capture the real-world random, internally generated variability, the temporal evolution of which cannot be captured in forced simulations. DA also accounts for unrealistic forcings, or for unrealistic responses to forcings in climate models. Moreover, it provides complete spatial fields, as well as information for variables for which no empirical estimates exist (Widmann et al., 2010). In general, the empirical data in the DA methods are used 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). In this thesis, the case of state estimation is considered.

The assimilation problem in palaeoclimatology is challenging and different to the respective problem in numerical weather forecasting in many respects. As already stated, in contrast to the wealth of direct and indirect measurements available for initialising the weather forecasts or for reanalysis projects, empirical estimates for past climate states are limited and the spatial coverage is incomplete. Climate proxies are only found in a few locations. Moreover, the type of the data is fundamentally different; the proxy data constrain seasonal and longer variability rather than individual weather states. The information propagation on these temporal scales is uncertain. The technical problems arising from the integration of observations over long periods lead to difficulties in the use of standard assimilation

methods, further to the large model and proxy errors. In addition, the need for the methods to be efficient enough for long simulations and the high computational cost required introduces another limitation to the use of palaeoclimate DA.

Only a few attempts of applying DA for palaeoclimate applications have so far been undertaken, and include different approaches that fall into two main categories. In the first approach, EMICs or GCMs are used to run ensemble simulations and the members that are most consistent with proxy-based reconstructions are selected, based on filtering (e.g. Goosse et al., 2006; Bhend et al., 2012). In an alternative approach, the model equations themselves are modified to add terms that push the model towards the observations. Atmospheric circulation target states based on proxy evidence are prescribed, using forcing singular vectors and EMICs (Van der Schrier and Barkmeijer, 2005), or pattern nudging and GCMs (Widmann et al., 2010). All these palaeoclimate DA approaches are formulated within the standard (Bayesian) framework of classical DA, but cannot be easily classified into either sequential or four-dimensional DA methods; this is because even though the assimilation is performed sequentially, the assimilated data are averaged over long time windows.

"Ensemble member selection" techniques, like the one implemented in this thesis, are the most frequently used by the community, having the advantage of being easy and straightforward to implement, with no strong modifications of model code required. The "forcing singular vectors" technique determines small perturbations to the time evolution of the prognostic variables of the model, leading the atmospheric model to a pre-defined target pattern. These perturbations, referred to as singular vectors, result in large perturbation growth during a short time period of several days (Van der Schrier and Barkmeijer, 2005). The approach of "pattern nudging" is similar to the forcing singular vectors; additional forcing (nudging) terms are added to the prognostic model equation to pull the states of

the model towards the observations (Widmann et al., 2010). The aforementioned techniques have only been the first step towards the assimilation of proxy-based data. The provided reconstructions have been found to be consistent with empirical knowledge over large parts of Europe, offering spatially complete information for many variables. It was however noted that if the target pattern is not within the patterns simulated by the model, a pattern close to the target state cannot be well approximated (Widmann et al., 2010). In the following subsection, some of the key studies employed for reconstructing the climate of the past one to two millennia are described in more detail, focusing on the methodology followed in each study.

1.3.3 Brief overview of palaeoclimate DA studies

The ensemble member selection techniques were pioneered by Goosse et al. (2006), who employed at first a simplified global 3-D climate model (ECBILT-CLIO-VECODE model). The coarse resolution of the model and the simplified dynamics made it one to two orders of magnitude faster than a state-of-the-art atmosphere-ocean GCM. A total of 105 simulations were performed, covering the last millennium or longer. The best model analogs were selected after the completion of the simulations, an approach named "off-line" DA. After comparison of each simulation with a set of 12 proxy-based Northern Hemisphere (NH) temperature time series, the optimal simulation for 10 to 50 years averaging sub-periods was selected. These simulations were grouped together in order to form a consecutive pseudo-time series, to allow for direct comparisons with the assimilated and independent climate reconstructions. An updated version of this technique was employed by Goosse et al. (2010), using a more advanced 3-D EMIC (LOVECLIM), combined with a set of 56 surface temperature proxy series derived from a comprehensive compilation of Mann et al. (2008). The main difference to the previous DA scheme was that a new ensemble was generated at each step of the assimilation procedure starting from the

analysis of the previous sub-period, an approach named "on-line" DA. An ensemble of 96 realizations was generated at each assimilation step. The revised method offered dynamical consistency between best model analogs of different periods, while the former benefited from its computational simplicity.

The two methods demonstrated positive reconstruction skill, with the DA-simulated time series following the temperature signal recorded in the proxies well, particularly in areas with high assimilated data coverage. Furthermore, the DA time series for most regions were strongly correlated with the corresponding instrumental data. In the Goosse et al. (2010) study, the agreement for Europe, Asia, North America, and the NH mean temperatures was reasonably good, with correlations between 0.72 and 0.86. The DA time series were also similar to previously published statistical proxy-based reconstructions. The results were poor for the temperatures of the North Atlantic region, which can mainly be attributed to the paucity of proxy data there to constrain the model results. Additional sources of bias were the inability of the EMIC to simulate large multidecadal variations there, and weaknesses of the DA technique itself, especially the fact that the initialization of the ensemble does not guarantee adequate perturbations in the ocean dynamics. The on-line DA method was also employed by Crespin et al. (2009) to analyse the 15th century Arctic warming. The comparison of the on-line and off-line DA approaches is the main objective of Chapter 3.

In addition to the above methods, where a single simulation having the best fit to the data was 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 ("full particle filter"). In this approach (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 are being discarded and the highly likely

ones are being copied proportionally to their likelihood. The technique was applied by Annan and Hargreaves (2012) using pseudoproxy experiments, to evaluate the precision of DA when limited observations are available. The study used 101 ensemble members generated by the LOVECLIM model, with one of the members being chosen each time as the "truth", i.e. the target. The pseudoproxy data were based on the screened network of Mann et al. (2008). The assimilation was performed off-line, i.e. after the completion of the simulations. The method achieved significant skill, shown by the correlations between targets and DA time series. Experiments with constant forcing, accounting for the internal variability of the model, showed that the analysis converged more towards the target when the number of the assimilated proxy data points increased. The skill on the large-scale mean values was not combined with good performance on regional scales, especially for regions far from the proxy locations and when low proxy data density was employed.

A similar "probabilistic posterior distributions" technique was used by Goosse et al. (2012) to investigate the role of forcing and internal dynamics during the Medieval Climate Anomaly, 950-1250 AD. The need for further investigation of this period stemmed from the fact that models usually underestimate the magnitude of the reconstructed regional temperature changes or disagree in timing with the reconstructions (Goosse et al., 2012). An ensemble of 96 simulations was generated using the LOVECLIM model, by adding noise to the atmospheric streamfunction of the model. The assimilation employed a temperature reconstruction based on a global temperature network of more than 1,000 proxy records, derived by Mann et al. (2009). After the first year of simulations, each particle was assigned a likelihood, calculated from its difference with the proxy-based reconstruction interpolated on the model grid, on all available points northward of 30°N. The members were either stopped or copied a number of times proportionally to their likelihoods, and the procedure was repeated for every year until the end of the

period. The outcomes led to distributions with larger overlaps and improved agreement with the proxy-based reconstructions. The same DA scheme was used by Mairesse et al. (2013) to reconstruct the climate of the mid-Holocene (6000 BP). Three sets of simulations were performed; one where air and sea surface temperature reconstructions constrained the model, and two others where only the continental or oceanic records were used. The simulation constrained by all reconstructions was found to be 15% closer to the assimilated reconstructions compared with simulations without DA. Finally, Klein et al. (2014) followed the same approach to reconstruct the mid-Holocene Arctic sea ice concentration.

Other ensemble-based DA studies employed the Kalman filter or variants of it. The off-line approach of DA was advanced by Bhend et al. (2012), through the use a high-resolution GCM (ECHAM5.4 model). The ensemble square root filter, a variant of the ensemble Kalman filter, was used to update the ensembles with climate proxy information. The use of an atmosphere-only GCM left no possibility for information propagation over long timescales; therefore, the assimilation was performed off-line. An on-line DA scheme would not have benefited the reconstruction skill, apart from leading to temporal consistency of the analysis. Thirty ensemble members were run for the period 1871-2005 AD. The 30th member was selected as the target time series and was used for validation, while the other 29 members represented the unconstrained ensemble. Pseudo-proxies for temperature at 37 different locations were generated from the target simulation. The assimilation improved the representation of temperature compared with simulations without DA, and also performed relatively well in other surface (and even upper air) features. Skill was higher in the NH, where the assimilated information was, and for variables in the boreal winter, increasing with the ensemble size. Considerable skill in regions close to the assimilated data was found for ensembles of 15 members or more, while a larger ensemble size was necessary for areas further away. The approach also allowed a generic quantification of the uncertainties, unlike most previous DA

approaches.

Dirren and Hakim (2005) used an idealized one-dimensional model to examine a novel algorithm for the case where only time-averaged observations are available. The algorithm was a natural extension of the ensemble Kalman filter and reduces to the ensemble Kalman filter in the limit of zero time averaging. Huntley and Hakim (2010) applied the new algorithm to test the method with a simple quasi-geostrophic model of a mid-latitude atmospheric jet that interacted with a mountain. The model was not sufficiently complex to capture palaeoclimate variability, but allowed testing over a range of observations and model configurations. The cycling of 50 ensemble members over 50 averaging periods made this approach far cheaper than a traditional implementation of the Kalman filter. Similarly, Pendergrass et al. (2012) tested two idealized models which captured adequate climate variability related to palaeoproxies. The skill of the technique was lower when the observation averaging time increased. 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. A single simulation was used as the background ensemble, with the ensemble members being the individual years instead of independent model simulations. This resulted in great computational cost savings, however the ensemble was not time-dependent and relied on the assumption that no information propagation exists in the atmosphere.

1.4 Present work and thesis structure

The primary aim of the research presented herein is the implementation and testing of DA methods using a GCM for the reconstruction of the past climate, focusing on some key periods of the last millennium and benefiting from the complementarity of simulations and proxy data. In this framework, we employ ensemble simulations

and select the ensemble members that are the closest to palaeoclimate proxy-based reconstructions. This approach has a sound theoretical foundation as it is equivalent to the aforementioned particle filter. The technique has already been successful in applications with EMICs (e.g. Goosse et al., 2010). The increase in computing power has now allowed for the method to be attempted using a GCM. Due to the considerably better representation of the mean climate and climate variability in GCMs compared to EMICs, it can be expected that the use of a GCM can provide more realistic reconstructions.

The thesis consists of four main studies (Chapters 2-5), followed by a conclusion (Chapter 6) in which the key findings are summarised and the scope for further development is discussed. Within each chapter, the relevant literature is reviewed, and descriptions of the model and data used as well as methodology details are provided. The research has led to three publications, in the journals "Climate of the Past" (Matsikaris et al., 2015a), "Climate Dynamics" (Matsikaris et al., 2015b), and "Climate of the Past Discussions" (Matsikaris et al., 2016), which correspond to the Chapters 3, 4 and 5 respectively. The climate model used throughout the thesis is the Max Planck Institute for Meteorology Earth System Model, MPI-ESM. The model is documented in Chapter 2, and evaluated against observations, reconstructions and other GCMs. This evaluation primarily focuses on large spatial scales, but also investigates the inter-regional patterns of temperature variability. Additionally, the content of Chapter 2 includes a general description of the climate evolution of the past millennium, as inferred from different models and studies based on proxy-based reconstructions or recent instrumental data.

Chapter 3 focuses on the comparison of the off-line and on-line ensemble-based DA methods. Ten ensemble members are run for each DA scheme. The test case corresponds to the climate of the period leading into the Maunder Minimum, i.e. 1600-1700 AD. The chapter provides a validation of the results of the two

approaches, discusses their limitations, and includes a significance test of the outputs. The study is based on the Matsikaris et al. (2015a) paper. The on-line approach is selected as the most suitable one, and is employed in the following two chapters. The focus of Chapter 4 is on the skill of the reconstruction using on-line DA and the added value gained compared with simulations without DA. The climate of the late pre-industrial period, 1750-1850 AD, is being simulated, employing 20 on-line ensemble members. The performance of the scheme is evaluated on large and small spatial scales. A maximum covariance analysis (MCA) of links between the NH temperature and sea level pressure (SLP) is also performed, to further investigate the potential of skill in the applied DA scheme in terms of small spatial scales. The chapter is based on the Matsikaris et al. (2015b) paper.

As the findings of Chapter 4 reveal a lack of information propagation to small spatial scales, Chapter 5 investigates whether the unsuccessful representation of regional temperature variability owes to the methodology or to errors in the assimilated data. Two different sets of ensemble simulations are conducted, this time assimilating gridded instrumental observations apart from proxy-based reconstructions. The period under investigation is 1850-1950 AD. Moreover, the study serves as a validation of the Matsikaris et al. (2015b) DA scheme against gridded instrumental observations. The validation of DA during the instrumental period can lead to substantial new results that are crucial for a well-rounded assessment of the method. The chapter is based on the Matsikaris et al. (2016) paper. Finally, a summary of the main conclusions of the thesis is provided in Chapter 6. This is followed by an identification of the main limitations of this research, and a discussion of the perspectives for future work.

Chapter 2

Climate evolution over the past millennium

2.1 Introduction

Earth's climate has always been varying on a wide range of timescales, either driven by external factors, or due to internal factors of the climate system. The Pleistocene climate epoch, which lasted from about 2.5 million to 12 thousand years ago, was marked by several major glacial and interglacial periods, characterized by oscillations in global ice volume on timescales of tens to hundreds of millennia. The latest glacial maximum was noted 21,000 years ago and the most recent glacial cycle ended abruptly 12,000 years ago (Jones and Mann, 2004). The successive interglacial period, which continues to the present, was marked by less impressive but still considerable variability on centennial and millennial timescales. This period is known as the "Holocene". During the past one to two millennia ("late Holocene"), the Earth's orbital geometry and global ice mass have not changed significantly, and thus the climate variations of this period can be expected to be representative of the natural climate variability expected today in the absence of anthropogenic

influences (Jones and Mann, 2004).

Climate variability on multimillennial timescales is dominated by external forcing, and particularly by changes in the insolation distribution at the Earth's surface, which is associated with long-term changes in the orbital position of the Earth relative to the sun (Jones and Mann, 2004). On timescales of the past one to two millennia, the forcings that dominate climate variability consist of external forcing due to natural changes in solar irradiance and volcanoes, as well as anthropogenic influences that arise from land use changes, greenhouse gases and sulphate aerosols. Like the climate itself, forcings need to be reconstructed from proxy information sources, and are therefore associated with substantial uncertainties. Solar and volcanic forcings are considered to be dominant amongst the drivers of the climate change in the last two millennia; however, the unprecedented warming of the late 20th century can only be explained by the anthropogenic impact.

The past millennium can advance our understanding of mechanisms that drive climate variability, as the abundance of high quality proxies allows reconstructions on all spatial scales (e.g. Jansen et al., 2007). Simulations and proxy-based reconstructions for the climate of the last 1,000 years reveal a long-term millennial cooling trend, followed by sustained recent warming in the 20th century. Two important periods are usually identified in many studies, the Medieval Warm Period (MWP), extending from 950 to 1200 AD, and the Little Ice Age (LIA), from 1450 to 1850 AD. However, many studies consider this picture to be too simplistic, describing a more complex pattern for regional temperature variations that is inconsistent with the hemispheric or global mean pattern (e.g. Williams and Wigley, 1983; Jansen et al., 2007).

During the MWP, our planet experienced warming similar to the recent period in some regions, but considerably less so on the global scale. Evidence shows that La Niña-like conditions prevailed in the tropical Pacific during this period,

with enhanced warmth in North America and the Eurasian Arctic, and cooling in central Eurasia (Mann et al., 2009). El Niño-like conditions prevailed later on, for example during the 17th century (Jones and Mann, 2004). The peak cooling during the LIA was noted over the extra-tropical NH continents. In the Past Global Changes (PAGES) 2K reconstruction, which gives a continental-scale description of the temperature evolution for the last two millennia (PAGES 2K Consortium, 2013), no adequate indications for well-defined global MWP and LIA exist. Centennial-scale temperature anomalies are heterogeneous in both time and space, and the timing of the peaks of the warm and cold periods varies regionally, with multi-decadal variability.

Robust conclusions for the past climate variability in the SH and parts of the tropics cannot be easily drawn, due to the sparsity of high-resolution proxy data in those regions, therefore most of the studies about the climate of the past millennium have focused on the NH. Recent studies that investigated SH and inter-hemispheric climate variability (e.g. Neukom et al., 2014) have reported substantial differences in the climate of the two hemispheres. For instance, no globally coherent warm phase was observed during the pre-industrial period (before 1850 AD), while a globally extended cold period was only observed between 1594 and 1677 AD (Neukom et al., 2014; Bothe et al., 2015). The only period in the past 1,000 years during which both hemispheres experienced simultaneous warming appears to be the late 20th century.

The global warming in the 20th century was approximately 0.6°C to 0.9°C according to most proxy-based reconstructions, whereas temperature changes in any other century of the past two millennia did not exceed 0.2°C per century (Jones and Mann, 2004). This century was also affected by the largest changes in radiative forcing, with the greenhouse gas concentrations playing the dominant role in the observed warming. According to most studies, the late 20th century warming has

been unprecedented at least during the past 1,800 years, especially for the NH. It is also possible that the warming of this century occurred at an average rate about 10 times faster than the overall warming of 4°C to 7°C since the last glacial maximum (Jansen et al., 2007). Proxy-based studies also indicate that internal modes of climate variability, such as the El Niño Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO), may have exhibited anomalous behaviours in the last part of the 20th century (Jansen et al., 2007).

This chapter provides a general description of the temperature variability of the past millennium, combined with a model documentation study for the evaluation of the model employed in the thesis, through comparisons with observations, reconstructions and other GCMs. Model documentation, although an essential source of knowledge dissemination, is often neglected in model development, as model applications and improvement are usually considered to be more important. Verifying through documentation that the model produces realistic simulations is essential in the present thesis, otherwise the ensemble-based DA technique performed in the following chapters will be worthless.

The structure of the chapter is as follows: Section 2.2 provides the general characteristics of the model used throughout the thesis. An evaluation of the model output against data and simulations for the last millennium on large spatial scales is presented in Section 2.3. Next, Section 2.4 examines the inter-regional patterns of temperature variability in different models and reconstructions, to investigate whether the simulated links between continental regions are similar to those identified in other simulations and proxy-based reconstructions. Section 2.5 presents further contributions to the understanding of climate variability of the past millennium from recent DA studies. Finally, a summary of the main findings is provided in Section 2.6.

2.2 Model overview

The model employed throughout the different parts of this thesis is the Max Planck Institute for Meteorology Earth System Model (MPI-ESM), which couples the atmosphere, ocean and land surface through the exchange of energy, momentum, water and important trace gases such as carbon dioxide. The configuration used here, MPI-ESM-CR, is a coarse-resolution version of the model used for the Coupled Model Intercomparison Project, Phase 5 (CMIP5) simulations. The MPI-ESM comprises of the general circulation models ECHAM6 (Stevens et al., 2013) for the atmosphere and MPIOM (Marsland et al., 2003) for the ocean/sea-ice. ECHAM6 was run at T31 horizontal resolution ($3.75° \times 3.75°$), 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 is used to couple the ocean and the atmosphere daily without flux corrections. The land surface model is JSBACH (Raddatz et al., 2007). No ocean biogeochemistry model was employed, and the dynamic vegetation module was switched off, as in the model version described by Jungclaus et al. (2010).

The simulation performed here covers the last millennium (850-2005 AD) and follows the "past1000" protocol of the Paleoclimate Modelling Intercomparison Project, Phase 3 (PMIP3). As part of this protocol, different choices for external forcing and boundary conditions are summarized by Schmidt et al. (2011). Tables for the orbital parameters and the well-mixed greenhouse gases (CO_2, CH_4, N_2O) are also provided. According to the PMIP3 protocol, the "past1000" integration period is the period between 850 and 1849 AD. The "past1000" simulation was continued over the "historical" period, 1850-2005 AD, following the CMIP5 protocol, except for the use of the Pongratz et al. (2008) land-cover data throughout the 20th century. The "past1000" simulation was started after a 700 year-long

spin-up simulation with constant 850 AD boundary conditions. The multi-century spin-up simulation was run until the time when the drift in the ocean temperature at a depth of 2,000 m was negligible. The "past1000" simulation was also the basis of the ensemble simulations employed in the following chapters.

The simulation is forced by reconstructions of natural forcing (total solar irradiance and volcanic forcing) and anthropogenic forcing (land cover changes, greenhouse gas emissions and aerosol forcing). For solar radiation, the reconstructed variations of total solar irradiance over the Holocene by Vieira et al. (2011) are employed. The volcanic effect on radiation is considered by applying the Crowley and Unterman (2012) reconstruction for volcanic aerosols. The orbital forcing is prescribed by a table providing annual values for eccentricity, obliquity and perihelion. In contrast to the millennium simulations described by Jungclaus et al. (2010), which featured an interactive carbon-cycle, prescribed CO_2 and concentrations of the most important greenhouse gases (Schmidt et al., 2011) are used. Prescribed external forcings also include the Pongratz et al. (2008) reconstruction of global land-cover and agricultural areas. Monthly average ozone concentrations for the period 850-1849 AD were calculated using the 1850-1860 AD monthly climatology of ozone concentrations from the AC&C/SPARC Ozone Database as a basis and representing the ozone dependency on solar irradiance through regression coefficients between historical ozone concentrations and the annual 180.5 nm solar flux (Jungclaus et al., 2014).

2.3 Large-scale temperature variability

In this section, some of the main characteristics of the climate of the last millennium are described, while the "past1000" and "historical" simulations of the model used in this thesis (MPI-ESM-CR) are evaluated against instrumental records, proxy-

based reconstructions and other GCM simulations, on hemispheric and continental-scale temperature variability. The comparisons serve as a documentation of the model, for testing whether it can reproduce important features of reconstructed climate indicators in the 20th century and the last millennium, and investigating the consistency with other GCMs. This documentation is important, as it is the first time that the specific version of the model is used. The validation of the model presented in this section looks into the temperature response of the model to forcings. DA attempts to capture both internal variability and the response to forcings, therefore the validation is necessary, as if the model's response to forcings is unrealistic, DA employed in the following chapters will simply produce erroneous results.

2.3.1   Inferences from proxy-based reconstructions

Proxy-based temperature reconstructions of the NH for the last millennium differ significantly among each other, while large discrepancies are also found when they are compared with simulations (e.g. Bothe et al., 2015). Climate periods such as the MWP and the LIA are therefore often under debate regarding their spatial and temporal extent. The evolution of the NH simulated MPI-ESM-CR temperature over the last 1,200 years is compared to the range of reconstructions, redrawn from Jansen et al. (2007), in Figure 2.1. The reconstruction data used are those featured in Fig. 6.10 of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Jansen et al., 2007). The simulated temperatures fit well within the range of the reconstructions, showing that the model responds in the right way to the forcings, w.r.t. the NH mean temperature. Nevertheless, this cannot be a strict validation, as the grey band is rather wide.

The main feature of the NH climate, in both model and reconstructions, is a temperature decline after around 1200 AD, which lasts until the early 19th

NH surface air temperature
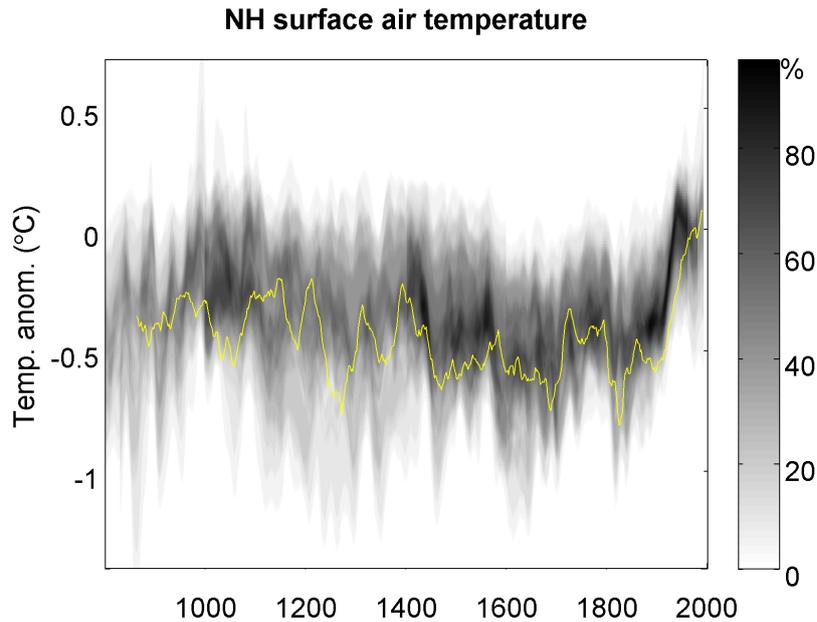
Figure 2.1: NH 2 m temperature anomalies w.r.t. the 1961-1990 AD mean for the MPI-ESM-CR model (yellow line) compared with the range of reconstructions (gray scale), redrawn from Jansen et al. (2007). The time series are smoothed by a 31-year running mean. Data time series and the shaded representation of overlap (consensus) were obtained from: http://www.cru.uea.ac.uk/datapages/ipccar4.htm.

century, followed by intense warming in the past two centuries. The influence of volcanic eruptions is also evident throughout the millennium. Most proxy-based NH temperature reconstructions (e.g. Jansen et al., 2007; Mann et al., 2008, 2009) indicate the centre of the MWP around the turn of the millennium. Their temperatures start to decrease after 1150 AD. According to the MPI-ESM-CR simulation, the warmest pre-industrial temperatures prevailed during the 10th, 12th and late 18th centuries. The MWP is less pronounced compared with the proxy-based reconstructions, in which the temperature maximum occurs during the 11th century. Different hypotheses have been expressed in order to explain the occurrence of the warm MWP conditions over Europe and North America. For example, according to Trouet et al. (2009), a positive NAO/Arctic Oscillation (AO) anomaly, as a dynamical response to the solar forcing anomaly, occurred continuously over several centuries and caused this rise in the NH temperatures.

The coldest anomalies in the simulation are identified during the 13th, 15th, 17th and early 19th centuries and agree relatively well with the consensus of reconstructions. The proxy-based reconstructions indicate the coldest part of the LIA to be during the first part of the 17th century (Frank et al., 2010), at about the same period as the MPI-ESM-CR simulation. Severe volcanic eruptions, such as the ones noted in 1258 AD, 1453 AD and 1815 AD, cause a sharp drop in the mean temperature during the years that follow, as well as a long-lasting effect on NH climate, recorded in both proxies and the simulation. Detection and attribution studies showed that the main driver of the anomalously cold conditions of the LIA is the volcanic forcing (e.g. Hegerl et al., 2007). Weak evidence suggests an additional contribution from a small decrease in CO_2 concentration (e.g. Schurer et al., 2014).

2.3.2   Comparison with instrumental data

The instrumental period is ideal for testing the ability of the model to reproduce important climate features, due to the high fidelity of the available observational record. Figure 2.2 compares the evolution of the simulated 20th century NH (land and ocean) temperatures with the HadCRUT3v instrumental dataset (Brohan et al., 2006), obtained from the Climatic Research Unit, http://www.cru.uea.ac.uk/cru/data/temperature. The two time series are in very good agreement (correlation is 0.83) and exhibit a warming trend between 0.6°C and 0.7°C over this period. Distinct variability on the multi-decadal timescale is superimposed over the main warming trend. The simulation arrives at slightly colder temperature towards the end of the period compared with the observed dataset. The simulated time series includes temperature minima in the 1880s and 1900s and a maximum in the 1960s.

The warming trend of the 20th century stands out against all pre-industrial trends
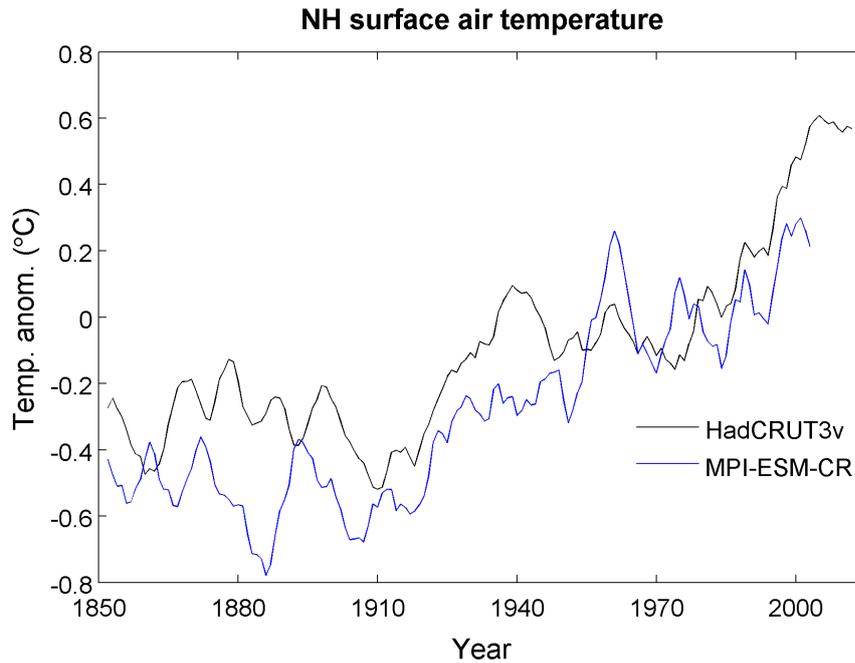
Figure 2.2: 20th century NH (land and ocean) 2 m air temperature anomalies w.r.t. 1961-90 AD, smoothed by a 5-year running mean, as simulated by the MPI-ESM-CR model, compared to the HadCRUT3v dataset (Brohan et al., 2006), obtained from the Climatic Research Unit, http://www.cru.uea.ac.uk/cru/data/temperature.

in both observed and simulated temperatures of the last millennium. In most studies, the 20th century was ranked as the warmest or nearly the warmest century everywhere in the world except Antarctica (e.g. PAGES 2K Consortium, 2013). The large thermal inertia of the ocean that surrounds this continent may be the explanation for this discrepancy, as it dampens the warming. In the PAGES 2K continental reconstructions, the warming in the regions of the NH was about twice that of Australasia and South America (about 0.5°C compared with 0.2°C), owing again to the strong ocean dominance in the SH (PAGES 2K Consortium, 2013). The largest warming was noted in the high latitudes, with the Arctic being warmed by 0.9°C.

Reasonably good agreement between the MPI-ESM-CR simulation and the instrumental data is also found when comparing the sea surface temperature evolution over the 20th century. Figure 2.3 illustrates the simulated 20th century NH sea

surface temperatures (anomalies w.r.t. the 1961-90 AD mean, smoothed by a 5-year running mean), in comparison with the HadSST dataset (obtained from the Climatic Research Unit). Model and observations follow similar trends, with the correlation between the two time series being again high (0.78). The warming trend of the 20th century is evident in both temperature series, and similarly to Figure 2.2, the simulation arrives at slightly lower sea surface temperatures towards the end of the simulation period than the observational record.
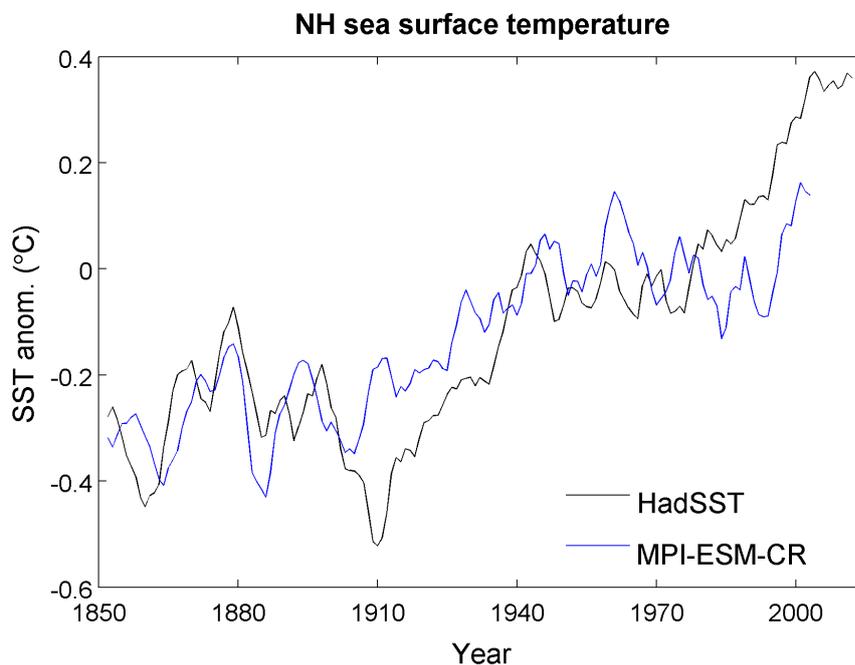
Figure 2.3: 20th century NH sea surface temperature anomalies w.r.t. 1961-90 AD, smoothed by a 5-year running mean, as simulated by the MPI-ESM-CR model, in comparison with the HadSST dataset (obtained from the Climatic Research Unit, http://www.cru.uea.ac.uk/cru/data/temperature).

2.3.3 CMIP5 and PMIP3 simulations

CMIP5 comprises an enhanced set of historical and palaeoclimate simulations and offers vast opportunities for detailed model evaluation (Taylor et al., 2012). These simulations were performed by more than 20 modelling groups, and more than 50 models were employed, aiming to address different problems and

scientific questions within various communities. Two main types of climate change experiments were performed; long-term integrations, with centennial timescales, and near-term integrations, with timescales between 10 and 30 years (decadal prediction experiments). Compared with earlier phases, such as the PMIP (http://pmip3.lsce.ipsl.fr) or the CMIP3, CMIP5 includes more comprehensive models and a wider set of experiments. The models have higher spatial-resolution and a richer set of output fields is archived. The spatial resolution of CMIP5 coupled models is between $0.5°$ and $4°$ for the atmosphere and $0.2°$ to $2°$ for the ocean (Taylor et al., 2012). The simulations cover the period 850-2000 AD, while some models have performed future projections as well. A full suite of forcings was employed, i.e. solar, volcanic, orbital, greenhouse gases and land-use changes forcings. In this chapter, the MPI-ESM-CR simulation is compared with the following CMIP5 and PMIP3 models:

- CESM: Community Earth System Model, developed at NCAR (National Center for Atmospheric Research), USA.

- CCSM4: Community Climate System Model, developed at NCAR, USA.

- COSMOSe1 (low solar plus other forcings), COSMOSe2 (high solar plus other forcings) and COSMOS: Max Planck Institute Millennium Activity pre-PMIP3 simulations, Germany.

- CSIRO-Mk3L-1-2: Commonwealth Scientific and Industrial Research Organisation model, University of New South Wales (UNSW), Australia.

- GISS-E2-R: NASA Goddard Institute for Space Studies model, developed at the Goddard Institute.

- HadCM3: Hadley Centre Coupled Model, version 3, Hadley Centre, UK.

- IPSL-CM5A-LR: Institut Pierre Simon Laplace model, France.

- MPI-ESM-P: Max Planck Institute PMIP3 simulations (higher resolution version of the MPI-ESM model used in the present thesis), Germany.

The evolution of the simulated temperature for the NH mean over the last

millennium using the MPI-ESM-CR model is compared with the temperatures simulated by the MPI-ESM-P, CCSM4, IPSL-CM5A and HadCM3 models, in Figure 2.4. The time series are anomalies w.r.t. the 1961-1990 AD mean, smoothed by a 31-year running mean. The correlations obtained are relatively high (0.76, 0.69, 0.60 and 0.73 respectively). Figure 2.5 compares the 20th century temperature anomalies for the MPI-ESM-CR simulation, the four GCMs, as well as the instrumental data (5-year running means). The temperatures simulated by the MPI-ESM-CR model are again in good agreement with the other simulations and the observations. All models follow similar trends.
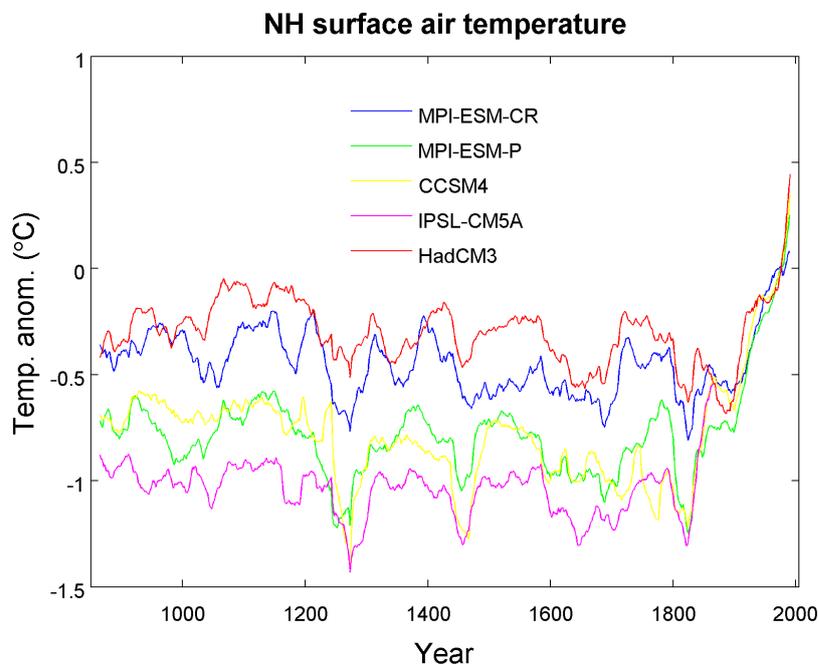
Figure 2.4: Simulated NH 2 m temperature over the last millennium (anomalies w.r.t. the 1961-1990 AD mean), smoothed by a 31-year running mean. Comparison of the MPI-ESM-CR simulation with the MPI-ESM-P, CCSM4, IPSL-CM5A and HadCM3 simulations.

A prominent common feature in all simulations is the pronounced negative temperature anomalies in the decades of large volcanic eruptions. The tropical eruptions in the 1250s, 1450s and 1810s are the most remarkable cases. The MPI and CCSM4 simulations register the largest response to the eruptions among the simulations. Proxy-based reconstructions reflect these eruptions to a smaller extent.
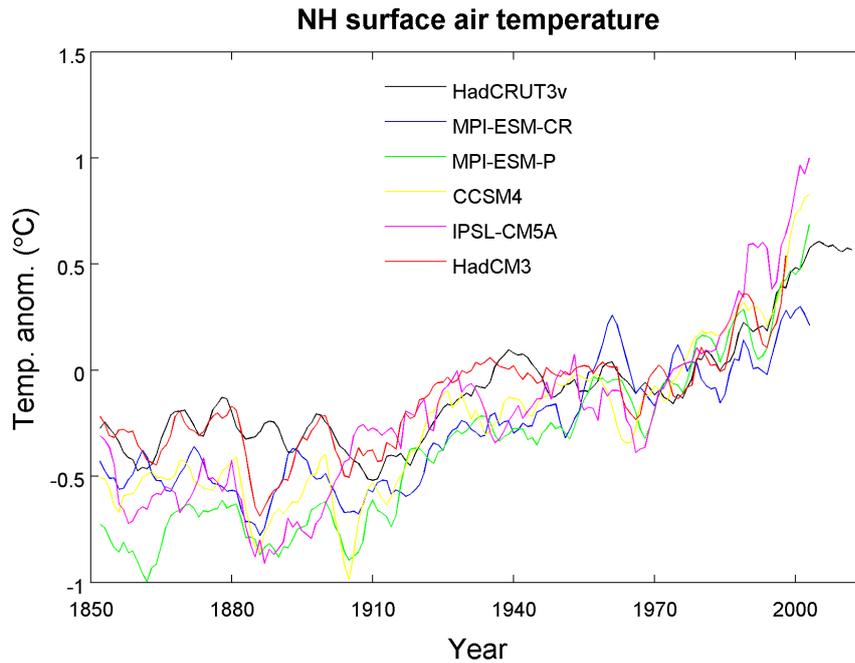
Figure 2.5: NH 2 m land and ocean temperature anomalies in the 20th century w.r.t. the 1961-1990 AD mean for the MPI-ESM-CR model, four different GCMs and the instrumental data (5-year running means).

Only the 1815 AD Tambora eruption is clearly registered in most reconstructions, but still, the cooling in most simulations is larger than that of the early instrumental observations or reconstructions (Brohan et al., 2012). These discrepancies are likely due to uncertainties in the forcing reconstructions employed by the models. Furthermore, some simulations, such as the IPSL-CM5A and CCSM4, exhibit a stronger recent warming than the observed data.

High-frequency variability is different in each simulation, owing mainly to the internal climate variability, and in addition, to the fact that the models have different resolutions and use different representations of the system dynamics. Every simulation represents only one of the many possible scenarios of the internal variability evolution of the climate system. Most probably, the latter is not the evolution followed in the real world. The choice of initial conditions in each model affects the variability. The large contribution from internal variability is evident in simulations of different ensemble members that use the exact same physical model

and forcing setup, but slightly different initial state (also known as "initial value ensembles"). The ensembles give a larger range of independent evolutions of internal variability. As an example, Figure 2.6 compares the MPI-ESM-CR simulation with three realisations of the higher resolution version of the MPI-ESM model (MPI-ESM-P). Naturally, there is a wider range of temperature evolutions simulated by the MPI-ESM-P ensemble at a given time, compared with the single simulation of the MPI-ESM-CR model. In ensemble simulations, the ensemble mean can isolate the forced response of the climate, due to the filtering of internal variability after the averaging procedure (Jungclaus et al., 2010). The figure also shows a more pronounced overall warming for the coarse resolution model compared with the members of the higher resolution one. This difference might be partly explained by a stronger sensitivity to volcanoes that has been noticed in the MPI-ESM-P version, particularly for the first realisation of the ensemble.
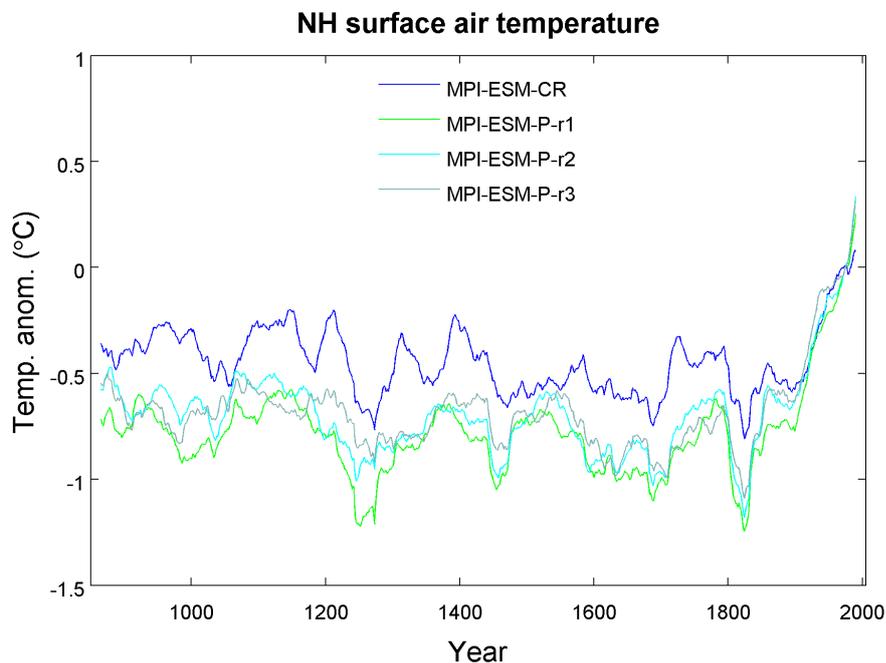
Figure 2.6: Comparison of the MPI-ESM-CR 2 m air temperatures with a three-member ensemble of the higher resolution MPI-ESM-P model (anomalies w.r.t. the 1961-1990 AD mean) over the last millennium.

Finally, the consistency between the different models is investigated, w.r.t. the

temperature anomalies on the continental scale. Figure 2.7 compares the correlations of the continental mean temperatures between the model used in this thesis and other GCMs during the past millennium. For the correlation calculations, the definitions of the continental regions were in accordance with the PAGES 2K supplement (PAGES 2K Consortium, 2013), given in Table 2.1, motivated by the use of the PAGES 2K reconstructions in the assimilation employed in the subsequent chapters. The ocean was masked for the simulated land-only PAGES 2K equivalent regions, namely Antarctica, Asia, Europe and South America, using binary land masks. The longest common periods between the forced simulations were used for the calculation of the correlations.

Continent	Season	Latitude range	Longitude range
Antarctica	annual	90S - 60S	180W - 180E
Arctic	annual	60N - 90N	180W - 180E
Asia	June - August	23.5N - 55N	60E - 160E
Australasia	September - February	50S - 0S	110E - 180E
Europe	June - August	35N - 70N	10W - 40E
North America	annual	30N - 55N	130W - 75W
South America	December - February	65S - 20S	75W - 30W

Table 2.1: Seasonal and spatial definition of continents in the PAGES 2K reconstructions.

The correlations of the continental mean temperatures between the models are not very high for most of the continents. Higher correlations are found for the Arctic, Asia, Australasia and South America (around 0.4), while Europe and Antarctica appear to be very differently simulated by the MPI-ESM-CR model and the other GCMs. Differences between the models can be attributed to the different random internal variability, which plays a more important role on smaller spatial scales, such as the continental regions, in comparison with the hemispheric or global scales, as well as the different response to the forcings in each model. It is important to note that, owing to random internal variability, simulations and reconstructions for the continental scale often disagree substantially in many parts of the millennium,
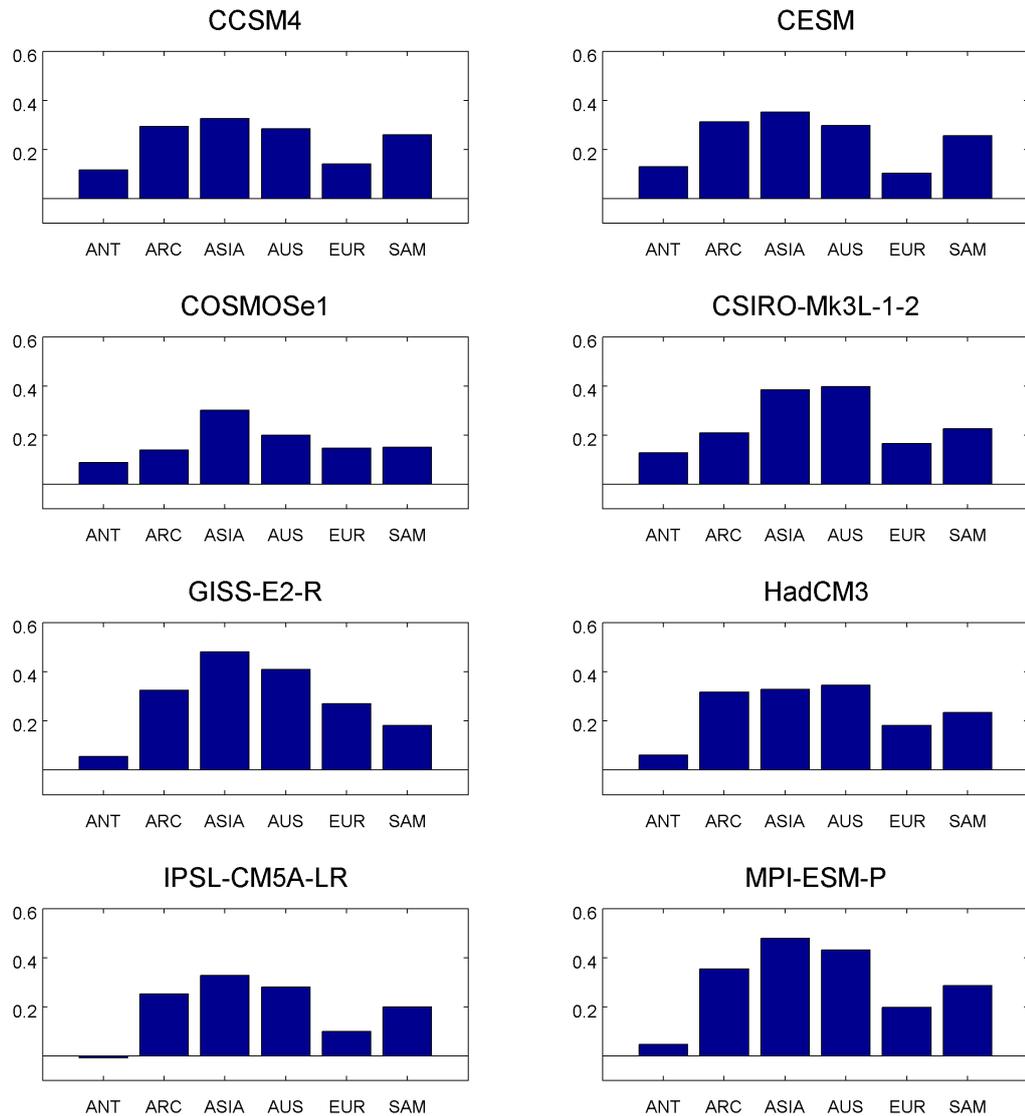
Figure 2.7: Correlations of the continental mean temperatures between MPI-ESM-CR model and other GCMs (for the PAGES 2K regions) for the last millennium.

in terms of both the timing and the amplitude of climate variability. Simulations thus often fall outside the uncertainty range of continental reconstructions (e.g. PAGES 2K Consortium, 2013), while agreement is generally higher at hemispheric and global scales.

2.4 Inter-regional patterns of spatial variability

While the consistency between models and reconstructions on hemispheric scale temperatures has been examined in several studies, the inter-regional patterns of temperature variability in reconstructions and in different simulations have not been thoroughly investigated. These were addressed by Bothe et al. (2015), who compared the cross regional relations of temperature variations in the PAGES 2K reconstructions and the CMIP5 and PMIP3 simulations. In this section, the above study is extended by using the low resolution version of the MPI-ESM model (MPI-ESM-CR), which was not part of that analysis. The definitions of the continental regions are the same as in the previous section (Table 2.1) and the longest common periods for forced simulations and reconstructions are used for each comparison.

The spatial variability structure of the climate is affected by both forced signals and teleconnections in the internal climate variability. Bothe et al. (2015) found large discrepancies in the covariability patterns of temperature variations in reconstructions and models. The simulations were found to be much more regionally coherent than the reconstructions, in which the correlations between different continents were rather low. In principle, these differences can be explained by several potential reasons: (i) the use of too strong forcings in the models or too strong response to realistic forcings, which leads to much larger forced climate variability in the models than in the real world; (ii) unrealistic internal variability in the models (some regions are, for example, less connected by modes of variability in reality compared with the link suggested by the models); and (iii) the fact that the PAGES 2K reconstructions might be dominated by non-climatic noise. Correlations in the reconstructions are expected to decrease due to any non-climatic

noise affecting the proxy records or errors in the methods employed to reconstruct each region (Bothe et al., 2015) .

Here, as a first step, the correlations between the mean continental temperatures of the PAGES 2K regions for eight different GCM simulations are calculated and presented in Figure 2.8. Either positive or negative inter-regional correlations significantly different from zero can appear due to teleconnections in internal variability, or due to a common response to the forcings. Data from North America were not included in this study, as the proxy-based reconstruction for this region was based on decadal and 30-year averages. Inter-regional correlation ranges vary considerably in all forced simulations but are in general high, with some models, such as the CESM, CCSM4 and IPSL, producing correlations above 0.7. Some regional characteristics are simulated differently to a large extent; for example, the correlations between Antarctic temperatures and the other continents, which are very high in CESM and CCSM4, but much lower in HadCM3 and COSMOSe1. Such differences owe to the different ratio of forced and unforced variability, and the different response to external forcing in the different models.

Figure 2.9 displays the temperature correlations among the continental regions in the PAGES 2K reconstructions. The reconstructions, which represent the real world, and thus a forced climate, attain very low inter-regional correlations. The highest correlations are found for the NH continents, but as in Bothe et al. (2015), these are still substantially inconsistent with the much more homogeneous structure found in the models. A notable example is the correlation between Australia and South America, which is very high in most models but low in the reconstructions. We must however note that the evaluation of the spatial correlations in this study does not take into account the reconstruction uncertainties, which can weaken the links between the continents.

The same analysis of links between regions was performed employing the forced
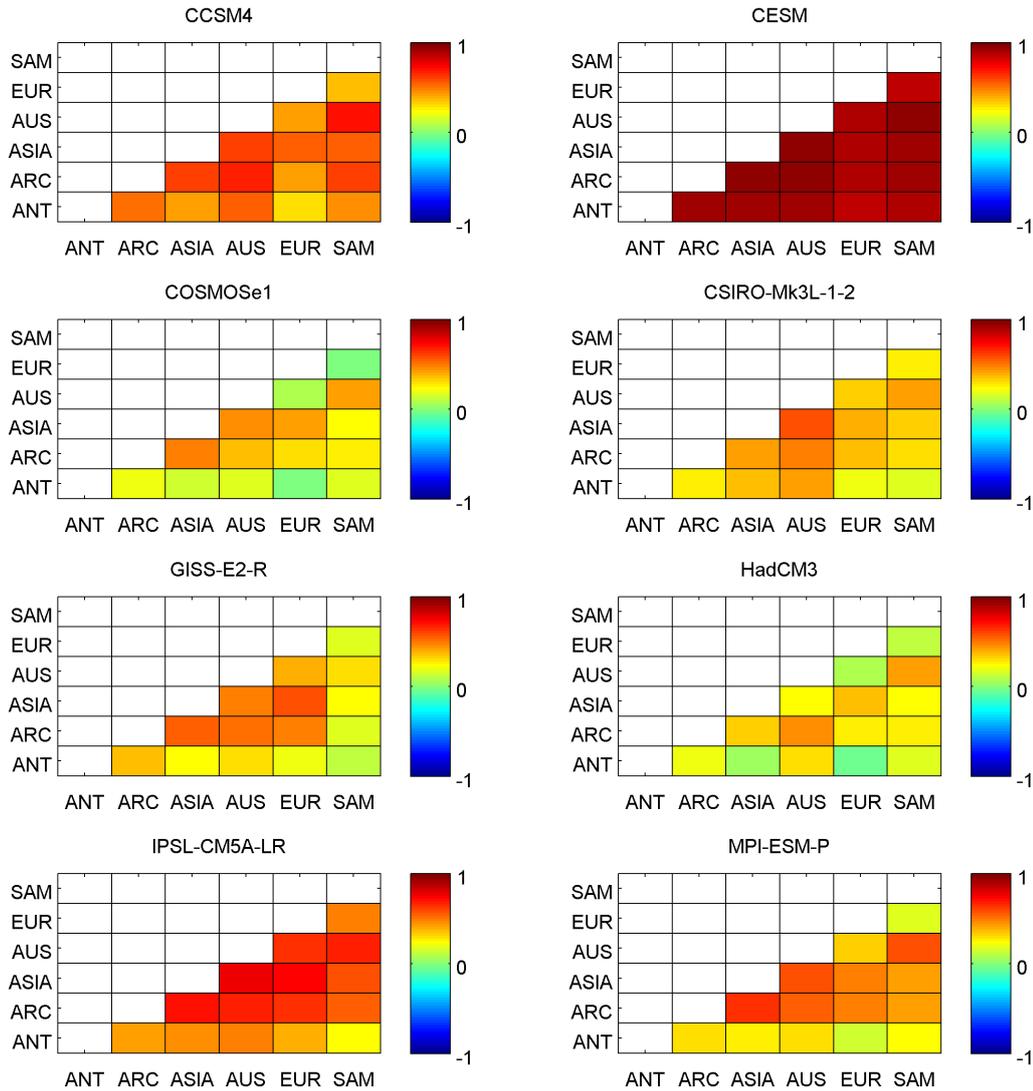
Figure 2.8: Inter-regional correlations among the PAGES 2K regions in different CMIP5/PMIP3 simulations for the last millennium.

simulation of the MPI-ESM-CR model for the entire analysis period (850 AD to present) and the pre-industrial period (before 1850 AD), as well as using the unforced control simulation. This aimed at disentangling the contributions from teleconnections in the internal variability and from the response to the forcings. The inter-regional correlations among the PAGES 2K regions over the past millennium are displayed in Figure 2.10a, and compare well with the range of correlations of the other CMIP5 and PMIP3 models (Figure 2.8). The correlations are the highest between the Arctic and Europe, Arctic and Asia, and Europe and Asia. This is
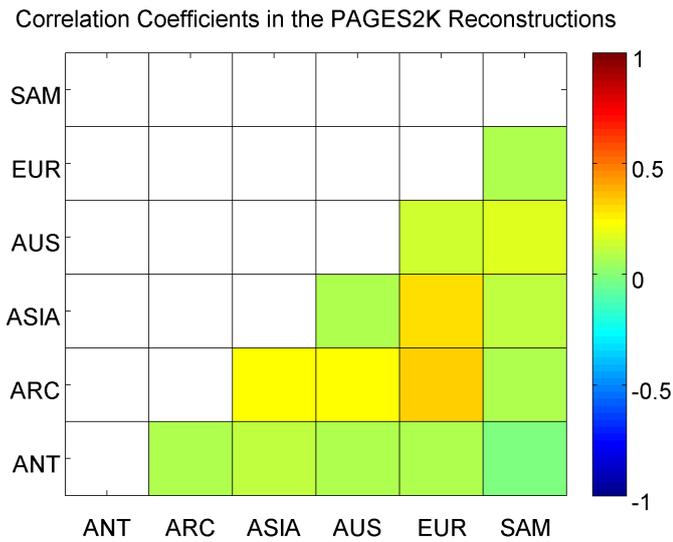
Figure 2.9: Inter-regional correlations among the PAGES 2K regions in the PAGES 2K reconstuctions for the last millennium.

likely to be due to the larger contribution of forced variability in the NH, and particularly the high latitudes.
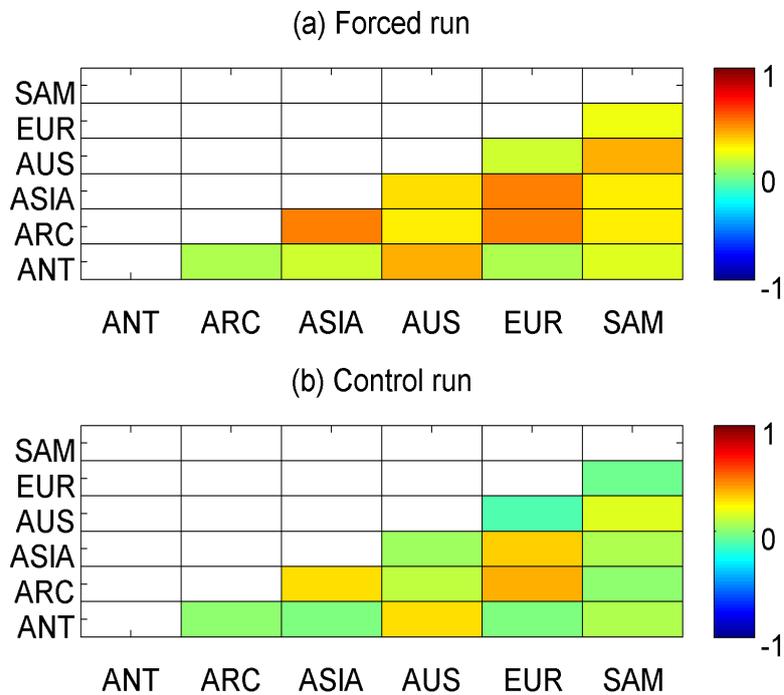
Figure 2.10: Inter-regional correlations among the PAGES 2K regions in the MPI-ESM-CR simulation for the last millennium. (a) Fully-forced simulation. (b) Control simulation.

The correlations for the unforced control simulation (Figure 2.10b) are found to be close to zero, indicating that teleconnections between continents are weak and do not contribute to the overall correlations. The correlations in the case when only the pre-industrial period is considered (not shown) are also high, but lower than the ones for the entire millennium. These results show that the strong inter-regional links in the model mostly owe to the homogeneous and strong anthropogenic warming, while the response to volcanic, solar, orbital and land use forcings also contributes to the inter-regional correlations. Based on the results, it can be argued that the discrepancies in the covariability patterns between the MPI-ESM-CR model and the PAGES 2K reconstructions owe to either the use of too strong forcings in the model (or a too strong response to the forcings), or to the existence of non-climatic noise in the reconstructions.

2.5 Contribution from DA studies

As stated in Chapter 1, the relatively large number of proxy-based reconstructions and simulations covering the past millennium allows for an investigation of the consistency between them. Model outputs and data were often found to agree reasonably on the hemispheric and global spatial scales, but the inconsistencies observed during some periods and especially on the local scale are large (Bothe et al., 2015). Assimilation of proxy-based reconstructions into models provides reconstructions that are complementary to those provided by the two methods. Moreover, the representation of atmospheric and oceanic circulations in models allows DA to provide insights into mechanisms that could be responsible for past climate changes (Goosse et al., 2006). This is however a relatively new field, and few DA studies have so far been performed. Among them, most studies have focused on climate variability of the past millennium, while others (Mairesse et al., 2013; Klein et al., 2014) examined the climate of the mid-Holocene (6000 BP).

The dominance of forcings on the large scales limits the benefits that DA can offer for large-scale temperature reconstructions, as climate models can perform well on these without the use of DA. Added value can however be given on smaller scales and process understanding. For instance, even on decadal timescales, DA can provide climate states that are closer to reality, as it can capture the contribution of internal variability. In addition, DA can shed light on climate reconstructions for periods with large uncertainties associated with the models or the proxy data. Reconstructions based on DA were found to improve the agreement with proxy-based reconstructions and instrumental data at global and hemispheric scales for the last millennium (e.g. Goosse et al., 2006). Reasonable consistency was also found on the continental scale at the study of Goosse et al. (2010), apart from the Arctic and North Atlantic regions.

In terms of the benefits provided by DA studies regarding process understanding of the mid-Holocene, Mairesse et al. (2013) found a strengthening of the westerly winds at mid-latitudes that led to a warming of northern Europe. Klein et al. (2014) identified a reduction of the southward winds in the Barents Sea and an increase in the westerly winds in the Canadian Basin as potential drivers of the reconstructed changes in Arctic sea ice concentration.

In studies focusing on the past millennium, Crespin et al. (2009) employed DA to examine the substantial surface temperature variations that occurred in the Arctic. Even though the last part of the 20th century was probably the warmest period in this region, strong warming episodes have also occurred in the past. The 1470-1520 AD Arctic warming was largely attributed to internal climate variability. In principle, the warm anomaly was likely produced by changes in atmospheric circulation, through enhanced south-westerly winds towards Siberia, northern Europe and Canada (Crespin et al., 2009). The extratropical NH climate variability was studied in Widmann et al. (2010). Three different DA methods

(ensemble member selection, forcing singular vectors and pattern nudging) inferred that the cold 1790-1820 AD event in Scandinavia was associated with anomalous northerly or easterly atmospheric flow. This flow was likely linked to a pressure pattern resembling a negative state of the Northern Annular Mode (NAM). Finally, Goosse et al. (2012) investigated the origins of the MWP and associated the spatial pattern of the warming noticed with relatively weak changes in radiative forcing and a modification of the atmospheric circulation.

2.6 Summary and conclusions

In this chapter, the temperature variability over the last millennium was examined using the MPI-ESM-CR model, as well as instrumental data, proxy-based reconstructions and other GCMs. The MPI-ESM-CR simulation showed good performance for multi-decadal temperature variability on hemispheric and global spatial scales when evaluated against the other sources of information. The simulation presented similar characteristics to those of the other model simulations and was found to fall in the range of the proxy-based reconstructions. In general, many of the features of the observed and proxy-based records were compatible with the MPI-ESM-CR simulation, and highlight the peculiarity of the present climate change. These include a gradual temperature decrease between the 13th and early 19th centuries followed by unprecedented warming, and severe cooling noted during the years that follow major volcanic eruptions. The consistency of the model results with instrumental data is also high, simulating a warming trend between 0.6°C and 0.7°C over the 20th century, superimposed by distinct variability on the multi-decadal timescale.

The coolest periods in the proxy-based reconstructions appear to be the 15th, 17th, and 19th centuries (Jones and Mann, 2004), in agreement with the MPI-ESM-CR

simulation, while the coldest part of the LIA is recorded during the early 17th century in both the MPI-ESM-CR simulation and the reconstructions. Yet, some other aspects of the climate evolution recorded by the proxies, such as the timing and magnitude of the MWP, do not agree well with the simulated ones. The last outcome was also noted in many of the last millennium simulations. Reconstructions and CMIP5/PMIP3 simulations are in reasonably good agreement in terms of the NH mean temperature. The pronounced cooling after large tropical volcanic eruptions is evident in all simulations, with a larger response recorded in simulations than in proxy-based reconstructions. In addition, the IPSL-CM5A and CCSM4 simulations show a stronger recent warming than that recorded by the instrumental data. Correlations of hemispheric-scale temperatures between the MPI-ESM-CR model and other GCMs are relatively high, but consistency on the continental scale and on high frequency variability is much lower, due to the dominance of internal variability on the smaller spatial scales.

Finally, in terms of the inter-regional patterns of spatial variability, the correlations between continents of the MPI-ESM-CR model over the past millennium are in general high, in agreement with the correlation ranges found in all forced simulations. On the other hand, the PAGES 2K reconstructions demonstrate very low inter-regional correlations. Higher correlations were found for the NH continents, but these were still much lower than in the MPI-ESM-CR or the CMIP5/PMIP3 simulations. The correlations were also calculated for the unforced control simulation of the MPI-ESM-CR model and were close to zero, revealing that teleconnections between continents are weak. Overall, forced simulations appear to have much more homogeneous inter-regional patterns than the reconstructions. Potential reasons for this discrepancy include an overestimation of the response to external forcing by the models, or the effect of noise and uncertainties associated with the proxy-based reconstructions.

Chapter 3

On-line and off-line DA: a case study

3.1 Introduction

Ensemble-based DA methods for palaeoclimate research were pioneered by Goosse et al. (2006), who employed a simplified global 3-D climate model. An updated version was employed by Goosse et al. (2010), using a more advanced climate model. In the first study, 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 study, 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. A conceptual schematic of the two approaches is given in Figure 3.1. The novelty of the present chapter is the focus on the comparison of the on-line and off-line ensemble-based DA approaches. We 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).
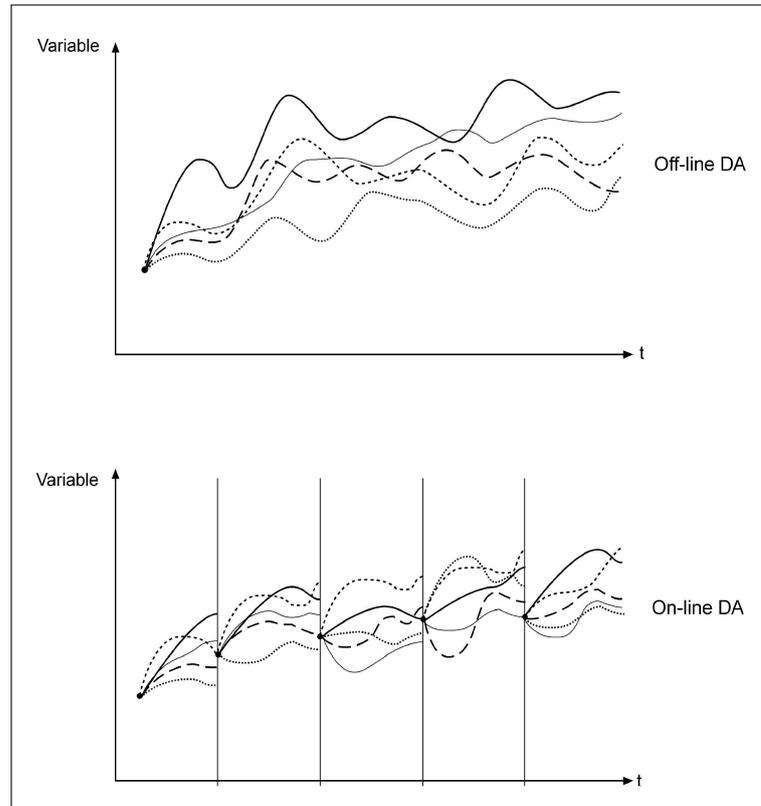
Figure 3.1: Time evolution of a simulated variable in the off-line and on-line ensemble-based DA techniques.

An advantage of the on-line compared with the off-line assimilation is the temporal consistency of the simulated states. The off-line approach on the other hand is computationally less complicated and can be computationally cheaper if one uses simulations that already exist. The main question addressed in this study is whether the on-line reconstruction is closer to the proxy-based reconstructions compared with the off-line version. This largely depends on the memory of the slow components of the climate system, e.g. 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 experimental design is set up with decadal assimilation, motivated by a number of reasons. Firstly, since we aimed for a complete NH reconstruction, the 10-year resolution of the North American proxy-based reconstructions did not allow the use of 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 with decadal variability. Decadal variability is composed by a forced and an internally generated component in the climate system, with the latter being often random and relatively large, particularly on sub-hemispheric scales (Hawkins and Sutton, 2009a). However, the slow components of the climate system, in particular the ocean circulation and to a lesser extent the ice cover and soil moisture, make some aspects of decadal variability predictable to some extent.

Potential for decadal ocean predictability and evidence for forcing of the atmosphere by the ocean on decadal timescales has been found, for example by Branstator et al. (2012), who demonstrated that GCMs exhibit up to decadal predictability in the North Atlantic. The ocean predictability can in turn lead to atmospheric predictability (e.g. Hawkins and Sutton, 2009a,b; Keenlyside and Ba, 2010). The extent of decadal predictability and the relevant mechanism behind are not yet clear (e.g. Hawkins and Sutton, 2009a,b; Keenlyside and Ba, 2010). Intensive work is being undertaken by leading researchers on this emerging issue, but the high cost of the subsurface observations and the small length of the observational record hamper the analysis of decadal processes (Hawkins and Sutton, 2009b).

The structure of the chapter is as follows: In Section 3.2, we review the model characteristics and the proxy datasets used, and give the details of our methodology. Section 3.3 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 3.4, we summarize, draw conclusions and discuss the benefits of each approach.

3.2 Experimental Design

3.2.1 Model Simulations

The ensemble simulations described here are based on the simulation described in Section 2.2, which covers the last millennium (850-1849 AD) and follows the "past1000" protocol of the PMIP3 (Schmidt et al., 2011). We used the MPI-ESM model, comprising of the GCMs ECHAM6 (Stevens et al., 2013) for the atmosphere and MPIOM (Marsland et al., 2003) for the ocean. Like in the previous chapter, ECHAM6 was run at T31 horizontal resolution (3.75°×3.75°) with 31 vertical levels, and MPIOM was run at a horizontal resolution of 3.0° (GR30) with 40 vertical levels. 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 2 m or sea surface temperature, but also in deeper ocean variables, such as the Atlantic Meridional Overturning Circulation (MOC). The selected ensemble generation method does not directly introduce any disturbance in the

47

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.
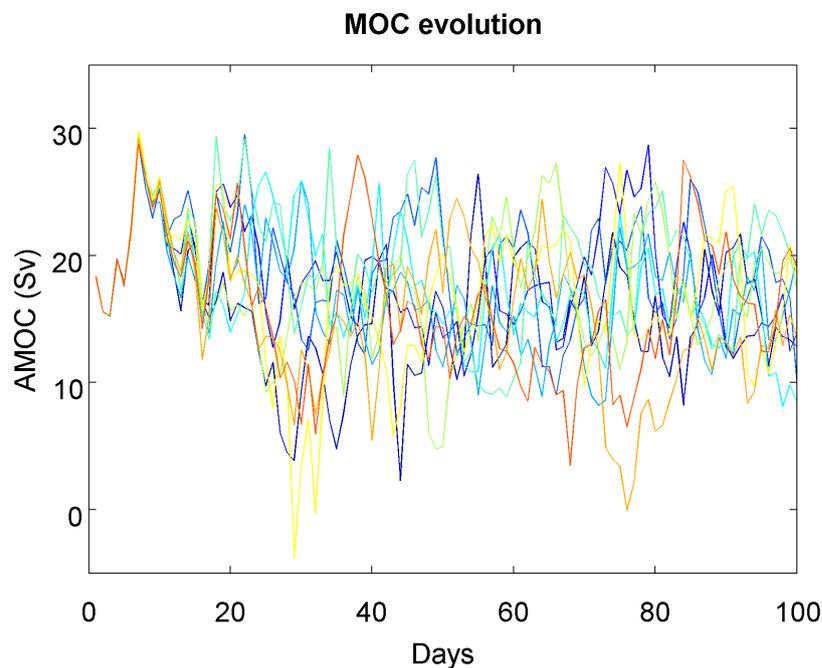
Figure 3.2: 26.5° MOC time series of the ensemble spread for the first 100 days after the initialisation of the ensemble in year 1600 AD, measured in Sverdrups (1 Sv = 10^6 m^{3}s^{-1}).

Figure 3.2 shows the MOC time series of the ensemble spread at 26.5°, for the first 100 days after the initialisation of the ensemble in year 1600 AD. The MOC is one of the main drivers of the northward heat transport in the North Atlantic Ocean, playing an important role in the NH climate. In principle, warm tropical water flows at the surface via the Gulf Stream and the North Atlantic Drift towards the North Atlantic, while a deep current of cold water is transported southward (Lozier, 2012). Variations in the strength of the MOC are believed to influence

the climate of the NH substantially. For example, the release of the transported heat in the atmosphere contributes to the warming of Northern Europe. Changes in the intensity of this circulation can also have global climate impacts, due to the associated modifications in the global ocean heat transport. Figure 3.2 illustrates the fast growth of the ensemble spread in ocean variables, which allowed us to assume that the ensemble spread is adequate.

3.2.2 Proxy Datasets

For our assimilation procedure, we used the "2K Network" of the International Geosphere–Biosphere Programme (IGBP) 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 two millennia 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 LIA. 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 PAGES 2K 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 proxy data have been used to reconstruct annual means for the Arctic (60-90N, 180W-180E), summer (JJA) means for Asia (23.5-55N, 60-160E), summer (JJA) means for Europe (35-70N, 10W-40E), and decadal means of annual values for North America (30-55N, 130-75W). The reconstructions have annual resolution, apart from North America, which is resolved in ten- and thirty-year periods.

3.2.3   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 2 m temperature were calculated for each of the NH continents. Using a root mean square (RMS) error-based cost function, the model outputs were compared with 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:

$$CF(t) = \sqrt{\sum_{i=1}^{k} \left(T_{mod}^i(t) - T_{prx}^i(t) \right)^2} \tag{3.1}$$

where i are the NH continents, namely the Arctic, Asia, Europe and North America, $T_{mod}^i(t)$ is the standardized modelled decadal mean of the temperatures in each NH continent and $T_{prx}^i(t)$ is the standardized proxy-based reconstruction for the decadal mean of the temperatures in each NH 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 NH 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 SH is affected by larger 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. Standardizing variables for multivariate analysis is important, otherwise variables measured at different scales do not contribute equally to the analysis. For example, in the calculation of our cost function, a continental time series that ranges between 0 and 1 would outweigh a time series ranging between 0 and 0.1. Using these time series in the cost function without standardization would give the continent with the larger range a weight of 10. After standardization, all the datasets are treated

as equally important contributors. Our standardization transforms both the model outputs and proxy data to have zero mean (debiasing) and unit variance (adjusting of variance). The standardized model and proxy time series were calculated by the equations:

$$X' = \frac{X - \mu_x}{\sigma_x}, \quad Y' = \frac{Y - \mu_y}{\sigma_y} \tag{3.2}$$

where X' and X are the 1600-1700 AD standardized and raw model output respectively, Y' and Y are the standardized and raw proxy-based time series respectively, μ_x and μ_y are the model and proxy-based mean temperatures for 850-1850 AD respectively, and σ_x and σ_y are the model and proxy-based standard deviations for 850-1850 AD, based on the decadal averages. 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 decadal mean temperatures of the model outputs were compared with 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 NH, 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 NH 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.

3.3 Results

The performance of the two schemes was assessed by computing the correlation and the RMS error for each NH continent between the simulated and the proxy-based reconstructions of the 2 m 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.

3.3.1 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 (raw 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 NH and lower RMS errors for the analysis compared with the individual members. The on-line DA scheme was also successful, improving the skill of the analysis time series compared with 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 3.3 shows the NH 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 NH continents and the NH mean for both DA schemes, as illustrated in Figure 3.4. 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 NH. 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 is 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 highest for the NH direct average (0.94), while the correlation for the NH 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 raw 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 highest for the NH direct average (0.93), and the NH 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
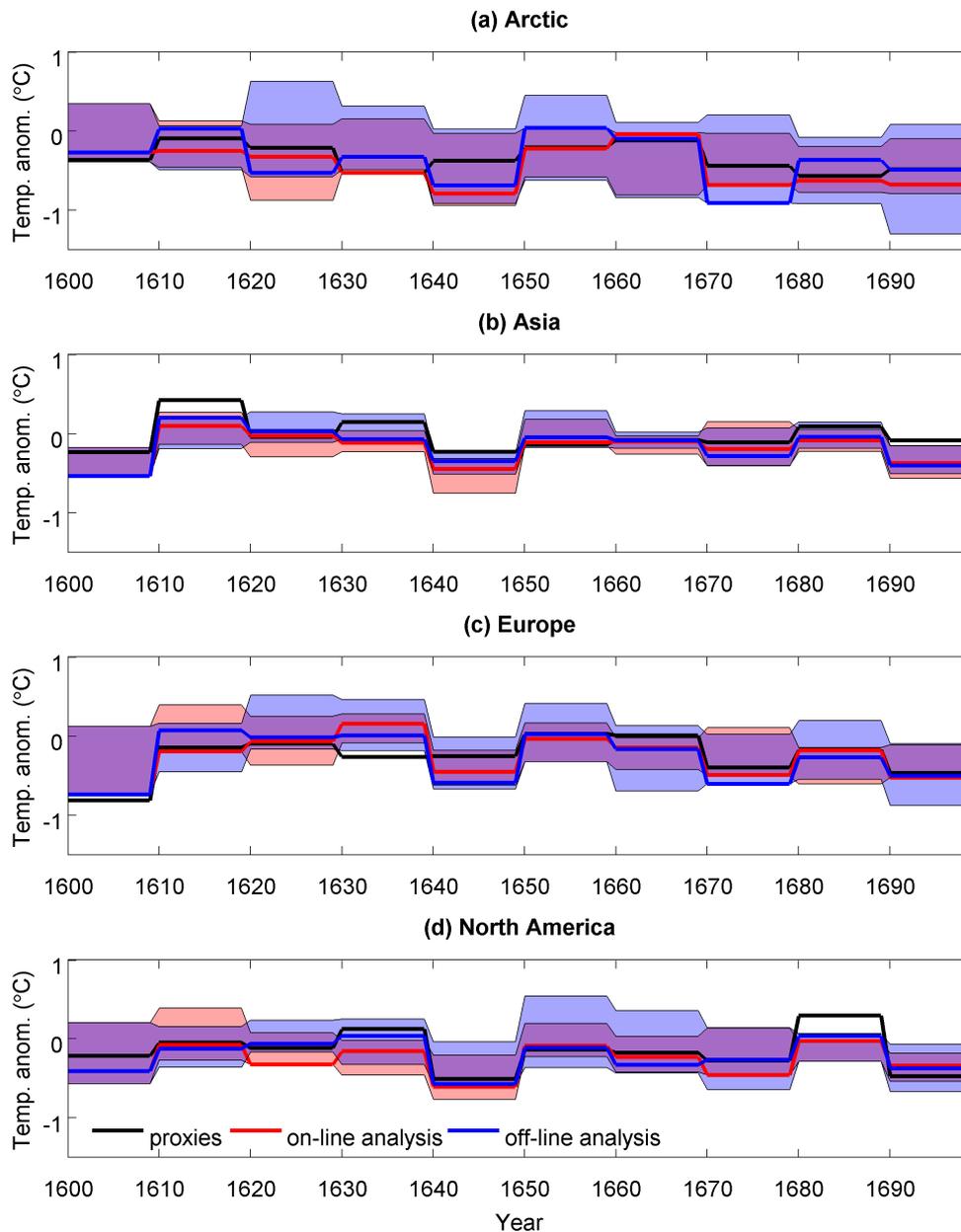
Figure 3.3: Continental decadal mean temperature anomalies w.r.t. the 850-1850 AD mean in the NH for the 17th century, for the on-line (red shading) and off-line (blue shading) ensemble members, the on-line (red line) and off-line DA analysis (blue line), and the proxy-based reconstructions (black line).

and the proxy time series for each continent. Figure 3.5 shows the RMS errors for the individual members, the ensemble mean and the analysis of the four NH continents in the two DA schemes. In both experiments, the RMS errors are either minimal or among the lowest for the analysis compared with all other members. The result is
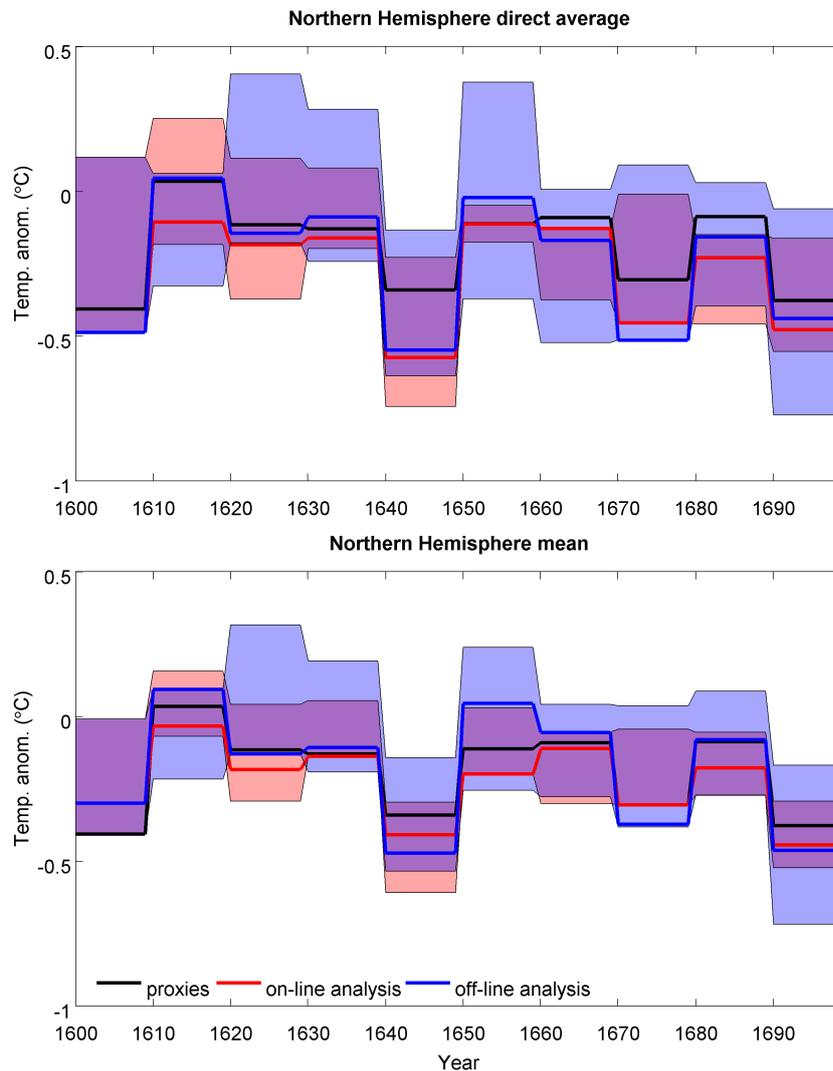
Figure 3.4: Direct average of the four NH continental temperatures (anomalies w.r.t. the 850-1850 AD mean) and NH mean for the 17th century, for the on-line (red shading) and off-line (blue shading) ensemble members, the on-line (red line) and off-line DA analysis (blue line), and the proxy-based reconstructions (black line).

even more evident when considering the RMS errors for the direct average and the mean of the NH (Figure 3.6). 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 with 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 NH continents is 0.12, not considerably different to the off-line one (0.11).
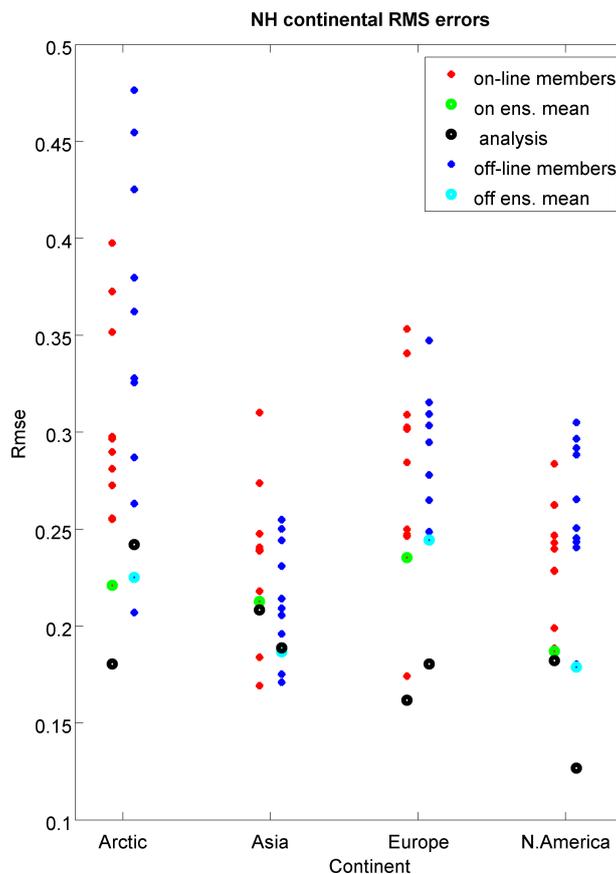
Figure 3.5: RMS errors for the four NH continents for the 17th century, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (green dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

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 raw anomalies as presented above. For the SH, 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
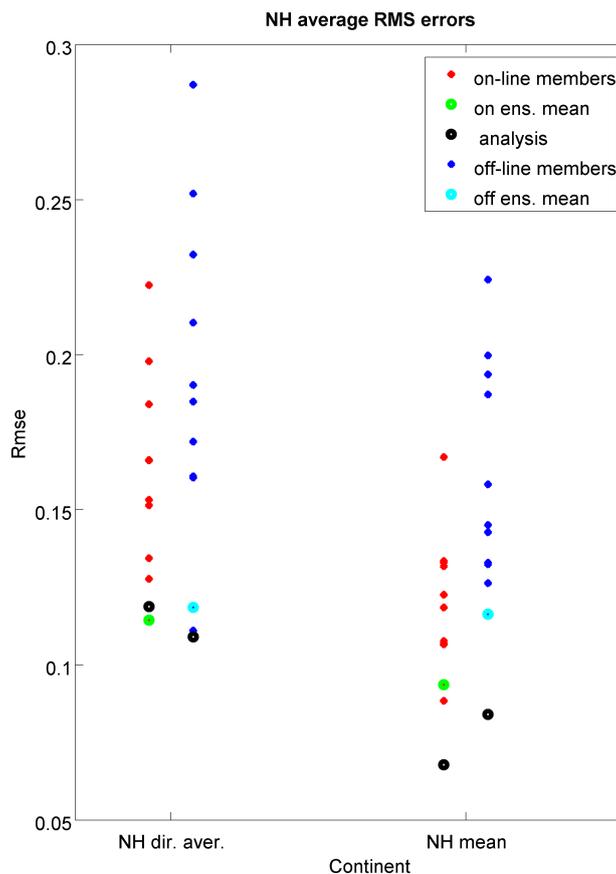
Figure 3.6: RMS errors for the direct average and the mean of the NH for the 17th century, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (green dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

in this case would result in non-comparable scales because of the different standard deviations between model and proxies. Figure 3.7 shows the SH continents' decadal mean temperature time series, standardized w.r.t. the 850-1850 AD mean, for the ten off-line ensemble members, the proxy-based reconstructions and the off-line DA analysis. In contrast to the good skill of the two schemes in the NH, the agreement between the analysis for the SH and the reconstructions is not good, as expected from the fact that SH data are not included in the cost function. This is also proved by both the RMS error and correlation calculations.

The construction of our cost function on the basis of decadal average temperatures of the NH, means that the analysis is not expected to be more skilful than the
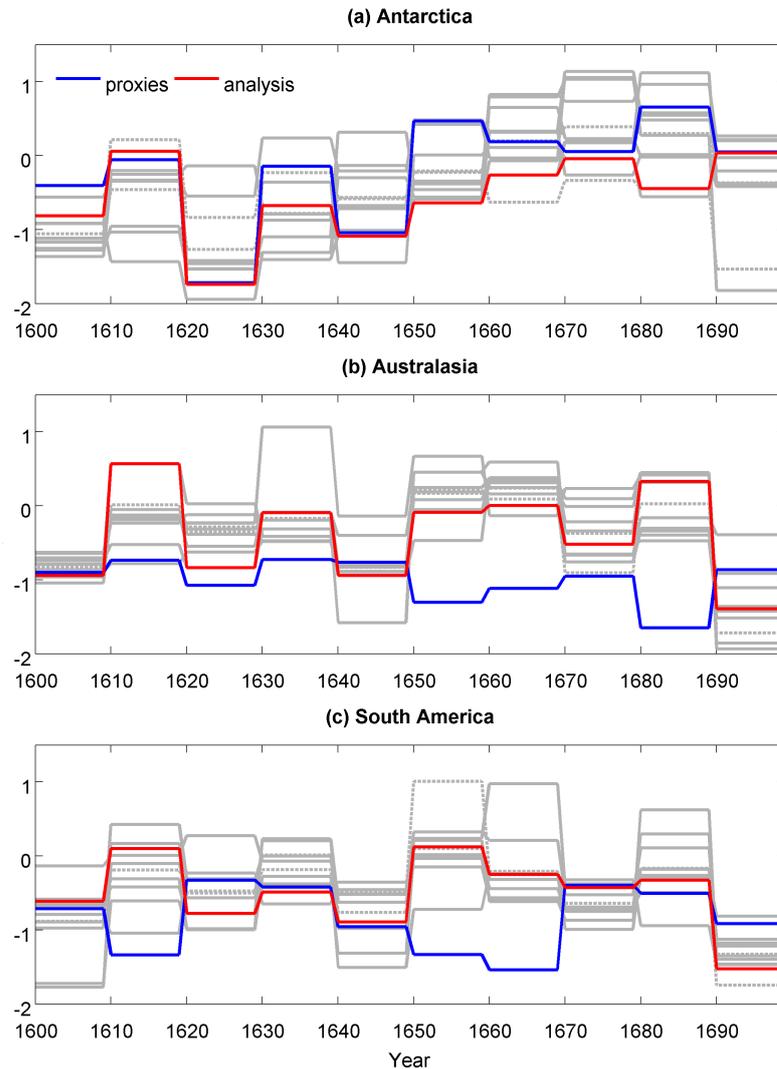
Figure 3.7: SH continents' decadal mean standardized temperature time series for the 17th century, for the off-line ensemble members (gray lines), the proxy-based reconstructions (blue line), and the off-line DA analysis (red line).

individual members when considering the hundred-year average. The absolute differences between simulated and reconstructed 17th century average temperatures, 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 3.8, 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.
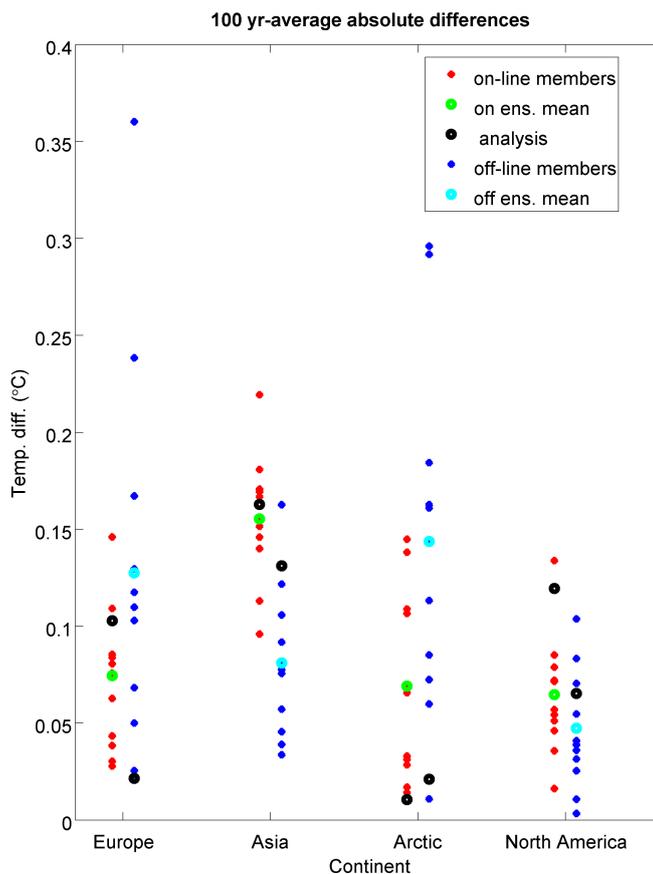
Figure 3.8: Absolute differences between simulated and reconstructed 17th century average temperatures, for the on-line (red dots) and off-line (blue dots) ensemble members, the on-line (green dots) and off-line (cyan dots) ensemble means, and the two analyses (black dots).

3.3.2 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 reconstruc-

tions 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 with the value of 0.94 from the off-line DA analysis. The above results are demonstrated by the histogram of Figure 3.9. 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.
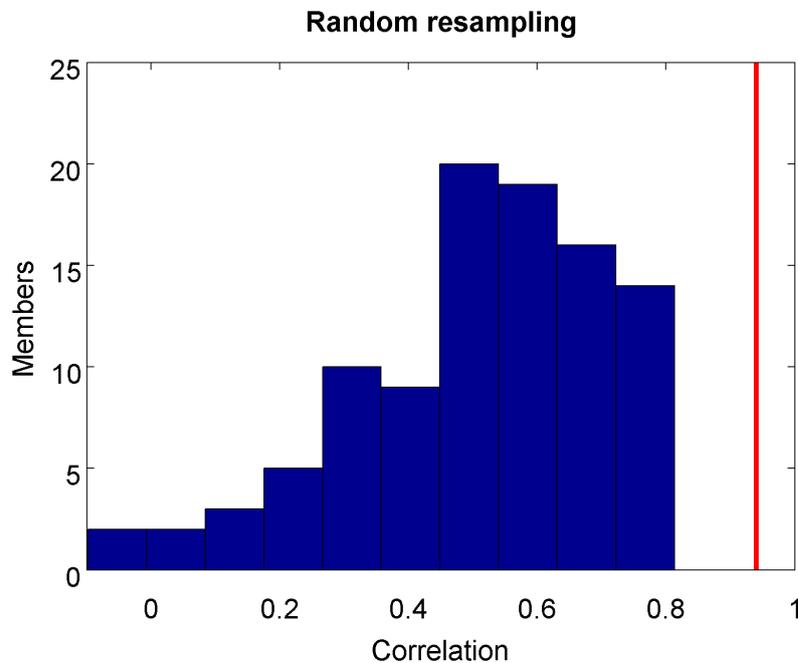
Figure 3.9: Correlations of the randomly sampled members with the proxy-based reconstructions (blue bars), compared with the correlation of the off-line DA analysis with the proxy-based reconstructions (red line).

The same resampling experiment was performed for the RMS error of the NH direct average, shown in Figure 3.10. The mean RMS error was 0.62 (with a standard deviation of 0.13), ranging between 0.4 and 0.9. 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 thus significantly different from the skill obtained from the random sampling.
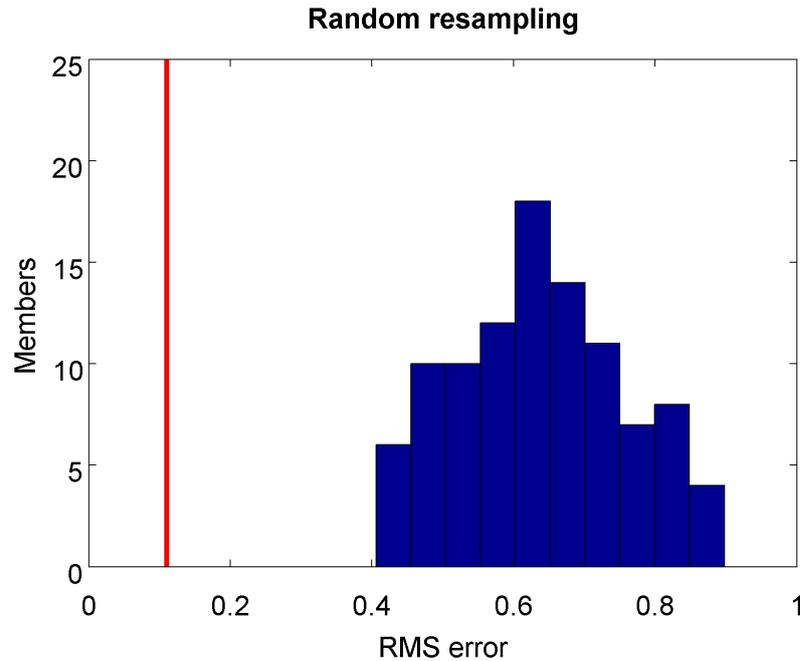
Figure 3.10: RMS errors between the randomly sampled members and the proxy-based reconstructions (blue bars), compared with the RMS errors between the off-line DA analysis and the proxy-based reconstructions (red line).

3.3.3 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 3.1). The respective ensemble mean (EM) cost functions are also shown in the table and are substantially larger than in the DA cases. Tables 3.2 and 3.3 summarise the NH correlations and RMS errors respectively, between simulations and proxy-based reconstructions for the

analysis and the ensemble mean of the two DA 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 3.11, which shows the 2 m mean temperature for the two analyses (anomalies w.r.t. the 1961-90 AD mean) and the 500 hPa 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. An interesting question to answer is whether there would be differences in the performance of the two schemes if we included circulation indices from proxies. It is also interesting to examine the usefulness of including more proxy information from the ocean (e.g. the North Atlantic), to get a better estimate of the ocean state.

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

	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 3.2: NH correlations between simulations and proxy-based reconstructions for the analysis and the ensemble mean of the two DA schemes.

There are three potential reasons for the fact that the on-line method does

	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 3.3: NH RMS errors between simulations and proxy-based reconstructions for the analysis and the ensemble mean of the two DA schemes.
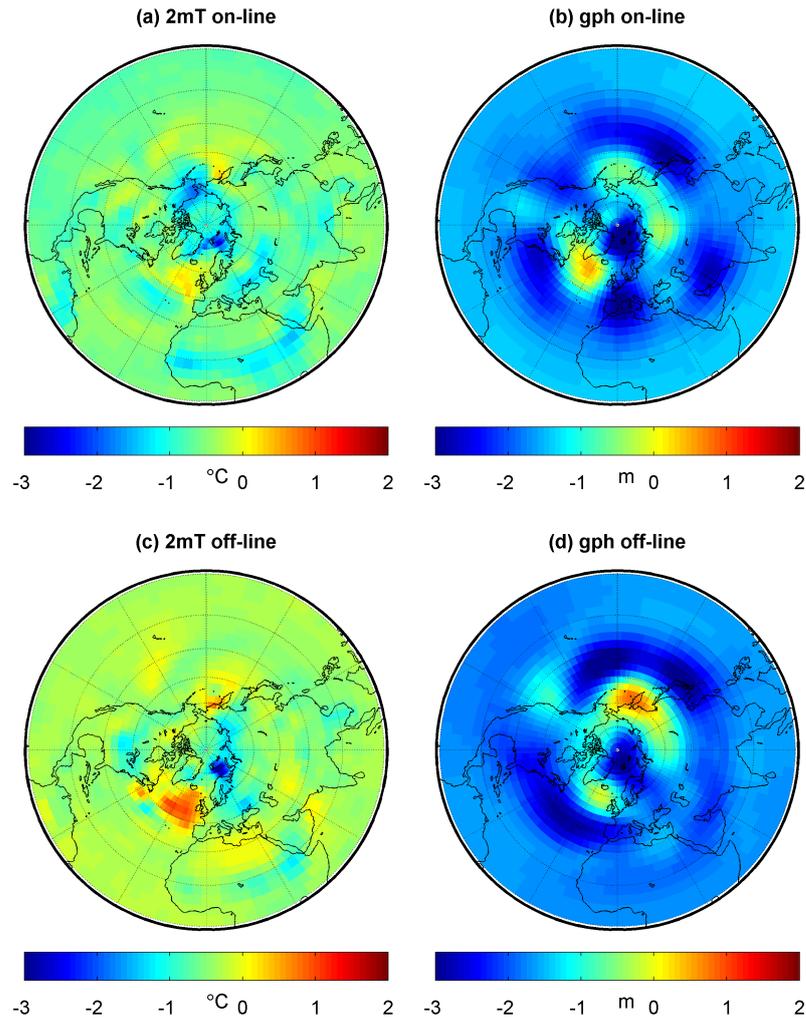
Figure 3.11: Analyses of the on-line and off-line DA methods for the 2 m mean temperature (anomalies w.r.t. the 1961-90 AD mean) and 500 hPa geopotential height (anomalies w.r.t. the 1961-90 AD mean) of the decade 1640-49 AD.

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 one 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 factors 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 temperature anomalies for each ensemble spread, for the different continents. The results are shown in Table 3.4. 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 with 0.30 for the off-line ensemble in the yearly test, and 0.28 compared with 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 with 0.23 for the off-line in the yearly test, and 0.22 compared with 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 3.3 and 3.4. It can also be noted that the all-year ensemble spread for the on-line method is consistently smaller than the respective last year spread. 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 with the off-line one, which starts from different ocean initial states, hints at 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.

	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

Table 3.4: Standard deviations of the ensemble spreads for the NH temperatures of the two DA schemes, calculated for all the years and for the last year of each decade.

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. The 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 assimilating 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. There could be various reasons why 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.

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.

3.4   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, 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 with 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 were 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 with 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.

However, 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 study. 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.

Chapter 4

Reconstructing the climate of the late pre-industrial period

4.1 Introduction

Despite the similar performance of the off-line and on-line DA schemes examined in the previous chapter and the fact that the on-line scheme is more difficult to implement, the advantage of the temporal consistency of the analysis makes the on-line selection technique more desirable than the off-line one. In this chapter, the selected on-line DA method is employed to simulate the climate for the period 1750-1850 AD. We analyse temperatures on the continental scale that has been used in the assimilation, as well as hemispheric and global scales. Moreover, we investigate the performance of the DA on smaller spatial-scale temperature variability within Europe and modes of atmospheric variability, to evaluate the added value on patterns and modes that are not assimilated.

The simulation period, 1750-1850 AD, is towards the end of the LIA, the spatial and temporal extent of which is often debated (e.g. Jones and Mann, 2004; Jansen et al.,

71

2007). This choice of the analysis period has several advantages. A relatively high number of proxies have contributed to the PAGES 2K continental reconstructions used in the assimilation, thus reducing the proxy-induced errors. Additionally, the period is directly before the instrumental period and a relatively high amount of climate data, e.g. early instrumental records and historical documents, are available. Moreover, a major feature of the chosen period is the occurrence of strong tropical volcanic eruptions, which are a major natural driver of the interannual climate variability (Robock, 2000; Cole-Dai, 2010). The eruptions, such as the tropical eruption VEI-6 (unknown location, 1809 AD), the Tambora (Indonesia, 1815 AD) and Cosiguina (Nicaragua, 1835 AD) eruptions, have various dynamical effects on the ocean and the atmosphere (Zanchettin et al., 2012, 2013a). Even though our DA scheme is applied with a time step of 10 years and the peak response to volcanic forcing is on the annual timescale, some forcing signal on the decadal timescale can be expected.

Volcanic eruptions impose short-term energy imbalances on the climate system, with a duration of one to two years, due to the ejection of large amounts of sulfur in the atmosphere, causing a temporary, strong near-surface global cooling (Robock, 2000). This is a combined effect of the aerosol absorption of outgoing infrared radiation and of the scattering of incoming solar radiation (Robock, 2000; Cole-Dai, 2010). The eruptions also trigger dynamical alterations of the atmospheric and oceanic circulation. Zanchettin et al. (2012) showed, via ensemble-based composite analysis based on Earth system model simulations of the last millennium (Jungclaus et al., 2010), the profound influence of volcanic forcing on the long-term climate variability. Typical post-eruption decadal features include a decadal-scale positive phase of the winter NAO, along with a delayed winter warming over Europe, that peaks approximately 10 years after a major eruption (Zanchettin et al., 2012, 2013a). However, the evolution of simulated near-surface atmospheric and oceanic dynamics after the eruption are strongly influenced by the background conditions, i.e. the

initial climate state and the external forcings (Zanchettin et al., 2013b).

For the simulations employed in this study, we use the MPI-ESM model and assimilate decadal surface temperature means of the NH continents from the PAGES 2K project (PAGES 2K Consortium, 2013), as in Chapter 3. Even though no information about the local temperatures is assimilated, we explore the performance of the DA on small spatial scales, since the assimilation of the NH continental averages might determine to some extent the state of the main modes of circulation variability, such as the NAM or the NAO. In principle, prescribing continental temperatures can be expected to constrain the phase of leading circulation modes if these modes have a temperature signal associated with the average continental temperatures. In turn, the temperature signal of these circulation modes can be expected to provide information on temperature variability on sub-continental scales.

The validation of the DA setup is performed against assimilated and independent data, while a MCA of links between temperature and pressure in the NH aims to explore the potential of added value in the assimilation run. The focus of the study is on the skill of the reconstruction using DA and the added value that can be achieved compared with simulations without DA. The structure of the chapter is as follows: in Section 4.2, we review the characteristics of the model, proxy and instrumental datasets, followed by the details of the experimental design. Section 4.3 validates the assimilation approach for large-scale temperatures by comparing the simulation with the assimilated and independent data, and other simulations. The circulation and temperature variability in the North Atlantic-European sector for the assimilation run is examined in Section 4.4. Finally, in Section 4.5, we summarize and draw the main conclusions.

4.2 Methodology

4.2.1 Model, proxies and instrumental data

The assimilation has been performed with the same version of the MPI-ESM model as in the previous chapters (ECHAM6 at T31 resolution and MPIOM at GR30 resolution). As mentioned previously, the model is a coarse-resolution version of the model used for the CMIP5 simulations, and hereafter is referred to as MPI-ESM-CR. The configuration follows the configuration for palaeo-applications (MPIESM-P) described in Jungclaus et al. (2014). The DA simulations consist of 20 ensemble members for each decade between 1750 and 1850 AD, and include all natural and anthropogenic forcing. Two different types of simulations using the MPI-ESM-CR are examined, namely simulations with and without DA. In addition to the forced simulations, a 1,000 year-long control run was performed and used for a MCA of links between temperature and pressure in the NH due to internal variability. In terms of the assimilation data, the DA simulations are constrained to follow the "2K Network" of the IGBP PAGES datasets (PAGES 2K Consortium, 2013), described in Section 3.2.2. Only the reconstructions of the NH continents are assimilated, for the reasons explained there.

For validation purposes, the skill of the DA simulations is assessed against the European seasonal surface air temperature reconstruction from Luterbacher et al. (2004). This statistical reconstruction provides gridded (0.5°×0.5° resolution) monthly (back to 1659 AD) and seasonal (from 1500 AD to 1658 AD) temperatures for European land areas (25W-40E, 35-70N). It is based on homogenized and quality-checked instrumental data, reconstructed sea-ice and temperature indices derived from documentary records and seasonally resolved proxy temperature reconstructions from Greenland ice cores and tree rings.

The DA simulations are also compared with the early instrumental record; specifically, the Berkeley Earth Surface Temperature (BEST) dataset (Rohde et al., 2012), which uses temperature observations from a large collection of weather stations in order to estimate the underlying global land temperatures. Temperatures are reported as anomalies relative to the 1951-1980 AD average, along with their uncertainties, which represent the 95% confidence interval for statistical noise and spatial undersampling effects. The uncertainties are larger in the earlier reconstructions and account for the effects of random noise as well as random biases affecting station trends and random shifts in station baselines. The BEST framework is expected to be robust against most forms of bias (Rohde et al., 2012).

4.2.2   DA experimental design

The method we employ for assimilating the PAGES 2K proxy-based reconstructions is similar to the ones followed in recent ensemble-based DA studies (e.g. Goosse et al., 2006; Crespin et al., 2009; Matsikaris et al., 2015a) and is based on the on-line degenerate particle filter (Section 3.2.3). We set up the DA by taking the last day of the year 1749 AD from a transient forced simulation starting in 850 AD (the "past1000" simulation) as the initial conditions of the ensemble. We generate 20 ensemble members by introducing small perturbations in an atmospheric diffusion parameter for the first year of the simulations, 1750 AD. After 10 years of simulations, a RMS error-based cost function (Equation 3.1) is used to compare the simulated decadal mean temperatures of the NH continents with the PAGES 2K continental proxy-based reconstructions. The member that minimizes the cost function is selected as the best member for that sub-period and is used as the initial condition for the subsequent 10-year simulation. A new ensemble consisting of 20 members is performed for the second decade, using the same method as before.

The procedure is repeated sequentially until the end of the simulation period, 1850 AD. We combine each decade's best member to form the "DA analysis". The high computational cost did not allow a larger number of ensemble members, however the ensemble is twice as large as that in the previous chapter. More details for the specific choices of the cost function are given in Section 3.2.3.

The proxy-based reconstructions are affected by various types of errors, which in turn affect the cost function and the member selection directly. More specifically, the statistical methods followed influence the reconstructions, the seasonal representativity is different in different proxies, non-climatic factors influence the proxies, while the poor spatial coverage induces large uncertainties in hemispheric or continental means, such as the ones we assimilate (e.g. Jones and Mann, 2004; Jansen et al., 2007). As a result of these errors, there is a high chance that the selected ensemble member is not in agreement with reality. This problem is often tackled by either weighing the cost function according to the errors of the proxy datasets, or by retaining several members that are close to the proxy-based reconstructions (e.g. Goosse et al., 2012; Annan and Hargreaves, 2012). Our simple assimilation scheme does not take the proxy errors into account, because they are not directly comparable for the different continents due to the different methods followed by each of the PAGES 2K groups. The use of proxies with minimum noise would give a better chance to capture the true climatic state.

4.3 Validation for continental to global-scale temperatures

The DA simulations are validated against empirical evidence for large-scale temperature variability, namely continental, hemispheric and global scales, in three ways. Firstly, we validate the simulations against the proxy-based reconstructions

used during the assimilation, i.e. the PAGES 2K data, to check the extent to which the two are consistent. Secondly, the assimilation results are tested against other proxy-based temperature reconstructions, including Luterbacher et al. (2004). The independence between those and the assimilated reconstructions is often not clear, as predictors used in different proxy-based reconstructions may be common. Nevertheless, proxy data from Luterbacher et al. (2004) have not been used by the PAGES 2K groups in their reconstructions. Thirdly, the DA analysis is compared with early instrumental records (BEST dataset). We also evaluate the accordance of the DA analysis with simulations that do not perform DA.

4.3.1 Comparison of simulations with the assimilated proxy data

We compare the simulated NH continental temperature series with the assimilated PAGES 2K proxy-based reconstructions. We initially compare the decadal means, which is the timescale used in the DA. Figure 4.1 shows the NH continents' decadal mean temperature anomalies for 1750-1850 AD w.r.t. the 850-1849 AD mean, for the DA analysis, the different ensemble members, the ensemble mean, the simulation without DA ("past1000") and the proxy-based reconstructions. The simulated data are for the same regions as described by the PAGES 2K reconstructions. The analysis follows the assimilated reconstructions well, which is a prerequisite for a skilful DA method and indicates a sufficiently large ensemble size. However, in decades with strong volcanic forcing, namely 1810-20 AD and 1830-40 AD, some agreement also stems from a common response to the forcings. Correlations between analysis and proxy-based reconstructions in most continents (apart from Asia) are higher than the respective correlations found in Section 3.3 (Matsikaris et al., 2015a) (0.93 instead of 0.79 for the Arctic, 0.64 instead of 0.76 for Asia, 0.95 instead of 0.79 for Europe and 0.96 instead of 0.81 for North America). This may be due to

the increase in ensemble size from 10 to 20 members, or indicate a more realistic or stronger forcing in the 1750-1850 AD period compared with the earlier period, 1600-1700 AD, analysed in Matsikaris et al. (2015a). The correlations of the "past1000" simulation with the PAGES 2K reconstruction are much lower (0.70, 0.45, 0.57 and 0.43 respectively) than those of the DA analysis, showing that DA improves the skill on the continental scale.

The "past1000" simulation includes the forcing signal and random internal variability. To focus on the forced variability we also investigate the ensemble mean for the DA simulations. However, it should be noted that the ensemble is generated sequentially, which means that all ensemble members are to some extent influenced by the assimilated data. How strong this effect is depends on how long the empirical information in the initial state for each decade is retained for in the system. It is also noteworthy that the ensemble mean is not a physically consistent state. The correlations between the DA ensemble mean and the PAGES 2K reconstructions are 0.73 for the Arctic, 0.52 for Asia, 0.92 for Europe and 0.71 for North America. These values are higher than the correlations of the PAGES 2K reconstruction with the "past1000" simulation, as the latter includes more random internal variability than the ensemble mean. The correlations of the DA analysis with the proxy-based reconstruction are highest. However, in the decades with strong influence of the volcanic forcing (1810-1820 AD and 1830-1840 AD), the ensemble mean is closer to the PAGES 2K data than the DA analysis. This is likely to be due to the relatively small ensemble size, which does not always allow the best member to capture the true internal variability. For a DA analysis that includes the forcing signal and the wrong internal variability, the cost function can have a higher value than that of a simulation with the forcing signal and much lower internal variability, such as the ensemble mean.

We now examine the skill of the DA on shorter timescales (5-yr running mean) in the
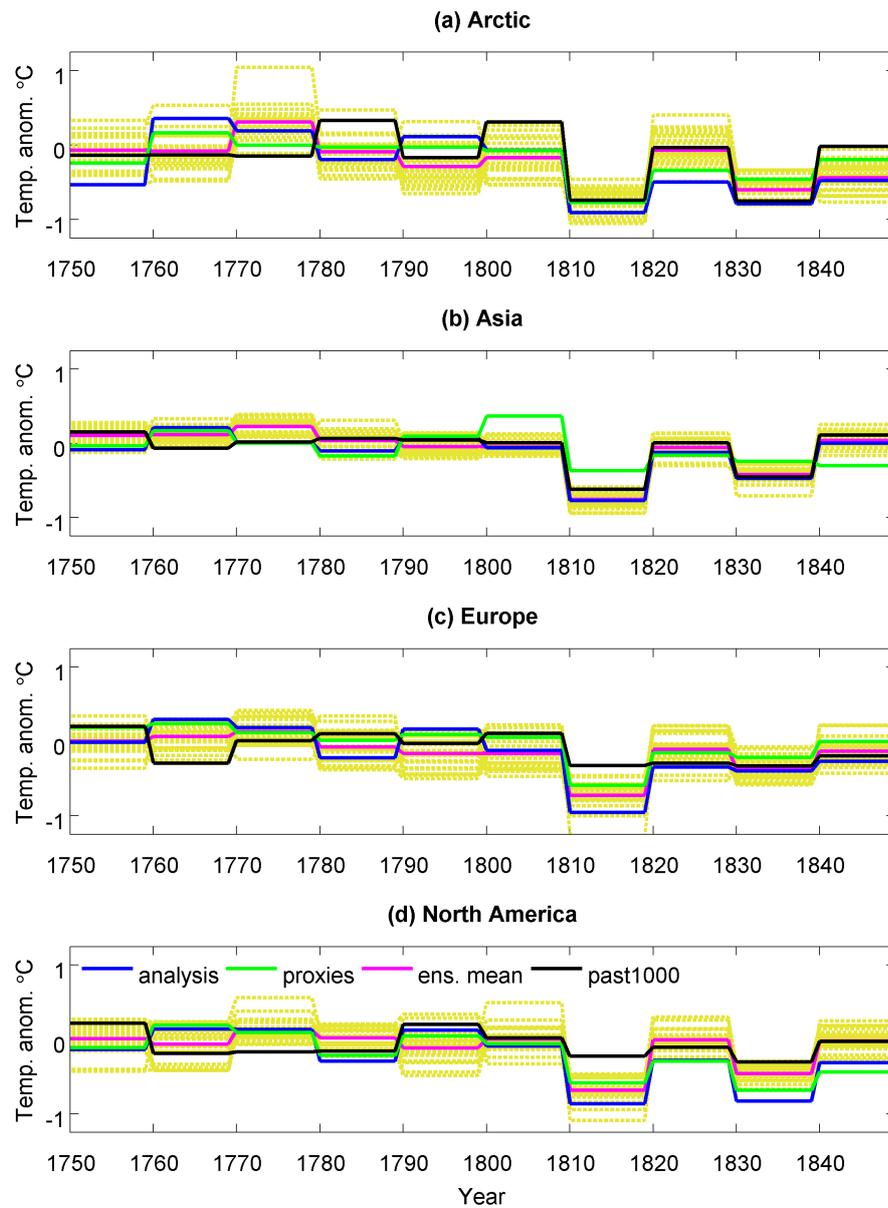
Figure 4.1: Continental decadal mean temperature anomalies for 1750-1850 AD w.r.t. the 850-1849 AD mean in the NH, for the DA analysis (blue line), the PAGES 2K proxy-based reconstructions (green line), the ensemble mean (magenta line), the simulation without DA (black line) and the individual ensemble members (yellow lines).

NH continents, apart from North America for which the PAGES 2K reconstruction has decadal resolution (Figure 4.2). Agreement between the DA analysis and the proxy-based reconstructions for the individual annual values is not expected, as only decadal average temperatures have been assimilated. However, some overall

agreement might be caused by the fact that the decadal averages are formed by the individual annual averages. The main possible source of consistency is the response of the model and the proxies to the forcings. Figure 4.2 indicates some agreement between the DA analysis and the assimilated temperatures, with moderate positive correlations (0.72 for the Arctic, 0.64 for Asia and 0.62 for Europe). These correlations are again higher than the correlations of the "past1000" simulation with the PAGES 2K reconstructions (0.63, 0.52 and 0.40 respectively). Thus, the response to the forcings, in particular to the volcanic eruptions, provides the main contribution to the correlations of the shorter timescales, but some small added value is obtained from DA.

The control experiment mean anomalies and the one standard deviation range in Figure 4.2 give an estimate of the internal variability. The differences between the DA analysis and the proxy-based reconstructions are smaller than the range of natural variability in some periods, showing the combined effect of the forcings and of the DA. The volcanic events of 1809-1815 AD leave a similar imprint in all continents, while the smaller scale event of 1835 AD greatly affects North America, where the volcano is located, and to a lesser extent the other continents. The eruption of Mount Tambora in April 1815 AD is the largest known historical strong tropical volcanic eruption (Oppenheimer, 2003; Cole-Dai, 2010) and followed the 1809 AD VEI-6 tropical eruption of unknown location (Cole-Dai et al., 2009), which occurred during the 1790-1830 AD Dalton solar minimum. The fact that the decade 1810-1819 AD was the coldest during at least the past 500 years in the NH and the tropics (Cole-Dai et al., 2009) is attributed to the combined effects of the 1809 AD and Tambora eruptions. Approximately 10 years after the major eruptions, most regions revert back to the natural variability range.
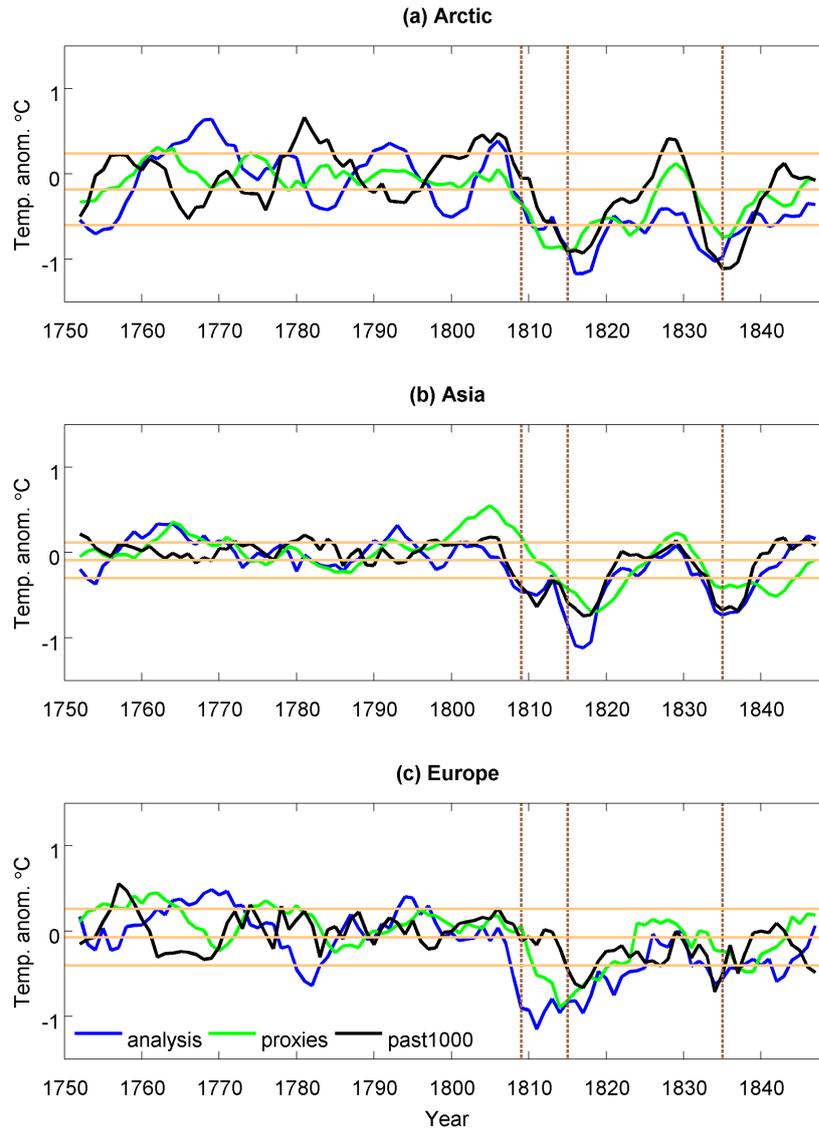
Figure 4.2: Continental temperature anomalies for 1750-1850 AD w.r.t. the 850-1849 AD mean, smoothed with a 5-yr running mean, in Arctic, Asia and Europe, for the DA analysis (blue line), the PAGES 2K proxy-based reconstructions (green line) and the simulation without DA (black line). Brown vertical lines denote the major volcanic eruptions. Yellow horizontal lines indicate the control experiment mean anomalies and the one standard deviation range.

4.3.2 Comparison with independent proxy data

Proxy-based temperature reconstructions for the NH for the last millennium differ significantly among each other and show discrepancies with simulations (Jansen et al., 2007). The period we examine, however, is more recent and

experienced strong forcing variations, thus the agreement among the proxy-based reconstructions is higher than in previous times. Comparing the DA analysis with independent proxy-based reconstructions allows to detect errors due to unrealistic assimilated temperatures. Moreover, if the validation data include information from areas that have not been assimilated, e.g. hemispheric means, the comparison with independent data provides some evaluation of information propagation in space.

Figure 4.3 presents the NH near-surface (2 m) air temperature anomalies w.r.t. the 1961-1990 AD mean for the DA analysis in relation to the range of reconstructions redrawn from Jansen et al. (2007). The grey shading includes published multi-decadal timescale uncertainty ranges of all temperature reconstructions identified in Table 6.1 (except for RMO2005 and PS2004) of Jansen et al. (2007). The proxy-based time series are smoothed with a 31-yr running mean. The reconstruction data used are those featured in Fig. 6.10 of the IPCC Fourth Assessment Report (Jansen et al., 2007). The DA analysis is presented in two ways; with a 31-yr running mean time series, to be directly comparable with the overlap of the proxy-based reconstructions, and with a 15-yr running mean series, to show the variability on shorter timescales. The 15-yr running mean DA analysis falls outside the range of the proxy-based reconstructions, which have averaged all the short timescale variations out, but is in good agreement with the PAGES 2K direct average for the NH, also presented with a 15-yr running mean (correlation 0.95). The correlation of the run without DA with the PAGES 2K direct average is about the same as the latter (0.96). The DA skill on the hemispheric scale is thus very good, but no additional skill compared with the simulation without DA is gained. The 31-yr running mean DA analysis lies well within the range of the proxy-based reconstructions. This is not a strict validation though, as the grey band is quite wide. We use the running mean filter to be consistent with the reconstructions, however in other parts of the study we prefer to use the Hamming window,

which has better filter characteristics. The coldest anomalies in the simulation are between 1810-1820 AD and agree well with the consensus of reconstructions. The severe volcanic eruptions of 1809, 1815 and 1835 AD cause a sharp drop in the mean temperature during the years that follow, as well as a long-lasting effect on the NH climate, recorded in both proxies and the simulation. We note that although the proxy-based reconstructions of Figure 4.3 seem to exhibit more multi-decadal variability than the DA, which is counter-intuitive given the tendency of regression-based reconstructions to reduce variance, this is not necessarily the case, as the grey shading shows the concentration of overlapping NH proxy-based reconstructions, which is different from individual trajectories. Additionally, not all of the proxy-based reconstructions are regression-based, as some of them have used the "composite plus scale" methodology, which does not affect the variance.

The DA analysis is also validated on the decadal timescale against the proxy-based European mean land temperature reconstruction by Luterbacher et al. (2004) (Figure 4.4). The spatial patterns of this reconstruction are also available and are compared with the DA analysis later. Agreement between the DA analysis and the Luterbacher et al. (2004) independent reconstruction is higher in summer, but the DA analysis shows a stronger response to the volcanic eruptions than the proxies and larger variability in both seasons. The correlations are 0.73 for summer and 0.61 for winter. The upper panel also includes the summer decadal means of the PAGES 2K reconstruction. The correlation between the two reconstructions is 0.70. The correlations of the "past1000" simulation with Luterbacher et al. (2004) are 0.75 summer and -0.24 for winter, indicating again an additional skill of DA on the continental scale, at least in winter. Validation against independent data was also performed using the annual time series. Moderate positive correlations were found as expected, due to the forcings, but they were lower than the ones in the decadal case.

NH surface air temperature
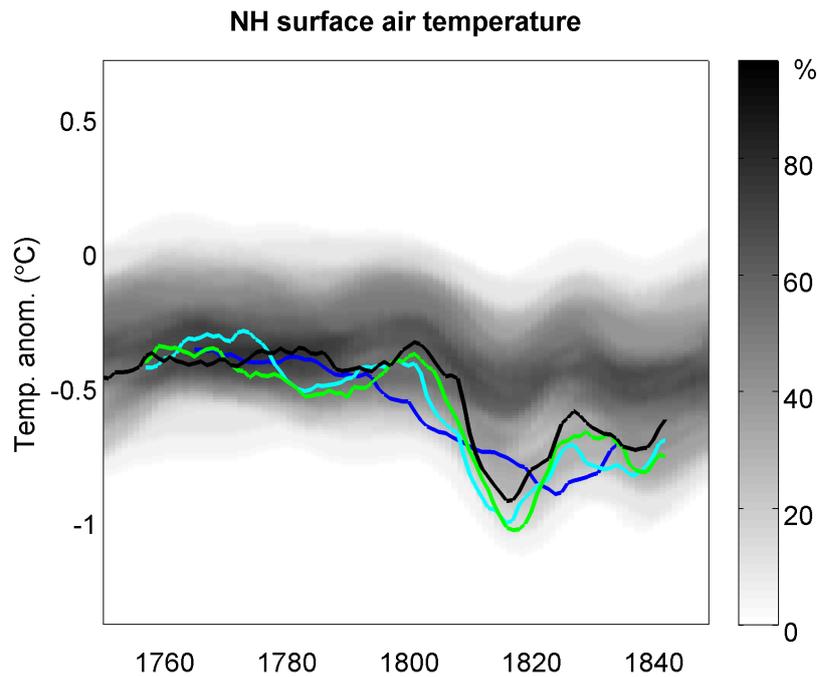
Figure 4.3: NH 2 m temperature anomalies for 1750-1850 AD w.r.t. the 1961-1990 AD mean for the DA analysis (blue line for 31-yr running mean and cyan line for 15-yr running mean) in comparison with the simulation without DA (15-yr running mean, black line) and the range of reconstructions (grey scale), redrawn from Jansen et al. (2007). Data time series and the shaded representation of overlap of proxy-based reconstructions (consensus) were obtained from: http://www.cru.uea.ac.uk/datapages/ipccar4.htm. The PAGES 2K direct average of the NH is also shown (15-yr running mean, green line).

4.3.3 Comparison with instrumental data

An advantage of the investigated period is that some early instrumental records are available, which can be used to evaluate the assimilation method. Here we validate the DA analysis against the instrumental BEST reconstruction (Rohde et al., 2012) for global mean temperatures. Figure 4.5 shows the global land 2 m air temperatures (anomalies w.r.t. the 1951-1980 AD mean), smoothed with a nine point Hamming window (approximately corresponding to a 5-year running mean), as simulated with and without DA, compared with the reconstruction from the BEST dataset. All three time series remain relatively constant over the period, interrupted by the cooling induced by the volcanic eruptions. The simulated DA
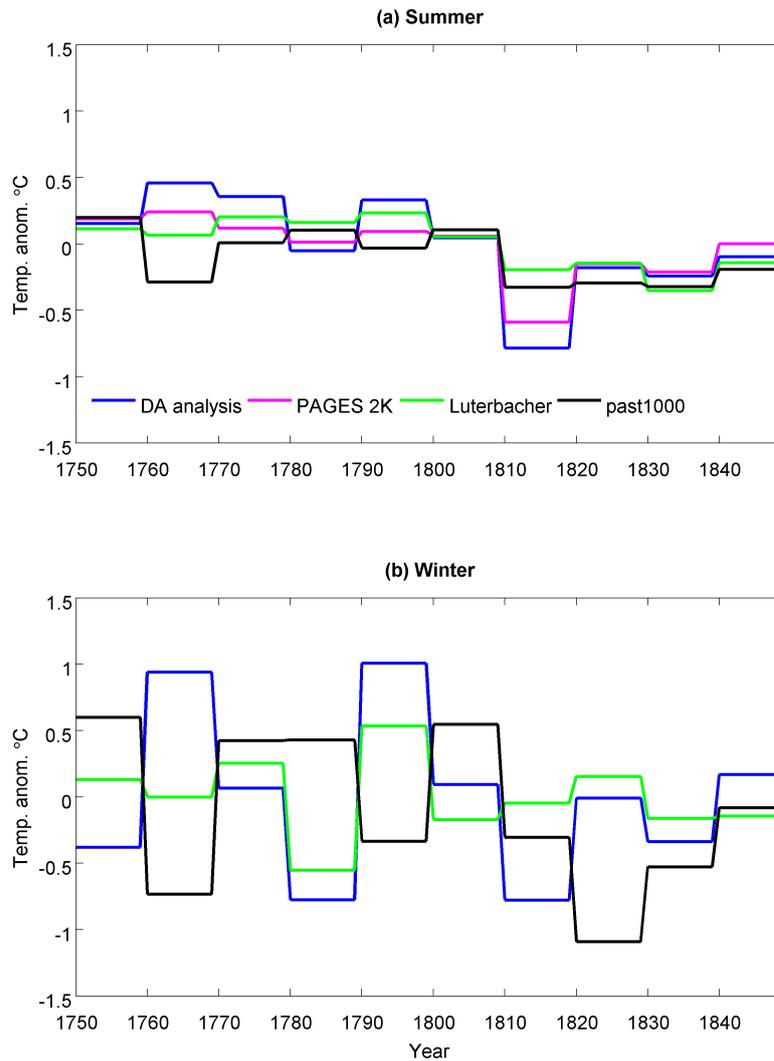
Figure 4.4: European summer (a) and winter (b) decadal mean temperature anomalies for 1750-1850 AD w.r.t. the 850-1849 AD mean, for the DA analysis (blue line), the simulation without DA (black line) and the Luterbacher et al. (2004) proxy-based reconstruction (green line). The PAGES 2K proxy-based reconstruction (magenta line) for the summer period is also shown.

time series includes minima in the 1810s and 1830s temperatures and maxima in the 1770s and 1800s. The correlation between the DA analysis and the BEST dataset is high (0.81), and the correlation of the BEST dataset with the "past1000" run without DA (0.79) is only marginally lower. The correlation of the DA analysis with the "past1000" simulation without DA is very high (0.88).

Global land-only surface air temperature
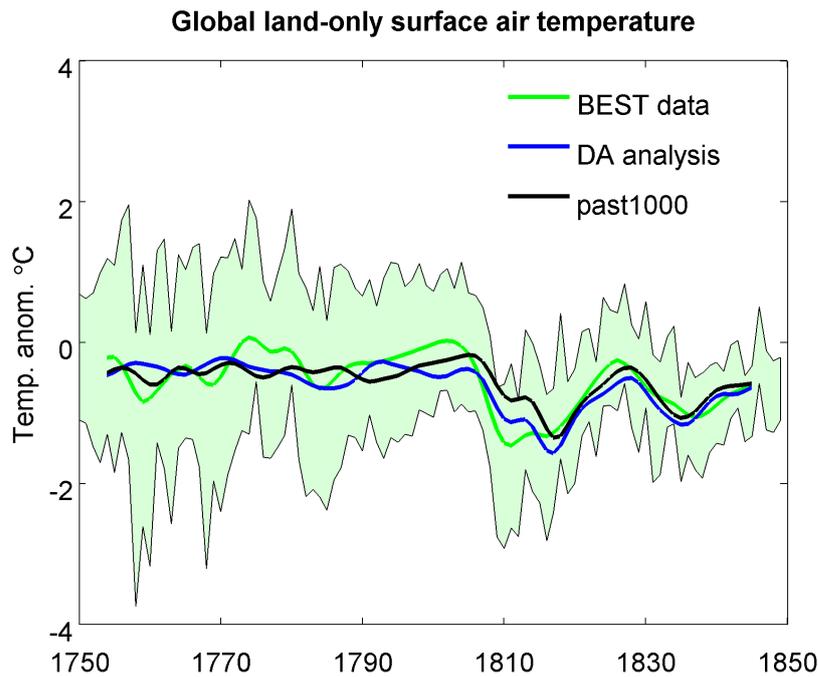
Figure 4.5: Global land surface air temperatures for 1750-1850 AD (anomalies w.r.t. the 1951-1980 AD mean), smoothed with a nine point Hamming window, for the DA analysis (blue line), the simulation without DA (black line), and the BEST instrumental dataset (green line). The green shading shows the 95% confidence interval of the BEST estimate, representing statistical and spatial undersampling uncertainties.

Overall, the performance of the DA scheme on the global scale is very good, but DA does not offer additional skill compared with the simulations without DA. This is similar to what has been found on the hemispheric scale examined earlier. The lack of added value of DA is likely due the influence of the strong forcings and the high ratio of forced to internal variability on large spatial scales. In principle, if the forced response in a model is unrealistic, DA can select an ensemble member that, due to internal variability, is closer to reality. Therefore, DA can correct to some extent for unrealistic response to forcing. However, if the response to the forcing is close to reality and the contribution of internal variability is low, then no systematic change to the response to the forcing can be expected from DA. On the continental scale, the contribution of interannual variability is larger, and if a DA scheme works properly, it can give additional skill compared with a simulation without DA. The

signal (response to the forcing) to noise (internal variability) ratio increases with the spatial scale. Therefore, forcings are sufficient to provide realistic simulations on hemispheric and global means, which have a relatively small contribution from internal variability.

4.3.4 Comparison with other GCMs

To further examine whether DA affects the long timescale (decadal to centennial) variability, we compare the NH surface air temperature anomalies of the DA simulations with other GCM simulations that have not used DA (Figure 4.6). This aims to investigate the behaviour of DA with respect to the forcings and the internal variability, in comparison to simulations without DA. The simulations were part of the CMIP5 or the PMIP3. We display the simulations performed with the models CCSM4 (Community Earth System Model, developed at NCAR, USA), HadCM3 (Hadley Centre Coupled Model, version 3, developed by the Hadley Centre, UK), IPSL-CM5A-LR (Institute Pierre Simon Laplace model, France), and MPI-ESM-P (Max Planck Institute PMIP3 simulations). The simulation performed with the model used in this study, MPI-ESM-CR, without DA, is also shown.

The correlation of the DA analysis hemispheric means with the PAGES 2K NH direct average for filtered (nine point Hamming window) temperatures is 0.85. Data for North America could not be included in the direct average because of their decadal resolution. The DA simulation is not closer to the proxy-based reconstruction than the non-assimilated MPI-ESM-CR and MPI-ESM-P simulations (correlations with PAGES 2K are 0.86 and 0.83 respectively). This shows again that the forcings play the dominant role in giving the skill on the hemispheric scale. The correlation of the DA analysis with the proxy-based reconstruction is higher than the correlations of the CCSM4 (0.24), IPSL-CM5A (0.64) and HadCM3 (0.80) simulations. The DA analysis arrives at slightly lower

temperatures compared with the MPI-ESM-CR without DA towards the end of the simulation period. Multi-decadal variability is similar in most simulations.
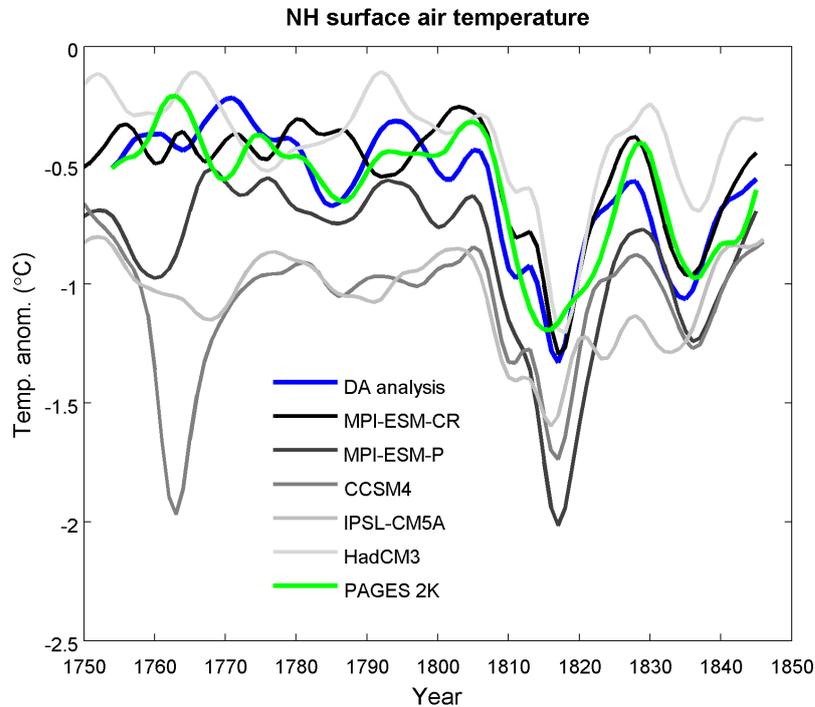
Figure 4.6: Evolution of the NH surface air temperature anomalies for 1750-1850 AD w.r.t. the 1961-90 AD mean, smoothed with a nine point Hamming window, for the DA analysis (blue line) and other GCM simulations without DA. The PAGES 2K direct average of the NH (green line) is also included (for Arctic, Asia and Europe, green line).

4.4   Circulation and temperature variability in the North Atlantic-European sector

One of the main reasons for performing DA is to obtain knowledge about variables that are not assimilated. We now investigate whether this is the case with respect to small-scale temperature variability in Europe and leading modes of circulation variability. The link between continental mean temperatures, which have been used in our assimilation, and large-scale circulation, is crucial for providing this added value. We examine the temperature-SLP link in the control simulation using MCA,

and check whether it is also visible when performing the assimilation.

4.4.1 Spatial European temperature patterns

In our DA scheme, only continental average temperatures have been assimilated. Hence, it is not guaranteed that skill on smaller spatial structures exists, since no information about the local temperatures was inserted in the model. The assimilation of the NH continental averages might however provide information that could potentially determine the state of leading circulation modes, which in turn may lead to skill in reconstructing the smaller-scale spatial patterns. Two possible reasons may lead to added value of DA on these scales: firstly, leading modes of variability, such as the NAM or the NAO, may have a link to the NH continental mean temperatures and be partially captured correctly with DA. Secondly, certain temperature mean values may be associated with tendencies for specific spatial patterns, in which case no explicit atmospheric circulation information is required. This would be the case if, for example, a low mean European temperature was associated with very low Northern European temperatures, less cold Southern European conditions, mild Eastern Europe temperatures etc. This hypothesis is not examined in the current study. In addition, another possibility for agreement on the small spatial scales could stem from the response of the circulation to forcing, such as the strong volcanic eruptions.

We compare the DA-simulated European temperature patterns with the patterns from the Luterbacher et al. (2004) reconstruction. Figures 4.7 and 4.8 show the European decadal 2 m temperature maps (anomalies w.r.t. the 1750-1849 AD mean) for the DA analysis and the reconstructions by Luterbacher et al. (2004) for the summer and winter periods of 1750-1850 AD respectively. The comparison of the patterns in most decades exhibits no agreement. If the Luterbacher et al. (2004) reconstructions are skilful, we can conclude that although the DA scheme
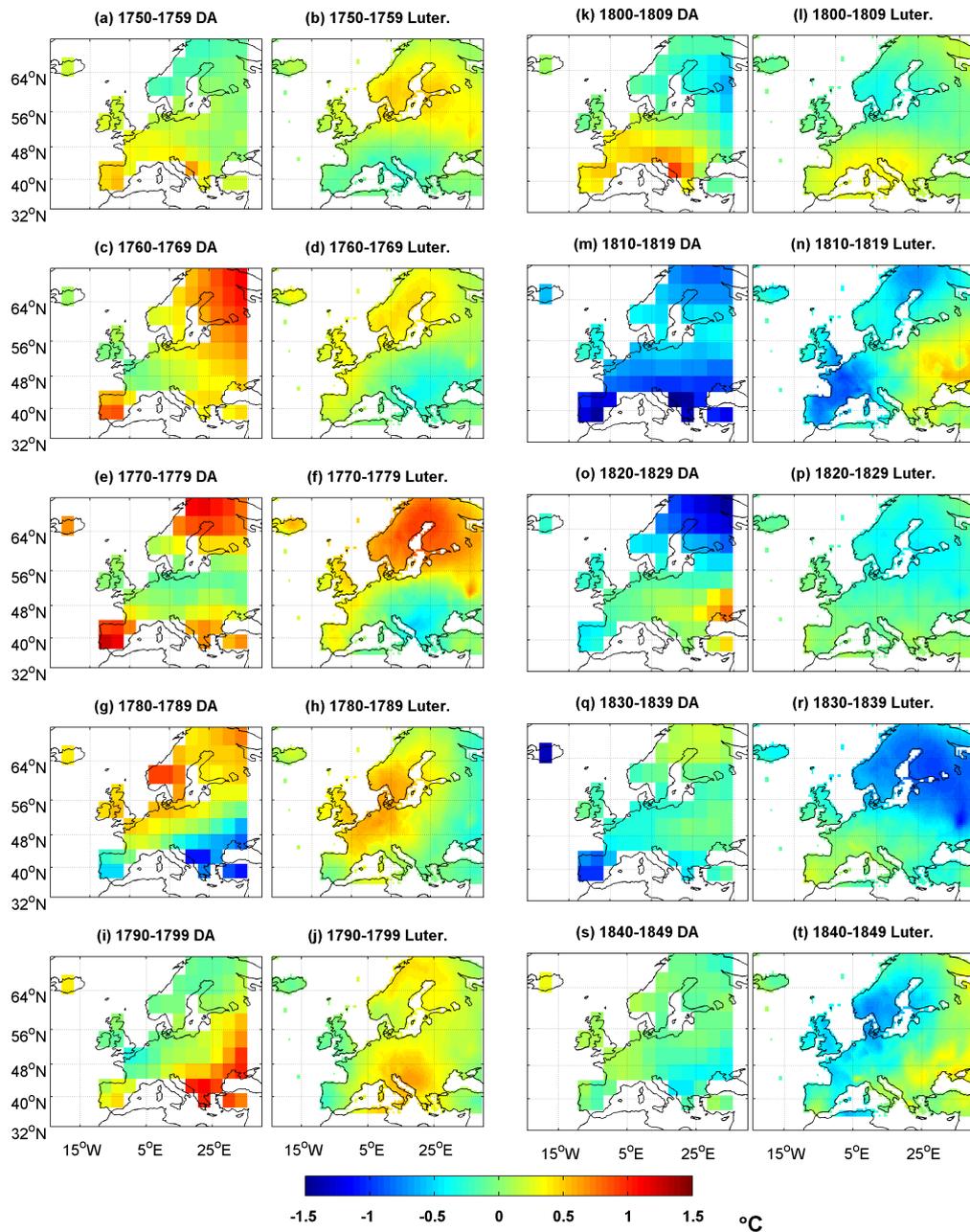
Figure 4.7: European decadal surface air temperatures (anomalies w.r.t. the 1750-1849 AD mean) for the DA analysis and the Luterbacher et al. (2004) reconstruction, for summer 1750-1850 AD.

is skilful in reconstructing the large-scale temperatures, it has no skill in capturing the spatial patterns within Europe correctly. The spatial correlations found for each decade are shown in Table 4.1. The mean correlations over all the decades are negligible (0.03 for summers and -0.03 for winters). We note that regression
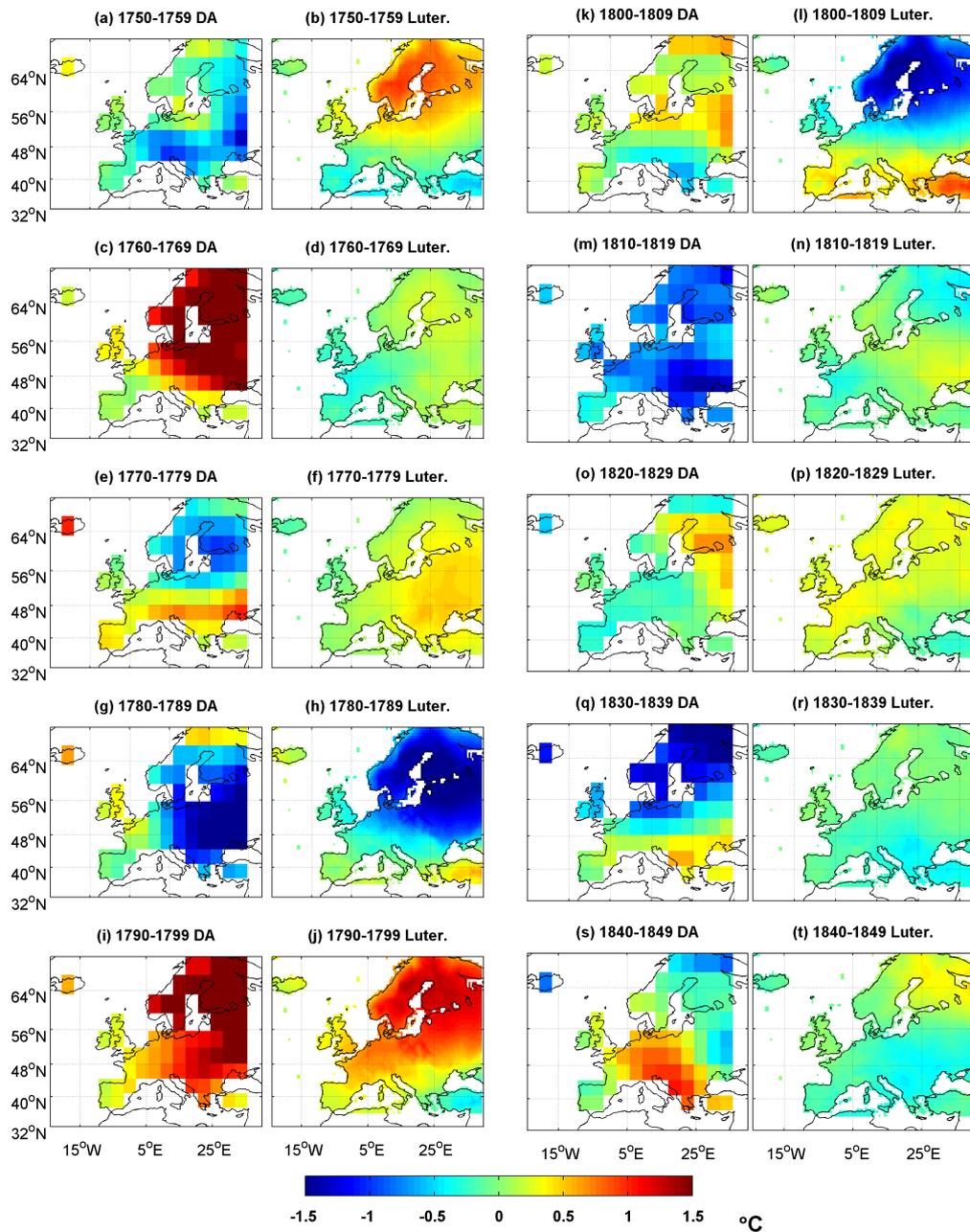
Figure 4.8: European decadal surface air temperatures (anomalies w.r.t. the 1750-1849 AD mean) for the DA analysis and the Luterbacher et al. (2004) reconstruction, for winter 1750-1850 AD.

generally underestimates variability, so the stronger variability appearing in the simulations than in the proxy-based reconstructions is to be expected.

Decade	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Summer	-0.62	0.19	0.21	0.59	0.02	0.70	0.09	0.40	-0.56	-0.70
Winter	-0.05	0.69	-0.04	0.48	0.77	-0.63	-0.31	0.02	-0.70	-0.54

Table 4.1: Spatial correlations between the DA analysis and the Luterbacher et al. (2004) reconstructions for European summers and winters, for the decades 1750-59 AD (1st) to 1840-49 AD (10th).

4.4.2   NH modes of variability

The atmospheric circulation variability in the North Atlantic-European sector is linked to modes of variability such as the NAO or the NAM. The NAO is one of the major modes of atmospheric variability in the NH and strongly influences the wintertime temperature over much of Europe. It is related to the pressure difference between Iceland and Azores, which generates the westerly winds that characterize the atmospheric circulation in the North Atlantic at mid-latitudes. Studies have shown a direct relationship of the NAO with the NH temperatures, particularly in winter (Hurrell, 1995). A positive NAO index corresponds to higher than normal pressure in the Azores High and lower than normal pressure in the Icelandic Low, leading to stronger than average westerlies across the Atlantic basin, which are associated with higher European temperatures. If the pressure is lower than normal in the Azores High and higher than normal in the Icelandic Low, the index is negative and the westerlies are weaker than average. The NAM, which is the dominant mode of the extratropical atmospheric circulation in the NH, is closely related to the NAO. Hence, obtaining the correct phase of the NAO index or other modes of variability with DA might lead to a skilful reconstruction of the small-scale temperatures in the NH.

We compare the decadal mean values of the simulated winter (December to March) NAO index with the available NAO reconstructions, namely the proxy-based reconstruction by Luterbacher et al. (2002) and the University of East Anglia (UEA) Climatic Research Unit instrumental-based reconstruction by Jones et al.

NAO reconstructions
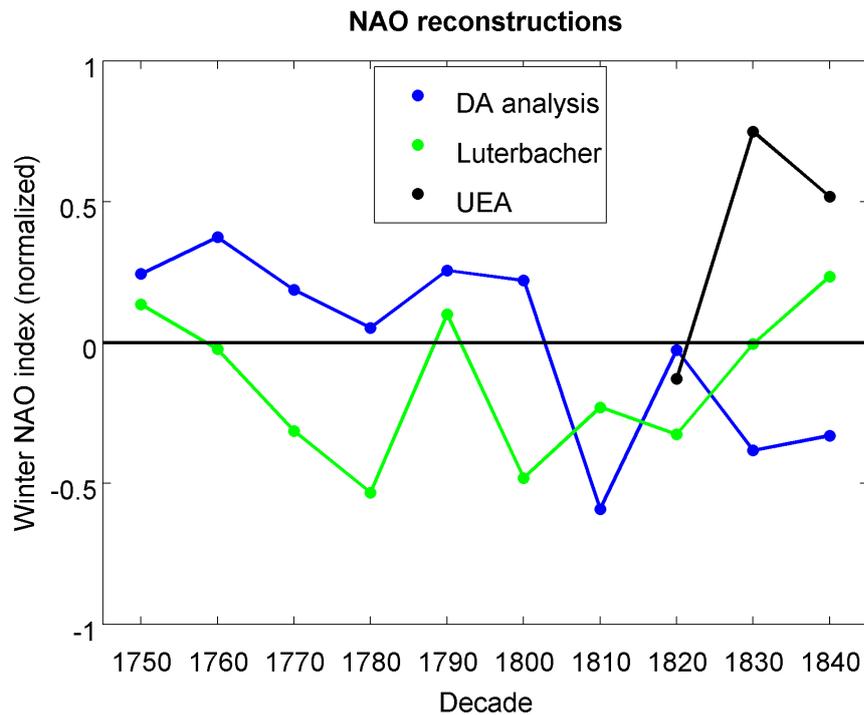
Figure 4.9: Normalized decadal mean winter NAO index for 1750-1850 AD, for the DA analysis (blue dots), the proxy-based reconstruction by Luterbacher et al. (2002) (green dots) and the instrumental reconstruction by Jones et al. (1997) (black dots).

(1997) (Figure 4.9). Luterbacher et al. (2002) estimated the NAO index based on a Principal Component Regression Analysis, as the standardized (1901-1980) difference between the SLP over the Azores and over Iceland. Jones et al. (1997) used early instrumental data to calculate the NAO index as the difference between the normalised SLP over Gibraltar and the normalised SLP over Southwest Iceland, and extended this index back to 1823 AD. Our NAO index (difference between the normalised SLP over Gibraltar and over Southwest Iceland), is not compatible with the two reconstructions, partly explaining the inconsistency between the simulated European temperature patterns and the reconstructions by Luterbacher et al. (2004). The correlation with Luterbacher et al. (2002) is -0.09. The annual NAO index (not shown) does not show any agreement either (correlation 0.21). We did not analyse the summer circulation as the NAO is less pronounced in this season. Zanchettin et al. (2012), using European climate reconstructions, found that the dynamic response to the volcanic eruptions is a decadal-scale positive

phase of the winter NAO, accompanied by winter warming over Europe peaking approximately one decade after a major eruption. There is no consensus among the NAO reconstructions however. Luterbacher et al. (2002) and Jones et al. (1997) reconstructions are also very different to each other, showing the limitation of proxy-based reconstructions, which affects the skill of DA too.

4.4.3 Link between continental temperatures and large-scale circulation

To explain the disagreement between the DA analysis and the proxy-based reconstructions regarding the temperature patterns over Europe and the NAO index, we use MCA to investigate the link between the NH continental mean temperatures and the SLP field. We analyse 1,000 years of the control MPI-ESM-CR simulation to focus on internal variability and to avoid the influence of forcings, because the main purpose of DA is to capture the internal variability. We use simulated decadal temperatures that represent the same continental regions and seasons as the PAGES 2K reconstructions, and NH decadal averages for the SLP field, based on annual and seasonal values. Because the PAGES 2K temperatures are seasonally mixed, it is not clear in which seasons the link of the seasonally mixed temperatures to the atmospheric flow is strongest, therefore we separate our analysis for the different seasons and the annual means. In all five cases the temperature data are the same.

MCA applies a singular value decomposition (SVD) to the cross-covariance matrix between two datasets to find pairs of patterns whose time expansion coefficients (TECs) have maximum covariance (Bretherton et al., 1992). In our case, the dimension of the cross-covariance matrix is 4×2304 (number of continental temperature means $\times$ number of grid-cells for the SLP field). To account for

the different size of the grid-cells, the SLP anomalies field (w.r.t. to the 1,000-year control average), SLP_i, has been weighed with weights w_i proportional to the square root of the grid-cell size, prior to calculating the cross-covariance matrix. The weights have also been used for calculating the TECs. For the continental temperature anomalies, T_i, no area weights have been used. The components of the cross-covariance matrix, C_{ij}, are given by the equation:

$$C_{ij} = \frac{1}{t-1} \sum_{k=1}^{t} \left(T_i(t_k) - \overline{T_i}\right)\left(SLP_j(t_k) - \overline{SLP_j}\right) \tag{4.1}$$

The temperature and SLP singular vectors, which are the MCA coupled patterns, are denoted by the vectors u_k and v_k respectively. The TECs (a_k and b_k) are given through area-weighted orthogonal projections of the data anomalies (T_i and SLP_i) onto the patterns, such that:

$$a_k(t_j) = \sum_{i=1}^{4} T_i(t_j)\, u_{ik} \tag{4.2}$$

$$b_k(t_j) = \sum_{i=1}^{2304} w_i\, SLP_i(t_j)\, v_{ik} \tag{4.3}$$

A key property of the SVD is that the patterns are constructed in a way that the covariance of the TECs is maximized, and the patterns are orthogonal to each other. The singular vectors are normalized and non-dimensional, while the TECs have the dimensions of the original data. The subsequent pairs of patterns are orthogonal to the previous ones. The first coupled pattern is the dominant one, as it explains the largest fraction of the covariance between the two datasets.

Figures 4.10 and 4.11 show the first coupled MCA patterns of the NH continental mean temperatures and the NH SLP respectively. The patterns are displayed as the singular vectors multiplied by the standard deviation of their TECs in order

to include information about their amplitude. For the annual means, winter and to a lesser extent for spring, the SLP MCA patterns look similar to the leading empirical orthogonal function (EOF) of the NH SLP field in the control run, shown in Figure 4.12. For summer and autumn the SLP MCA patterns are different from the leading EOF. Looking at the EOF structures, in all seasons except summer, the SLP leading EOF patterns resemble the annular mode patterns. This mode is basically the NAM, but it should be noticed that our EOF analysis is based on decadal SLP anomalies in the region 0N to 90N, whereas the standard definition of the NAM is based on the first principal component of monthly SLP anomalies in the region 20N to 90N. In summer, the SLP leading EOF shows a pronounced land-sea contrast, which is presumably linked to the intensity of the land-sea temperature contrast. From the above, we can conclude that for the annual means, winter and to a lesser extent for spring, the NAM is the pattern that is most closely linked to the NH continental temperatures. Hence, by assimilating continental temperatures, the state of the NAM should be determined to some extent. For summer and autumn, the MCA patterns do not resemble the NAM but have a wavelike structure. The correlations between the temperature and SLP TECs are 0.73 for the annual case, 0.48 for winter, 0.51 for spring, 0.36 for summer and 0.54 for autumn. Given the correct continental mean temperatures, the explained variance for the SLP TECs is 0.73^2 for the annual means, i.e. 52%, and much less for the seasonal means. In our case the estimate for the SLP TECs will actually explain less than 52% variance because the PAGES 2K reconstructions are different from reality, as they include noise. We also analysed higher coupled patterns. The correlations between their TECs were lower, and thus we focus on the first pair of coupled patterns.

To test whether the correlations between the two sets of TECs are significant, we applied a Monte Carlo method to calculate the distribution of correlations in the case of no link. We created a 1,000-year random temperature array for the four continents, 300 times, and performed a MCA analysis of the link between these
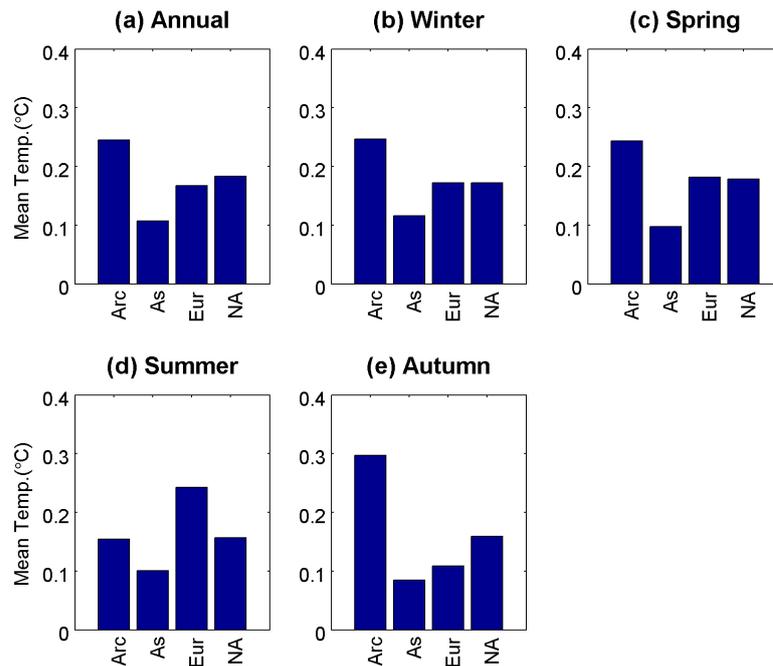
Figure 4.10: MCA between NH continental mean temperatures (seasons defined by PAGES 2K) and NH SLP (annual and seasonal means) from the MPI-ESM-CR control simulation: first temperature patterns for the Arctic, Asia, Europe and North America.

random temperatures and the control run's SLP field. In the annual case, the mean TEC correlation of the randomly sampled temperatures with the SLP field is 0.39, ranging between 0.25 and 0.55 (Figure 4.13a). These correlations are lower than the value of 0.73 of the case in which the simulated temperatures were used for the MCA. Similar results are found for all four seasons (Figure 4.13b-e), but the largest difference is noticed for the annual means. In all cases, the correlations of the simulated control run temperature TECs were either outside or at the very end of the randomly sampled distribution. The test shows that the link found between the NH continental mean temperatures and the NAM in the control run is significant, and thus the assimilation of continental mean temperatures gives the potential of small spatial-scale skill, although this was not achieved in the assimilation run. The same Monte Carlo experiment was performed for higher coupled patterns, the TEC correlations of which remain relatively significant but are gradually reducing.
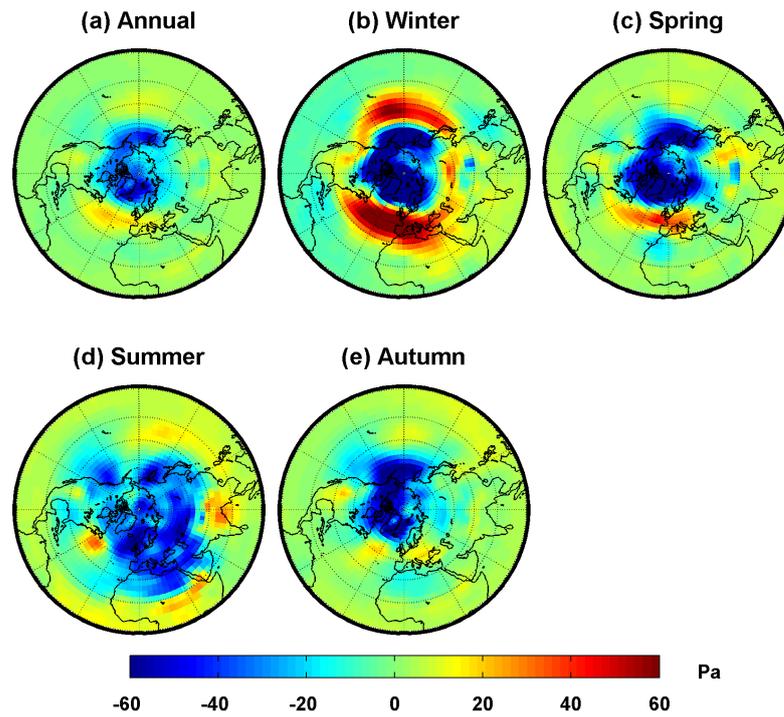
Figure 4.11: MCA between NH continental mean temperatures (seasons defined by PAGES 2K) and NH SLP (annual and seasonal means) from the MPI-ESM-CR control simulation: first SLP patterns.
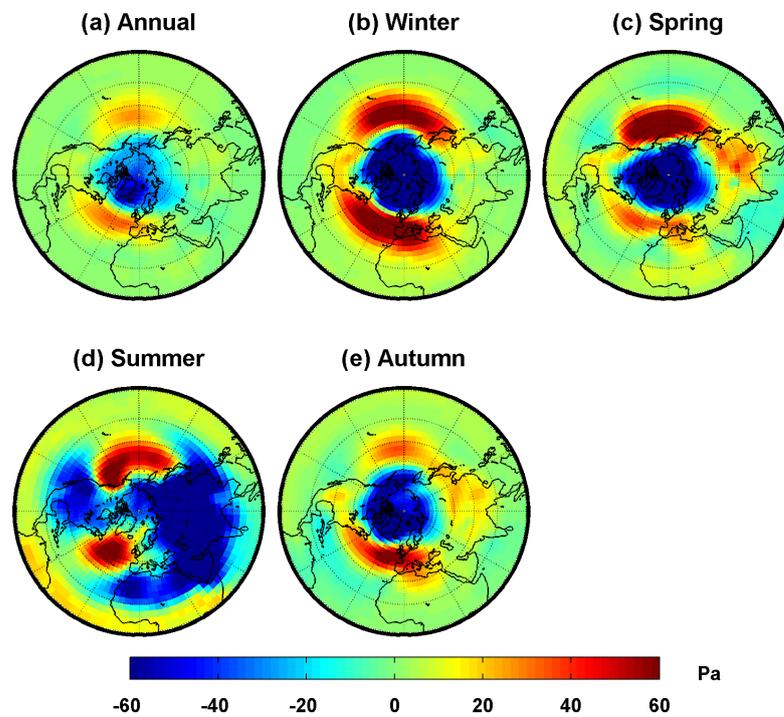
Figure 4.12: Leading EOF of the NH SLP decadal means in the MPI-ESM-CR control simulation, for the different seasons.
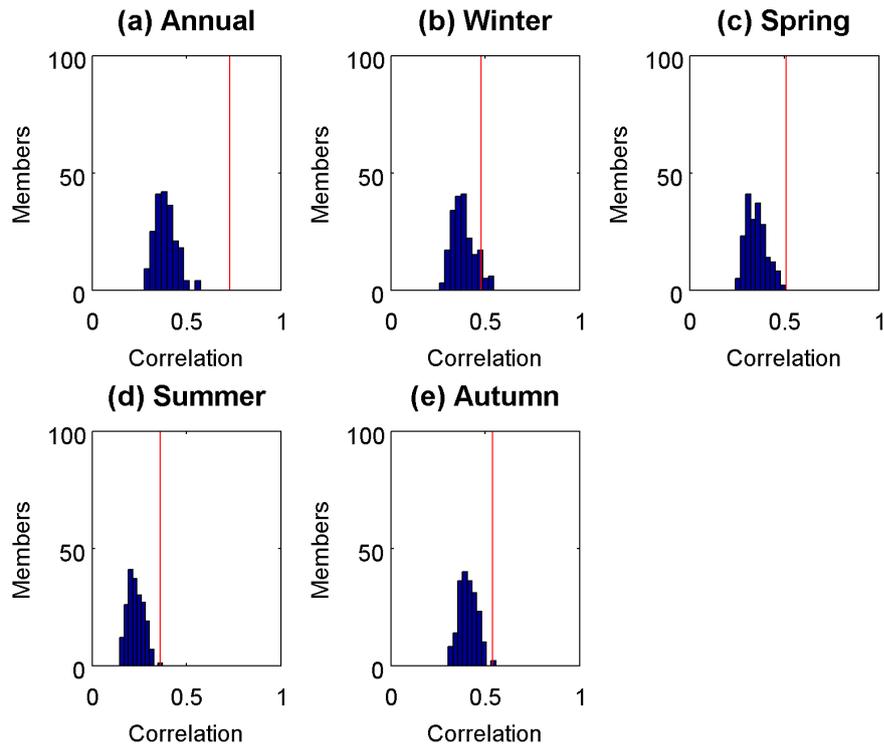
Figure 4.13: Monte Carlo experiment for the distribution of correlations in the case of no link between temperature and SLP, for the annual and seasonal means. Red vertical lines show the cases in which the simulated temperatures were used in the MCA

4.4.4 NH circulation link in the assimilation run

We now investigate whether the link between the NH continental average temperatures and the SLP field that has been found in the control run is also visible when performing the DA. The link between temperatures and SLP can be caused by two processes: i) for a given SLP anomaly the temperature is influenced through advection; and ii) given temperature anomalies might directly influence the pressure field, for instance through changes in the thickness of atmospheric layers and the associated upper-level flow anomalies which in turn affect the surface pressure. The former process can be expected to be dominant. Independent of the type of process that is relevant, the DA may reproduce the link because it is just selecting states rather than prescribing temperature anomalies, and the model captures both

process types.

To examine whether the DA actually reproduces the link, we project the DA decadal NH temperature and SLP anomalies of the period 1750-1850 AD onto the control experiment's temperature (u_k) and SLP (v_k) MCA patterns respectively, to get temperature and SLP TECs for the assimilation period, and calculate the correlation of the TECs. The anomalies were calculated w.r.t. the means of the control run to be consistent with the previous analysis. The correlations are 0.81 for the annual means, 0.82 for winter and spring, 0.17 for summer and -0.01 for autumn. After detrending the TEC time series, the correlations were only slightly reduced (0.70 for the annual means, 0.73 for winter, 0.70 for spring, 0.02 for summer and -0.21 for autumn). The fact that the difference between detrended and non-detrended correlations is very small shows that the correlations are mainly due to the decadal variations. When the simulated temperatures in the analysis are replaced by the PAGES 2K temperatures, which have been assimilated, very similar correlations are obtained. The correlations of the TECs in this case are 0.82 for the annual means, 0.81 for winter, 0.79 for spring, 0.12 for summer and 0.02 for autumn. This can be expected because, as shown earlier, the continental mean temperatures in the analysis follow closely the assimilated temperatures. The above analysis shows that the link between the NH continental mean temperatures and the SLP field in the assimilation run exists in winter, spring and the annual means, but not in summer and autumn.

A question that arises from the analysis is why the link that we found between temperatures and SLP in the control run is not reproduced in summer and autumn in the assimilation run. To investigate these seasonal differences, we take a closer look at the SLP MCA patterns (Figure 4.11). We note that the link in the assimilation run appears in the seasons when the SLP MCA patterns are similar to the leading SLP EOFs, which in turn resemble in these seasons the NAM. The

link does not appear when the MCA pattern does not look like the leading EOF. This observation hints to a potential explanation; it seems plausible that it is easier in a small ensemble to find analogues for circulation and temperature anomalies that resemble the dominant mode of variability than finding good analogues for anomalies that are different from the leading EOFs. The reason for this might be that the simulated anomalies have a tendency to look like the leading EOF patterns or linear combination of these, while they are usually not similar to patterns that are not in the space of the leading EOFs. Moreover, the link in summer is weaker than in the other seasons, as the TECs correlations reveal.

The MCA thus shows that the NH continental mean temperatures for the PAGES 2K regions are most closely linked to the NAM in winter, spring and the annual means, and that this link is reproduced in the DA for the same seasons. However, this does not imply that the continental temperatures are the best predictors for the NAM. To investigate this further, we have regressed the control run's grid-point NH temperatures onto the first SLP TECs. Linear regression between the temperature field and the SLP patterns demonstrates the linear relation between the strength of the SLP MCA pattern and the local temperatures. As shown in Widmann (2005), these regression coefficients are proportional to the weights for the temperature field that lead to an optimal linear estimation of the SLP MCA TEC in a one-dimensional MCA. Therefore, temperature information from areas with high values are needed for obtaining good estimates for the SLP MCA TEC. Figure 4.14 shows the regression maps of the control run NH temperatures onto the first SLP MCA TECs, multiplied by one standard deviation of the SLP TECs. In this case the map demonstrates how much the temperature would change when the SLP TECs change by one standard deviation. The green areas in the figure denote areas where no significant temperature change is associated with a change in the NAM. The regression maps show that this TEC, which is similar to the NAM index for annual data, winter and spring, has a strong temperature signal

only at the very high latitudes, e.g. Scandinavia for the case of Europe. Similarly, the temperature signal of the TEC is larger at the northern parts of North America and Asia. The regression maps for the second and third SLP TECs were also checked and corroborate that the first couple pattern explains the largest fraction of variability, as the temperature signal of the higher SLP TECs is substantially weaker.
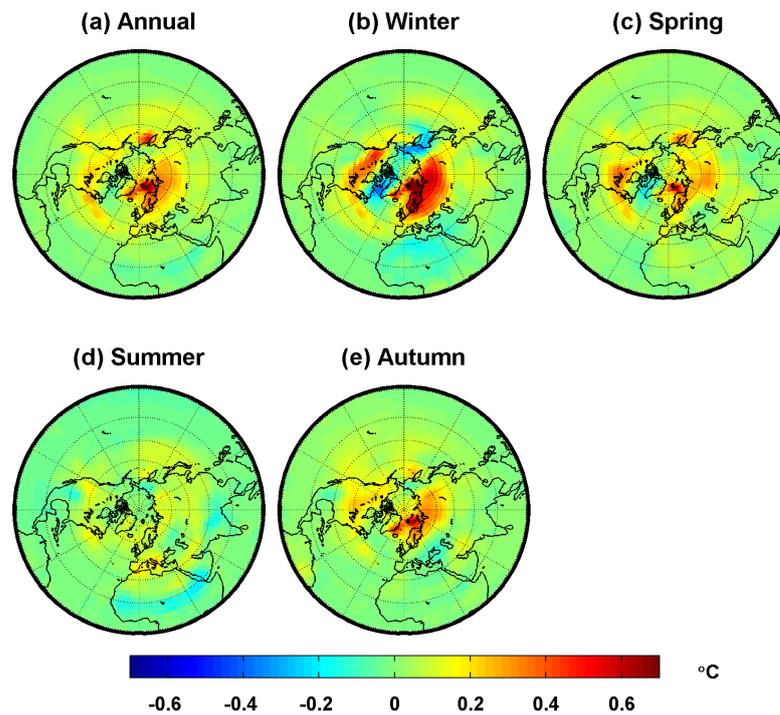
Figure 4.14: Regression maps of the MPI-ESM-CR control simulation's grid-point NH temperatures onto the SLP MCA TECs, multiplied by one standard deviation of the SLP MCA TECs, for the different seasons.

The correlation maps between the control run's local NH temperatures and the leading SLP MCA TECs for the different seasons (Figure 4.15), show that only at the very high latitudes, the SLP TECs have relatively high correlations (up to 0.7) with the local temperatures. The correlations at the mid-latitudes range from 0.2-0.4 for the annual means, and even lower values for the seasonal means. This means that a large amount of the local temperature variance cannot be explained by the leading SLP TEC. For instance, only 49% of the local temperature variance can be explained at the very high latitudes given a perfect estimation of the SLP

TEC, which is similar to the NAM index in the annual case, and almost no variance at the low and mid latitudes. This designates that even if the true amplitude of the leading MCA circulation pattern were simulated, the local temperatures would still not be strongly confined in many areas. It can further be noticed that the local temperature correlation and regression coefficients over Europe have the same sign everywhere, which means that the variability in the SLP TEC only leads to limited variability in temperature gradients across Europe.
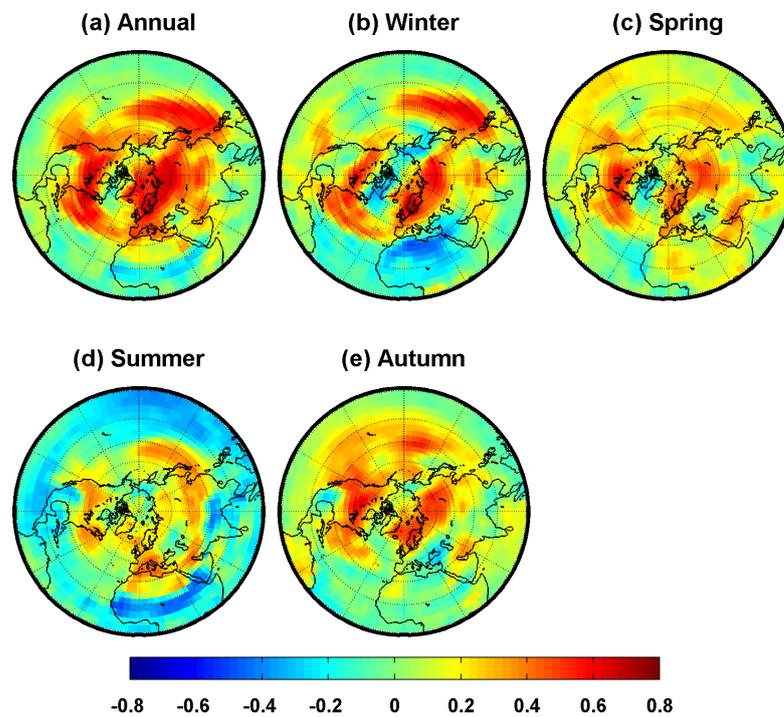
Figure 4.15: Correlation maps between the MPI-ESM-CR control simulation's grid-point NH temperatures and the SLP MCA TECs, for the different seasons.

To summarise, the continental temperatures for the NH PAGES 2K regions are not optimal for confining the amplitude of the leading MCA SLP pattern, which for annual values, winter and spring are similar to the NAM. The reason is that the temperature signals of the SLP MCA patterns are mostly concentrated in high latitudes, hence the continental mean temperatures include information from regions not related to these circulation anomalies. This limits the correlation between the temperature TECs and SLP TECs. In DA this can be expected to lead

to circulation states that are not strongly constrained by the assimilated continental temperatures, and in turn to a limited potential for realistically simulating small-scale temperature variability. Moreover, for summer and autumn the link between NH continental temperatures and the SLP field identified by MCA is not reproduced in the DA. This might be due to the fact that the leading SLP MCA pattern is not similar to the leading SLP EOF in these seasons, which in turn may lead to sampling problems if a small ensemble size is used. We have also shown that even if the amplitudes (i.e. the TECs) of the MCA SLP pattern were perfectly captured in the DA the local explained temperature variance would be limited. Therefore, the discrepancies of the assimilation output with the European temperature patterns of Luterbacher et al. (2004) and the NAO reconstructions of Luterbacher et al. (2002) are not surprising.

4.5   Summary and conclusions

The PAGES 2K NH continental decadal mean temperatures have been assimilated using the MPI-ESM model to get a climate reconstruction for 1750-1850 AD. The skill of the DA-based reconstruction was firstly evaluated against the assimilated data and other proxy-based temperature reconstructions on large spatial scales. We also validated the DA analysis for smaller-scale temperatures for Europe and for the NAO, to see whether there is added value from the DA. The assimilation showed good skill for large-scale temperatures. With respect to the NH continental means, the DA analysis followed closely the assimilated proxy-based reconstructions on the decadal timescale. Skill was also found on shorter timescales due to the strong forcings, especially the major volcanic eruptions. The DA was shown to provide added value on the continental scale, especially on the decadal, but also on shorter timescales, when compared with simulations without DA. The DA analysis was in agreement with independent proxy data and early instrumental

records for global and hemispheric-scale temperatures too, but no additional skill of the assimilation was found due to the dominance of forcings on these scales. Although no information about the smaller-scale spatial temperature anomalies was assimilated, the assimilation of the NH continental average temperatures might lead to some skill in simulating small-scale spatial temperature patterns because of the potential capturing of information that may determine the state of leading circulation modes, such as the NAM or the NAO. However, no agreement between the DA analysis and the proxy-based European temperature reconstruction of Luterbacher et al. (2004) or the NAO index reconstruction of Luterbacher et al. (2002) was found.

To examine the reasons why the DA does not provide added value on smaller spatial scales, we calculated the link between NH continental temperatures and the NH SLP field using MCA and decadal means from the MPI-ESM control simulation. The temperature data for the different continents represented different seasons or annual data in accordance with the seasons that had been used for the PAGES 2K reconstructions, while the SLP data were taken for each of the seasons and for annual data, and the MCA was performed separately for the five cases. The analysis showed that for annual, winter and spring SLP, the circulation pattern that is most closely linked to the NH continental mean temperatures strongly resembles the NAM, while for summer and autumn it is a wave-like pattern. The correlations between the temperature and SLP TECs are 0.73 for the annual data, 0.48 for winter, 0.51 for spring, 0.36 for summer and 0.54 for autumn. These correlations indicate potential for constraining to some extent the large-scale circulation if the continental temperature means are known, which in turn might lead to some skill for small-scale temperature variability in the DA simulations. However, the validation results showed that this potential was not achieved.

To further investigate the lack of small-scale skill in the DA simulation we also

checked whether the link found in the control run is visible in the DA simulations and found this to be the case for annual, winter and spring SLP, but not for the other seasons. This is possibly due to the fact that for the former, the NH continental temperatures are linked to SLP anomalies that resemble the first SLP EOF (which in these seasons is similar to the NAM), which might be better captured in DA with small ensemble sizes than variability patterns that are less dominant. It was also shown that even if the amplitudes of the leading MCA SLP pattern were perfectly known, the local explained temperature variance would still be less than 50% in the best case and much lower in many areas.

In summary, the analysis suggests the following potential reasons for a lack of skill of the DA in simulating small-scale temperature variability: i) the link between NH continental temperatures and large-scale atmospheric circulation anomalies might be too weak to sufficiently constrain the large-scale circulation; ii) the small ensemble size might make it difficult to find ensemble members that resemble the real circulation state that led to the PAGES 2K continental temperature anomalies; iii) noise and errors in the PAGES 2K temperature reconstructions contribute to unrealistic large-scale circulation states in the DA simulations; and iv) the local temperature variance explained by the large-scale circulation anomalies that can be estimated from the NH continental temperatures is substantially limited.

These potential contributing factors to a lack of small-scale skill give some guidance for improvements of the DA method. Firstly, the link between assimilated proxy-based temperature reconstructions and the large-scale circulation can in general be expected to get stronger if local or regional rather than continental-scale reconstructions are used, provided they cover suitable areas. However, it should be noted that with decreasing spatial scale of the reconstructions their error can be expected to increase, as there is less cancellation of random errors. Determining the optimal scale for the temperature data to be assimilated is thus a challenge.

In this context we also point out that the location of most of the proxy records that have been used for the PAGES 2K reconstructions are not in the high-latitude regions that have the strongest NAM signal, and thus the reconstructed continental temperature variability may have a disproportionally low contribution from these areas. For example, the European PAGES 2K mean temperature is reconstructed from proxies mostly in the southern and central parts of the continent, which are not sensitive to NAM variations, and fewer proxies in Scandinavia and Northern Europe. Using temperature reconstructions that provide a stronger constraint for the atmospheric circulation may also lead to constraining several modes of circulation variability and thus increase the local explained temperature variance. Secondly an increase in ensemble size will make it more likely to find simulated climate anomalies that are similar to the real world. The required sample size to have a good chance for finding close analogues depends on the dimensionality of the state space in which the analogues are defined, with the required sample size increasing very strongly with the dimensionality. This is an argument for using low-dimensional state spaces, such as the four continental means we have used, rather than temperature reconstructions with a high spatial resolution. However, other methods for reducing dimensionality such as principal component analysis might be an alternative to using continental averages and might lead to a stronger link with the atmospheric circulation. Finally, retaining more than one ensemble member, i.e. using a non-degenerate particle filter as in Goosse et al. (2012), might improve the performance of the DA.

Chapter 5

Influence of proxy data uncertainty on DA

5.1 Introduction

As a result of the fact that at small spatial scales random internal variability plays a dominant role compared to forced climate variability, good skill at these scales cannot be obtained in a forced, free running simulation. In order to realistically simulate regional-scale variability, some main modes of atmospheric circulation need to be constrained by DA. The study of the previous chapter, hereafter referred to as Matsikaris et al. (2015b), has shown that although DA simulations generally follow the large-scale target temperatures well, there is a lack of information propagation to smaller spatial scales. A potential reason for this lack of skill was that continental temperatures might not be ideal for constraining the atmospheric circulation and thus regional temperatures. Two main sources of error are present in this DA setup. The first one is the methodology error, caused by: the use of continental-scale, decadal mean temperatures; limited ensemble size; a RMS error-based cost function; and the selection of only the best member. The second one is the data

uncertainty error, i.e. differences between reconstructed and true temperatures in the assimilated data. As proxy-based temperatures, which may have substantial errors, have been assimilated in the Matsikaris et al. (2015b) study, it is an open question whether the lack of added value in small spatial scales is mainly due to the method or also to the errors in the assimilated data.

A fundamental problem in the validation of DA simulations for the past is that the true climate is unknown. This problem can in principle be addressed by simulations for the instrumental period, which allow evaluating the assimilation output against direct observations. In this chapter, to distinguish between methodology and data error in the DA setup of Matsikaris et al. (2015b), we assimilate instrumental observations (Brohan et al., 2006) and proxy-based reconstructions (PAGES 2K Consortium, 2013) for the period 1850-1949 AD using two different DA schemes. The two sets of DA simulations are referred to as DA-I (for the instrumental-based scheme) and DA-P (for the proxy-based scheme) respectively. Assuming that the instrumental observations constitute perfect input data for DA, the errors can be separated: the error in the DA-I scheme is mainly the methodology error, while the difference in the errors of the two schemes is the consequence of the errors in the proxy-based temperature reconstructions. Error separation is important, as it can lead to improvements of the method. The approach remains however challenging, as the global observational coverage is very incomplete early in the record, leading to an additional source of sampling uncertainty when estimating regional averages, which affects DA. The coverage is not complete even in the more recent years that have the most observations (Brohan et al., 2006), hence the gridded instrumental data cannot be considered perfect.

The aim of this chapter is twofold: firstly, to evaluate the DA scheme of Matsikaris et al. (2015b) against gridded instrumental observations; and secondly, to investigate the effect of the proxy error on the performance and the lack of

small spatial-scale skill in the Matsikaris et al. (2015b) setup, and examine the consequences of unrealistic targets for DA. We employ again on-line DA using the MPI-ESM model, and simulate 20 ensemble members with each DA scheme. The member selection is performed based on decadal mean temperatures for the NH continents. The structure of the chapter is as follows: Section 5.2 presents a brief overview of the simulations and data used here, together with the experimental design. The performance of the two schemes is evaluated against the target and the observed temperatures for large spatial scales in Section 5.3, and the obtained cost functions are compared. The skill of the two DA schemes with respect to the observed regional temperature variability is also examined. Finally, Section 5.4 provides a discussion of the results and summarizes the main findings.

5.2   Model, data and method

The ensemble simulations analysed here were conducted with the coarse-resolution version of the MPI-ESM coupled model, described in Section 2.2. Two sets of DA simulations are conducted, selecting the best among 20 ensemble members, in the first case based upon comparison with the PAGES 2K proxy-based reconstructions (DA-P scheme), and in the second case upon comparison with the HadCRUT3v instrumental observations (DA-I scheme). The PAGES 2K proxy-based dataset (PAGES 2K Consortium, 2013) has been described in Section 3.2.2. Only the NH reconstructions have been assimilated, which include annual mean temperatures for the Arctic (60-90N, 180W-180E), summer means (JJA) for Asia (23.5-55N, 60-160E) and Europe (35-70N, 10W-40E) and decadal means representing all seasons for North America (30-55N, 130-75W).

The HadCRUT3v historical surface temperature dataset was developed by the Climatic Research Unit of the University of East Anglia, in conjunction with the

Hadley Centre of the UK Met Office. It provides monthly instrumental surface temperature observations since 1850 AD interpolated on a 5°×5° grid (Brohan et al., 2006), as well as a comprehensive set of uncertainty estimates. These include measurement and bias errors, as well as sampling errors due to the estimation of a grid box mean from a small number of point values (Brohan et al., 2006). The uncertainties are larger in the nineteenth and early twentieth century and small in the more recent period, however, the temperature increase over the century is significantly larger than its uncertainty. In the assimilation of the HadCRUT3v temperatures, the same continental-scale regions and seasons as the ones defined by the PAGES 2K groups have been used.

The implementation of the method is the same as the one described in Section 4.2 (Matsikaris et al., 2015b). The initial condition of the two ensembles was taken as the last day of 1849 AD from a transient forced simulation starting in 850 AD. Twenty ensemble members were generated in each scheme by introducing small perturbations in an atmospheric diffusion parameter, and after 10 years of simulations, a RMS error-based cost function was used to compare the simulated decadal mean temperatures of the NH continents with the PAGES 2K and HadCRUT3v reconstructions respectively. The member that minimized the cost function was selected as the "analysis" for that period and was used as the initial condition for the subsequent 10-year simulation. The procedure was repeated sequentially until 1950 AD. Before being used in the cost function, the simulations, proxy-based reconstructions and instrumental data were all standardized in order to remove biases in the mean and variance, and to give equal weight to each continent regardless of the variance of the respective temperature time series. The standardization in the DA-P scheme used the long-term 850-1849 AD mean and variance, while for the DA-I scheme the reference period was 1850-1949 AD. The

cost function in the two DA schemes is:

$$CF(t) = \sqrt{\sum_{i=1}^{4} \left(T_{mod}^i(t) - T_{tar}^i(t) \right)^2} \tag{5.1}$$

where i denotes the NH continents, $T_{mod}^i(t)$ is the standardized modelled decadal mean of the temperatures in each continent and $T_{tar}^i(t)$ is the respective standardized target mean temperature (proxy-based in the DA-P scheme and instrumental-based in the DA-I scheme).

5.3   Results

We address two questions when evaluating the performance of the two DA schemes: i) how well the DA analyses follow the assimilated target; and ii) how well they follow the reality as given in the instrumental data set. The study employs two different metrics. The first one, correlation, measures the similarity of the temperature variations regardless of biases or scaling differences. Significant positive correlations are an indication of a common response to changes in external forcings or a similar realisation of internal variability. The second metric, RMS error, tests how close the DA analysis and the reconstructed or observed temperatures are. It is worth mentioning that the PAGES 2K reconstructions are calibrated against the instrumental data sets. For "composite plus scale" methods, the temporal evolution of the reconstruction is given by the proxy data composite, regardless of the calibration record, which is only used for scaling. The situation is different for regression-based methods, where the parameter fitting leads to reconstructions that have a temporal behaviour that is more similar to the instrumental continental temperature series than a simple composite. As the number and quality of proxy records decreases when going back in time, the discrepancies between the real climate and the proxy-based reconstructions can be expected to increase, with the

consequence of less realistic teleconnections for the reconstructions. The increase in reconstruction errors for periods further in the past might be larger for regression-based methods in cases where overfitting or temporal instabilities in the regression relationships occur. The skill found in our analysis for the DA-P scheme is thus an upper limit for the skill for periods before 1850 AD.

5.3.1 Consistency with the assimilated target data

For each DA scheme, the simulated NH continental temperatures are first compared with the assimilated target data, i.e. the PAGES 2K proxy-based reconstructions for the case of the DA-P scheme and the HadCRUT3v instrumental data for the case of the DA-I scheme. This evaluates the coherence of the DA analyses with the assimilated target. We compare the consistency for the NH continents and the NH mean using both the standardized time series, as this is the form of the data included in the cost function, and the raw (non-standardized) temperature anomalies, which is the standard way of presenting simulation output. The comparison is based on decadal means, which is the timescale used in the assimilation. Agreement on smaller timescales can possibly arise due to the common response of simulations and reconstructions (or observations) to the forcings, but it cannot stem directly from the assimilation (apart from some correlation inherited from the decadal timescale).

Figure 5.1 compares the NH continental decadal mean temperature anomalies for the DA-P analysis with the PAGES 2K reconstructions (anomalies w.r.t. the 850-1850 AD mean), as well as the DA-I analysis with the HadCRUT3v reconstruction (anomalies w.r.t. the 1850-1949 AD mean). The DA-P analysis follows the temperature variations of the assimilated proxy-based reconstructions very closely, showing that the DA method is skilful. The correlations between the DA-P analysis and the PAGES 2K reconstructions are 0.93 for the Arctic, 0.86 for Asia, 0.85 for

Europe and 0.91 for North America. The DA-I analysis and the instrumental data, HadCRUT3v, are also in good agreement, having significant positive correlations on the decadal timescale for the Arctic (0.96), Europe (0.85) and North America (0.88). They are however strongly inconsistent for Asia (correlation is 0.24). The same plots but using the standardized time series are shown in Figure 5.2. As correlations are independent of scaling, they are the same for the standardized and the raw temperature anomalies.

Based on the decadal values, the standard deviations of the DA-P analysis for the Arctic, Asia, Europe and North America are 0.47, 0.20, 0.30 and 0.27 respectively. Accordingly, the standard deviations of the DA-I analysis are 0.28, 0.12, 0.19 and 0.13. The smaller variance found for the DA-I analysis cannot be attributed to a difference in the variance of the input data, as the DA analysis does not follow the amplitudes of the assimilated data due to the standardization of the time series in the cost function. A cost function that gives each continent in the assimilated data weights based on the variance of the temperature variations rather than using standardized time series would allow more robust conclusions regarding the effect of the input data variance on the DA scheme. From the current study, it cannot be concluded why the DA-I analysis has a smaller temperature variance than the DA-P one.

The skill of the assimilation on the inter-annual timescale (five-year running means) with the two DA schemes is shown in Figure 5.3. The consistency of the DA-P analysis with the assimilated data is moderate (0.71 correlation for the Arctic, 0.61 for Asia, and 0.71 for Europe), mostly owing as mentioned previously to the response to the forcings. North American data were not evaluated on this timescale due to their decadal resolution in the PAGES 2K reconstructions. Similarly, moderate correlations are found for the five-year running mean time series of the DA-I analysis for the Arctic (0.74) and Europe (0.54), unlike the case of Asia (0.20), where data
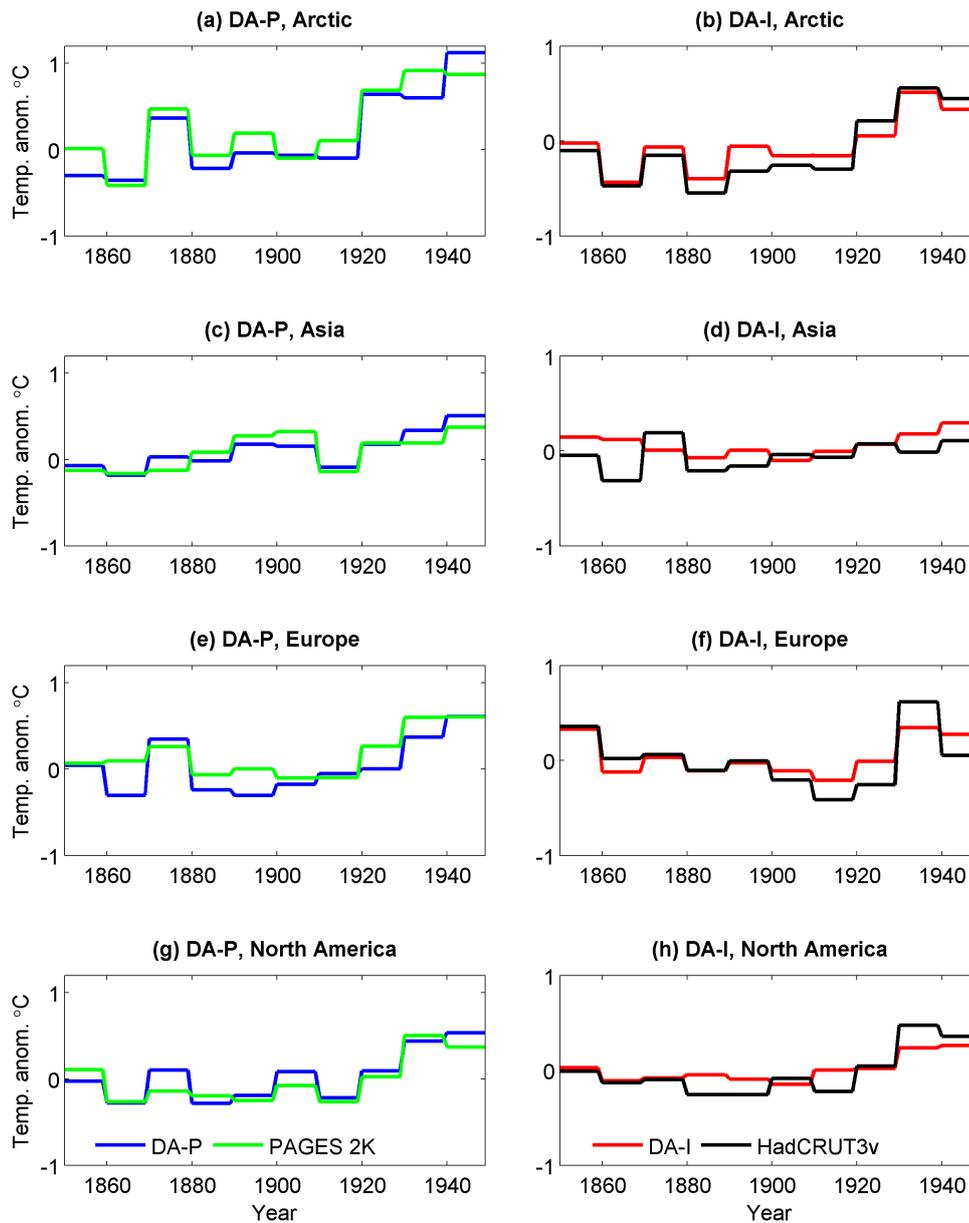
Figure 5.1: Left column: Continental decadal mean temperature anomalies in the NH for the DA-P analysis (blue) and the proxy-based reconstructions (green). The reference period is 850-1849 AD. Right column: Continental decadal mean temperature anomalies in the NH for the DA-I analysis (red) and the instrumental data (black). The reference period is 1850-1949 AD.

and DA-simulation disagree strongly.

The RMS errors are calculated from the decadal mean differences between the DA analyses and the target time series for each continent. With the DA-P scheme, the RMS errors of the non-standardized time series are 0.20 for the Arctic, 0.11 for
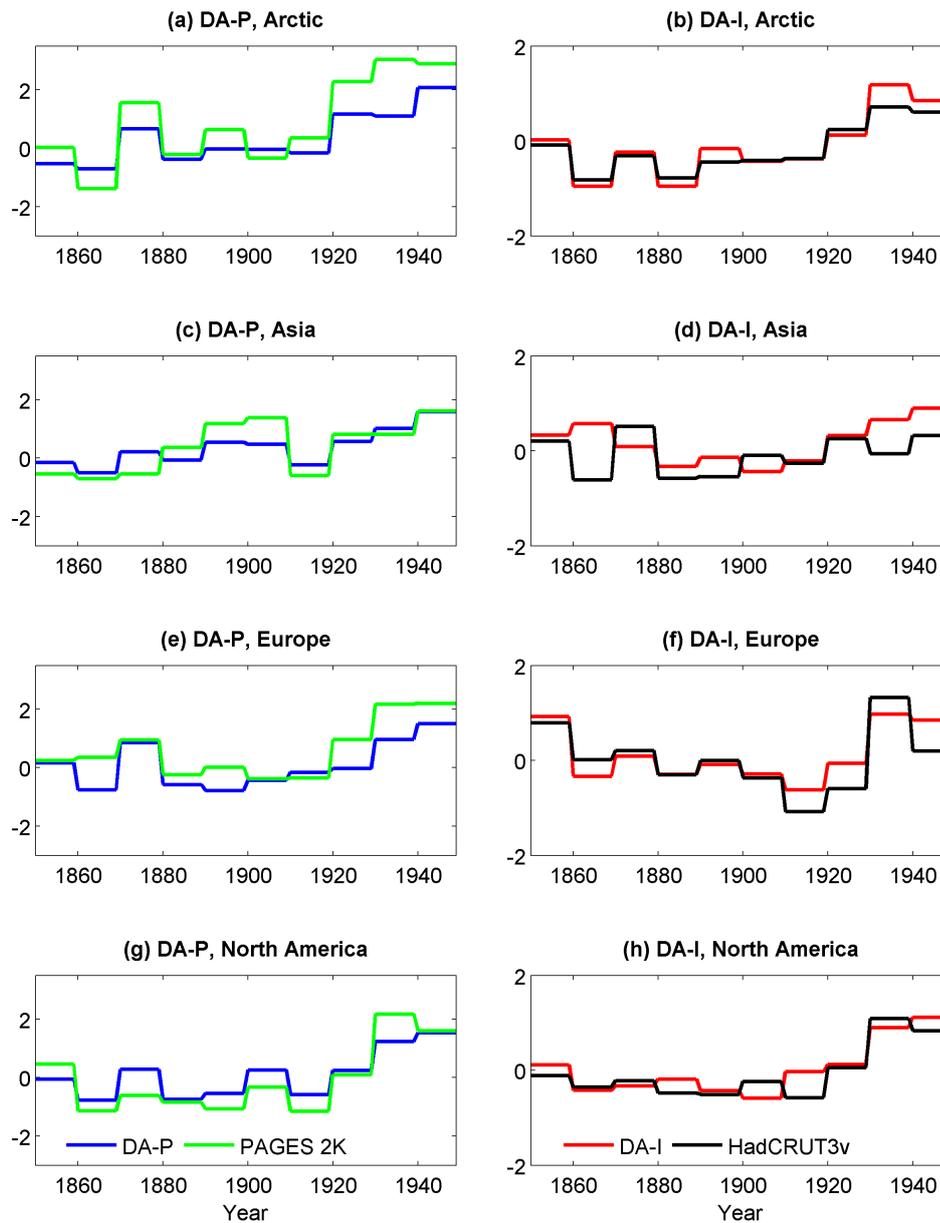
Figure 5.2: Consistency of the DA analyses with the assimilated data for 1850-1949 AD (as Figure 5.1) for the standardized time series.

Asia, 0.21 for Europe and 0.19 for North America, while with the DA-I scheme, they are 0.13, 0.20, 0.16 and 0.14 respectively. However, unscaled RMS errors (and cost functions) cannot be compared due to the very different variance of the datasets. The RMS errors for the standardized anomalies with the DA-I scheme (0.21 for the Arctic, 0.53 for Asia, 0.35 for Europe and 0.27 for North America) are lower than the DA-P ones (0.89, 0.49, 0.70 and 0.56 respectively), in all continents except
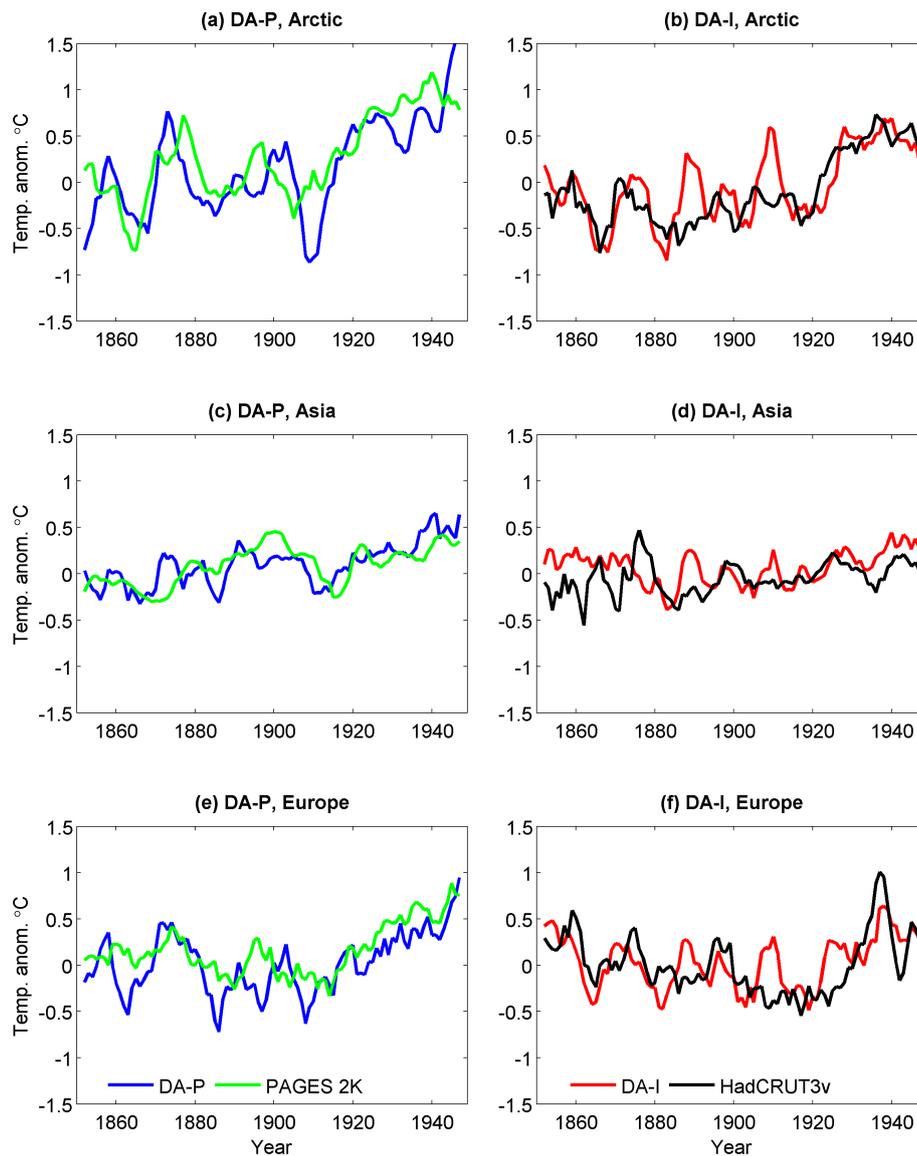
Figure 5.3: Consistency of the DA analyses with the assimilated data for 1850-1949 AD (as Figure 5.1) on the interannual timescale. All time series were smoothed using a five-year running mean filter.

Asia. Despite the fact that the standardisation of the DA-I and the DA-P time series is based on different periods, the standardised anomalies are approximately comparable, as the period 1850-1949 AD used in the DA-I scheme is out of the impact of the strong anthropogenic forcing of the late 20th century. The same holds for the cost function comparisons. The fact that the RMS errors are (apart from Asia) lower for the DA-I scheme than for the DA-P scheme shows that the analysis follows the target more closely when assimilating instrumental data.

The large discrepancy between the DA-I analysis and the instrumental data for Asia is owing to the fact that the mean temperature for the continent in the instrumental dataset is not representative of the region due to the sparse observations, as discussed below. The agreement between the instrumental time series and the PAGES 2K series is also much lower than for the other continents. A visual comparison of the temperature variations in the DA-I analysis with the instrumental data for Asia shows that they are in better agreement towards the end of the simulation period, when more observations are available.

5.3.2 Impact of the low observational coverage

The incomplete coverage of the PAGES 2K continental areas by the HadCRUT3v grid cells leads to a spatial sampling error in the continental mean temperatures, in addition to the error due to the uncertainty in the individual grid cells. This spatial sampling error depends on the number and position of the grid cells in the continental region, and increases with the lower observational coverage. Such a case is the region of Asia, where the coverage is extremely sparse at the beginning of the period and progressively becomes more complete (Figure 5.4). The grid cells that have observations during the first decade are only 47, whereas missing data are found in 121 grid cells. Conversely, 148 grid cells have data during the last decade. The expected error is exacerbated by the uneven spatial resolution during the first few decades, when coverage was concentrated in the western Pacific Ocean and East China Sea. The instrumental coverage for the whole NH in the earlier and later period is illustrated in Figure 5.5.

To check the impact of the poor observational coverage on the mean temperature of Asia, we subsampled the data of the last four decades of the simulation period (1910-1949 AD) to have the same coverage as the first decade (1850-59 AD), and calculated the difference between the complete and the subsampled average temperatures.

a) Winters 1850-59 AD

b) Winters 1940-49 AD
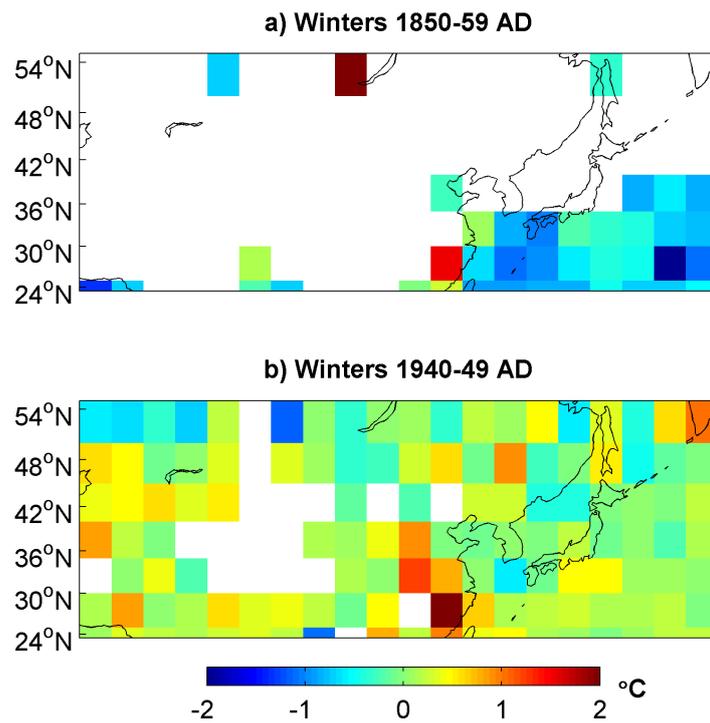
Figure 5.4: HadCRUT3v instrumental coverage in the PAGES 2K box for Asia during the first (1850-1859 AD) and last (1940-1949 AD) decade of the period. The values are decadal winter mean temperatures as anomalies w.r.t. the 1850-1949 AD mean.

a) Winters 1850-59 AD

b) Winters 1940-49 AD
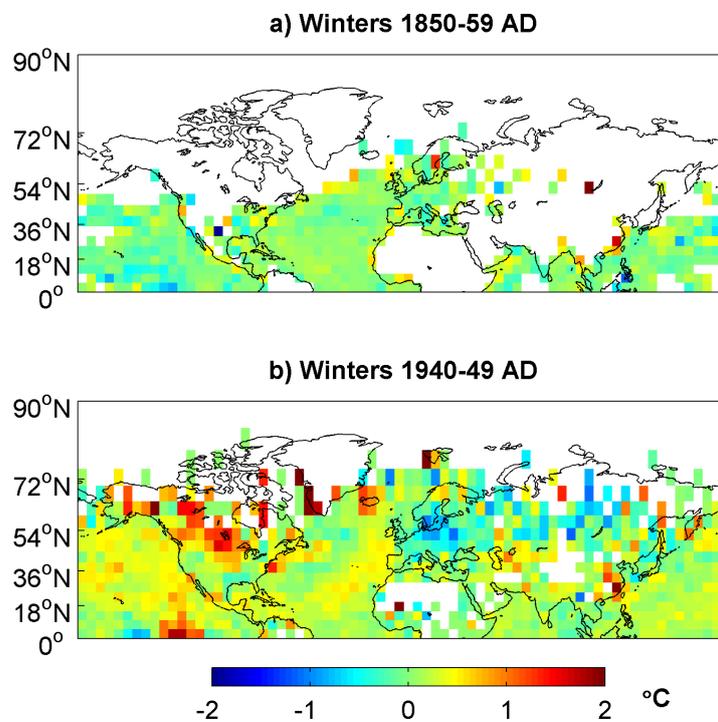
Figure 5.5: As Figure 5.4 but for the NH.

The four "low data coverage" means are compared with the respective higher data coverage means in Table 5.1. The eight temperature means (high and low resolution) are also plotted in Figure 5.6. The differences are on the order of 0.1°C, which is similar to the standard deviation of the DA-I analysis in Asia (0.12, Figure 5.1), indicating a significant impact of the poor observational coverage.

Decade	1910-1919	1920-1929	1930-1939	1940-1949
High density mean	0.01	-0.10	-0.01	0.11
Low density mean	-0.11	-0.14	-0.06	0.18

Table 5.1: Decadal winter mean temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the high and low density coverage of Asia in HadCRUT3v, during the last four decades of the simulation period.
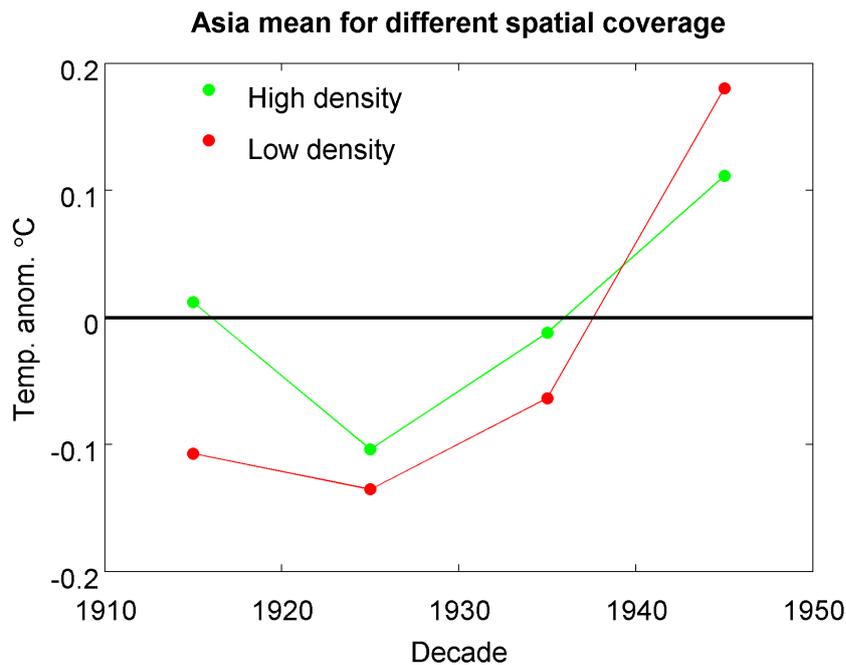
Figure 5.6: Decadal winter mean temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the high and low density coverage of Asia in HadCRUT3v, for the last four decades of the simulation period.

The sampling error in the gridded instrumental data leads to possible inconsistencies with reality and with the forcings, including errors in the dynamical links between different continents (e.g. correlations). The limited instrumental coverage is thus expected to lead to errors in the target data for Asia, which can explain the low skill

of the DA-I scheme for this continent. As the gridded observational record is an imperfect input dataset for our assimilation scheme, the difference between the skill of the assimilation of proxy-based and instrumental reconstructions is not identical to the additional error caused by the proxies. Our initial objective of obtaining a clear error separation can therefore not be fully accomplished.

5.3.3 Cost functions comparison

Another way of measuring the closeness between the simulated and assimilated time series in the two DA schemes is the cost function, which is the RMS error-based measure that has been used to select the best ensemble member for each decade. Table 5.2 provides the values of the best cost functions for the two DA schemes in each decade. The cost functions (shown in Figure 5.7) are based on decadal mean standardized anomalies for the four NH continents, using the mean and variance of 850-1849 AD in the DA-P scheme and 1850-1949 AD in the DA-I scheme. The best cost functions of the DA-I scheme are consistently lower than the DA-P ones in all decades, in agreement with the lower RMS errors found for the DA-I scheme. The mean of the cost functions for the closest member of the DA-P scheme is 1.27, whereas it is 0.61 for the DA-I scheme. Consistent with the results found in Section 5.3.1, this suggests that less realistic targets, such as noisy proxy-based reconstructions, cannot be followed as well by the model as more realistic ones, such as the instrumental data, especially in a small ensemble.

Decade	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
DA-P	0.87	1.36	1.48	0.58	1.33	1.13	0.87	1.52	2.47	1.07
DA-I	0.51	0.63	0.73	0.26	0.38	0.39	0.71	0.57	1.03	0.93

Table 5.2: Best cost functions for the DA-P and DA-I schemes, for the decades 1 (1850-1859 AD) to 10 (1940-1949 AD).

Our cost function does not include weights that account for the different levels
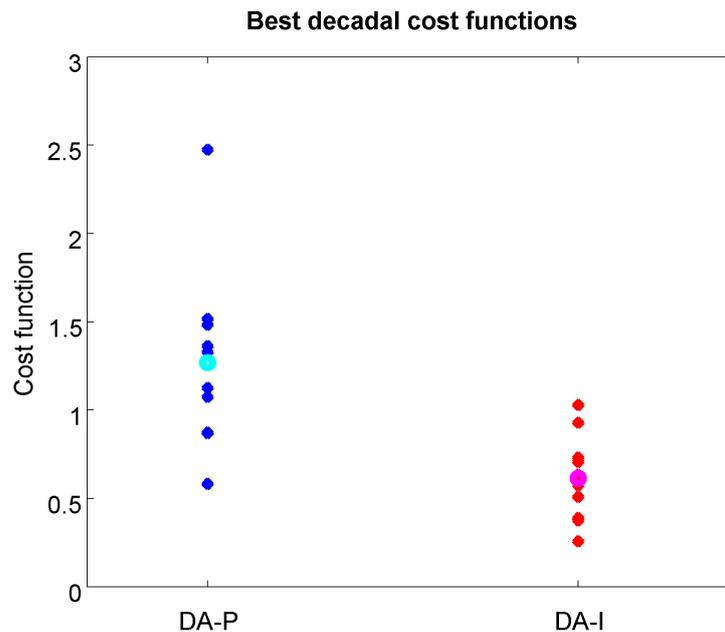
Figure 5.7: Best decadal cost functions with the DA-P (blue) and DA-I (red) schemes. The mean of the best cost functions for each scheme are also shown (cyan and magenta respectively).

of uncertainty in the assimilated data or for the different size of the regions represented. For the case of the proxy-based reconstructions, this is mainly because the uncertainty estimates have been calculated by the PAGES 2K groups using different methods for the different regions and thus the errors are not directly comparable. In addition, the published PAGES 2K uncertainties are defined for data at annual resolution. The decadal means used in our cost function can be expected to have lower uncertainties, but the exact level is difficult to determine due to the autocorrelation of the non-climatic noise in the reconstruction (Moberg and Brattstrom, 2011). For consistency, the instrumental errors (e.g. measurement and bias errors as well as the effect of limited observational coverage on the continental averages) have not been taken into account either.

5.3.4   Evaluation against the observed climate

The DA-I analysis has already been evaluated against the observed climate, i.e. the instrumental data. We now evaluate the DA-P analysis against these data, and compare the performance of the two schemes. The reference period for the simulation anomalies is 1961-90 AD, taken from a historical run with the MPI-ESM model, while for the gridded observations, anomalies are calculated w.r.t. the 1961-90 AD mean of the HadCRUT3v data set. The decadal mean temperature anomalies for the DA-P and the DA-I analysis as well as the instrumental data in the NH continents are shown in Figure 5.8. The correlations between the DA-P analysis and the HadCRUT3v data set, given in Table 5.3, are much lower than the ones between the DA-I analysis and HadCRUT3v in Europe, slightly lower in the Arctic and North America, but higher in Asia. For the NH mean, the DA-I analysis is again more consistent with the observations. Finally, the skill on the inter-annual timescales (annual values smoothed by a nine point Hamming window) is also better with the DA-I scheme (correlation is 0.76 compared with 0.67 for the NH mean).

	Arctic	Asia	Europe	N. America	NH mean
DA-P	0.90	0.62	0.45	0.84	0.78
DA-I	0.96	0.24	0.85	0.88	0.91

Table 5.3: Correlations between the NH continental mean temperatures from the DA analyses and the gridded instrumental data, for the two DA schemes.

It may not appear surprising that the correlation of the observed climate with the DA-I analysis is higher than with the DA-P analysis, because the observations have been assimilated in the DA-I scheme. However, this is not trivial as the cost function was based on RMS error and not on correlation. A limitation of this part of the analysis is that the HadCRUT3v dataset serves a twofold purpose: it is used as both the validation data and the assimilation target for the DA-I scheme. In a hypothetical case where the PAGES 2K reconstructions for the NH mean were
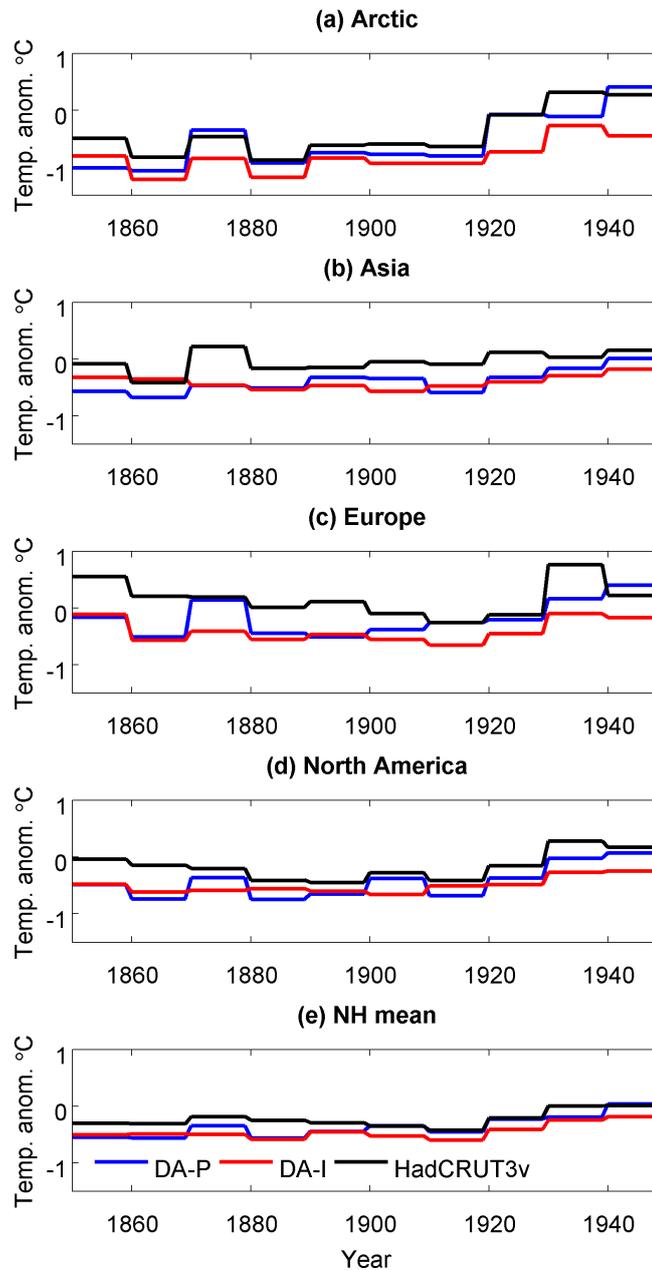
Figure 5.8: NH continental and mean decadal temperature anomalies for the DA-P analysis (blue), DA-I analysis (red) and the instrumental data (black). The reference period is 1961-1990 AD.

closer to reality than the HadCRUT3v mean, the DA-I analysis may still follow the HadCRUT3v reconstruction better. The overall consistency of the DA-P analysis with the instrumental data on the large spatial scales shows that given realistic targets, the Matsikaris et al. (2015b) DA setup, which followed the same scheme for a previous period (1750-1849 AD), performs well.

The comparison of the RMS errors, based on decadal mean differences between the DA analyses and the instrumental data, does not have much importance here, as the RMS errors are strongly influenced by the choice of the reference period. Unlike the correlations found above, the RMS errors were not lower for the DA-P analysis (e.g. 0.18 for the NH mean of the DA-P analysis and 0.23 for the DA-I analysis). A different reference period would however shift the time series and change the RMS errors drastically. The period 1961-90 AD was chosen because it is not included in any of the two DA cost functions, but it is a peculiar period with strong anthropogenic warming.

The correlations of the decadal mean temperature anomalies between the DA-P analysis and the DA-I one are relatively high in all continents, and modest even in the case of Asia (0.80 for the Arctic, 0.48 for Asia, 0.71 for Europe and 0.78 for North America). The correlation for Asia is lower than the correlations of the other three continents, but it is higher than the correlation between the DA-I analysis and the HadCRUT3v observations. The assimilation of an unrealistic mean temperature for Asia in the DA-I scheme causes the minimization of the cost function to be dominated by the variations in the other three continents. In general, the DA-I analysis is affected by both the target temperatures and the teleconnections between the continents. As the assimilated mean temperature for the Asia region in the DA-I scheme is unrealistic due to the large sampling error, the influence of the teleconnections leads the DA temperatures being inconsistent with the target data.

5.3.5 Skill on spatial patterns

Simulated temperature variations at regional scale cannot be in agreement with the observed ones in a free running simulation, due to the dominant contribution of random internal variability compared to forced climate variability. Good skill on

small spatial structures would therefore be added value and is one of the aims of DA. This skill is not easy to achieve in our setup, as it relies on constraining large-scale circulation modes by the assimilated continental temperatures. A useful tool for the evaluation of the temporal correlations between the DA analyses and the instrumental data are the correlation maps, nevertheless, we did not plot such maps because of the low number of time steps involved in our study (only ten). We have examined the spatial maps for each decade for the DA analyses and the observations, for Europe and the NH in winters and summers. The DA patterns for the NH in most decades show no agreement with the observed patterns in either of the two schemes (not shown). The patterns in the European sector (35N-65N, 10W-30E), where the instrumental dataset is almost complete, are also inconsistent (Figures 5.9 and 5.10 for the DA-P scheme, and Figures 5.11 and 5.12 for the DA-I scheme, for the winters and summers of 1850-1949 AD respectively). For an easier comparison between the two schemes, Figure 5.13 shows the European decadal surface air temperature maps for the two DA analyses and the HadCRUT3v observations for the summers of 1850-1859 AD and 1940-1949 AD, as an example.

The spatial correlations of the DA analyses with the HadCRUT3v data are not significant (Table 5.4 for the DA-P scheme and 5.5 for the DA-I scheme). As the HadCRUT3v dataset has a coarser resolution than the model (72×36 compared to 96×48 grid), the spatial correlations were calculated after interpolating the model grid onto the observational one. The average decadal correlations for the European summers, winters, NH summers and NH winters are -0.12, -0.14, 0.05 and 0.04 respectively with the DA-P scheme, and 0.24, -0.05, 0.07 and 0.01 with the DA-I scheme. The DA-I analysis has a higher correlation than the DA-P one only for Europe in summer, but this improvement is likely to be happening only by chance, especially as the NH teleconnections in summer are weaker than in winter and the performance of the DA scheme probably not as good. The fact that the assimilation of observational data does not provide any added value on regional scales shows that
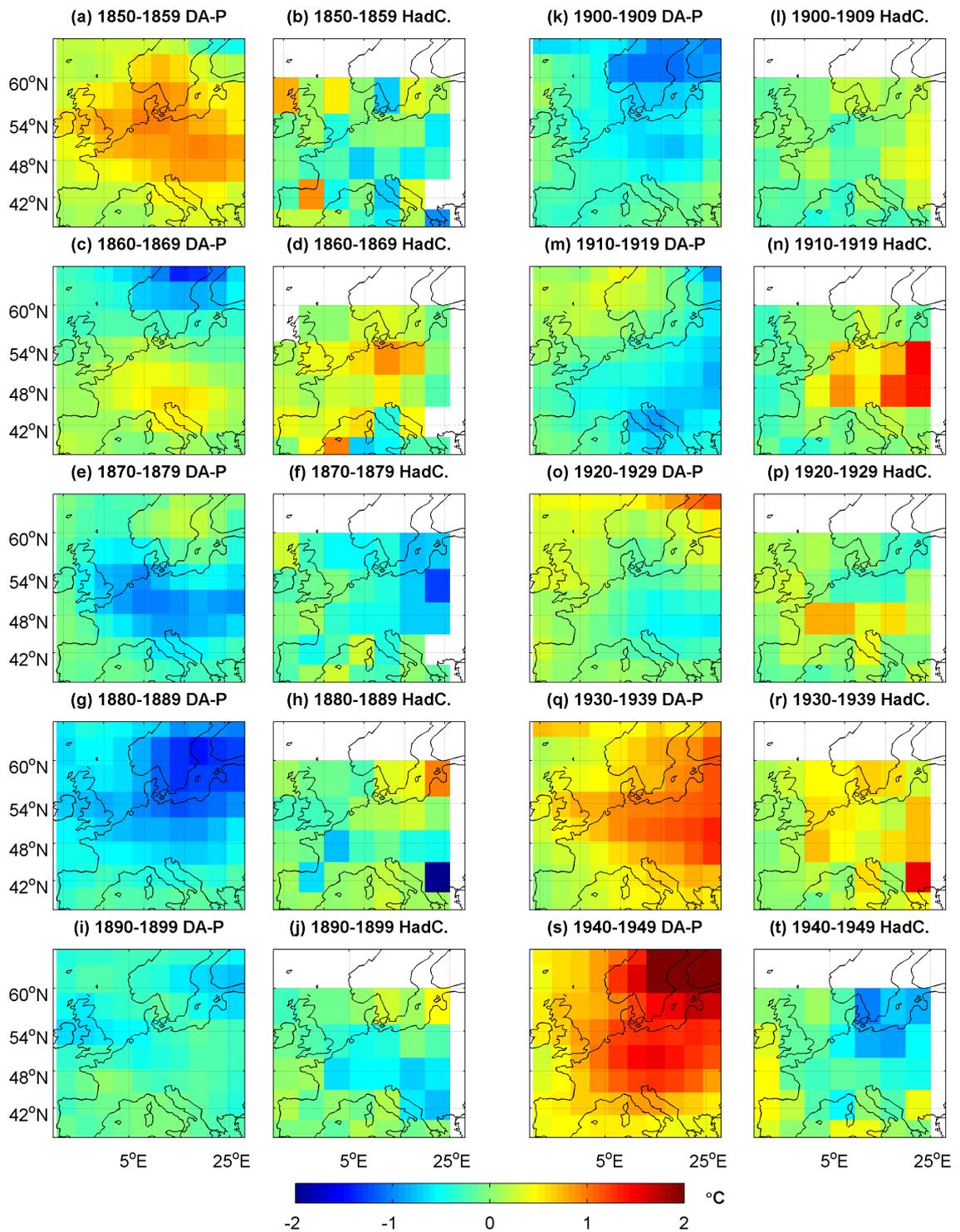
Figure 5.9: European decadal surface air temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the DA-P analysis and the HadCRUT3v reconstructions, for the winters of 1850-1949 AD.
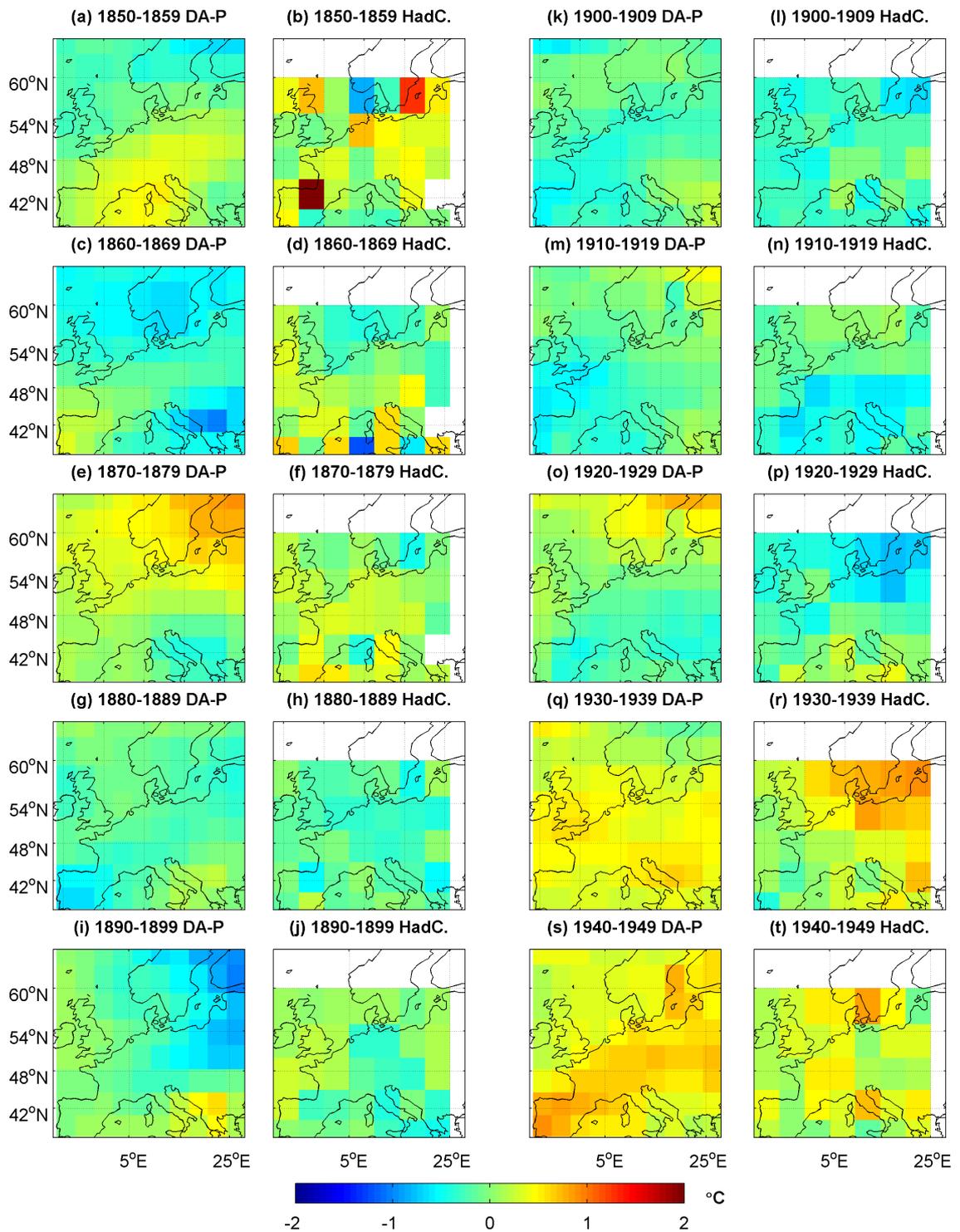
Figure 5.10: European decadal surface air temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the DA-P analysis and the HadCRUT3v reconstructions, for the summers of 1850-1949 AD.
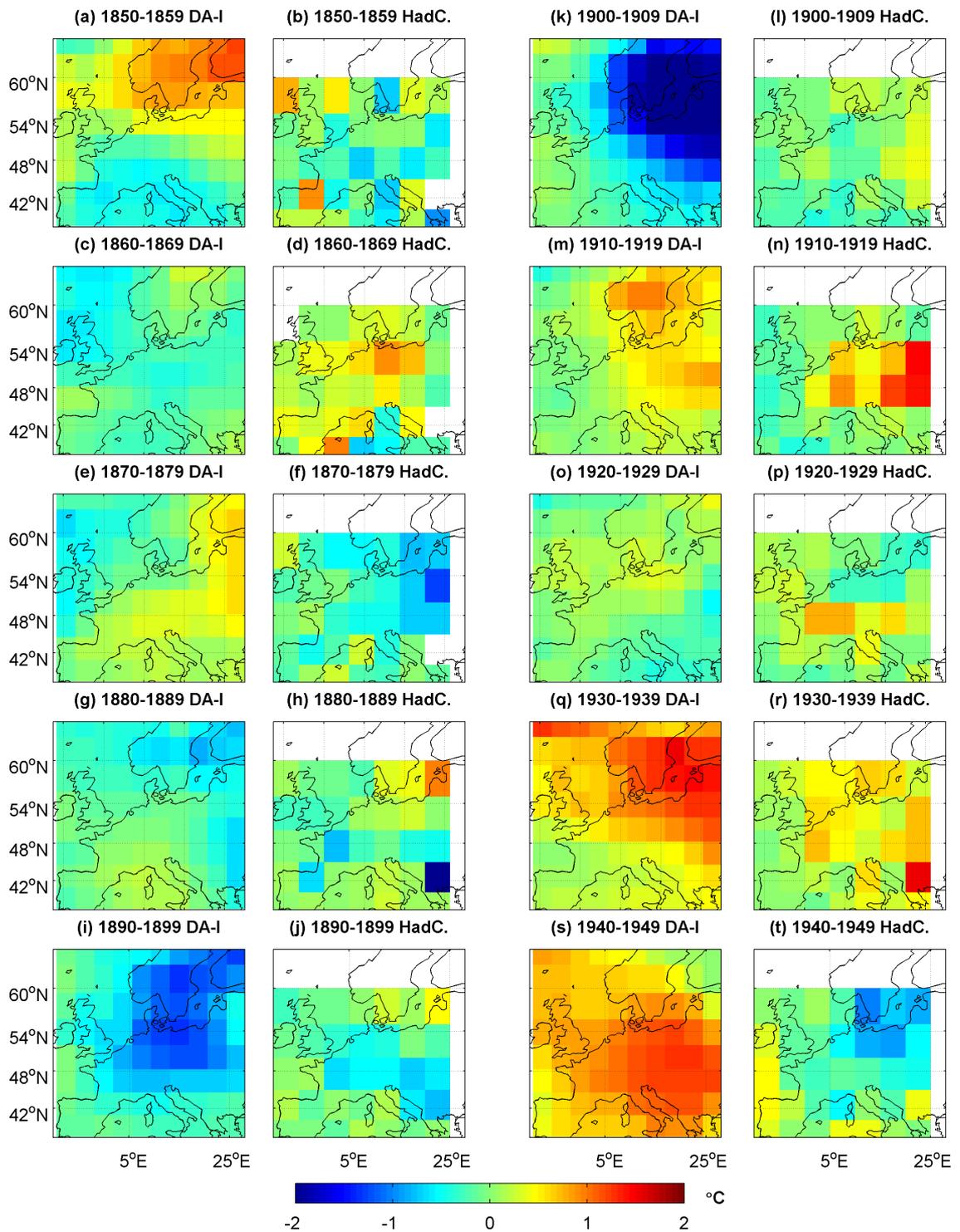
Figure 5.11: European decadal surface air temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the DA-I analysis and the HadCRUT3v reconstructions, for the winters of 1850-1949 AD.
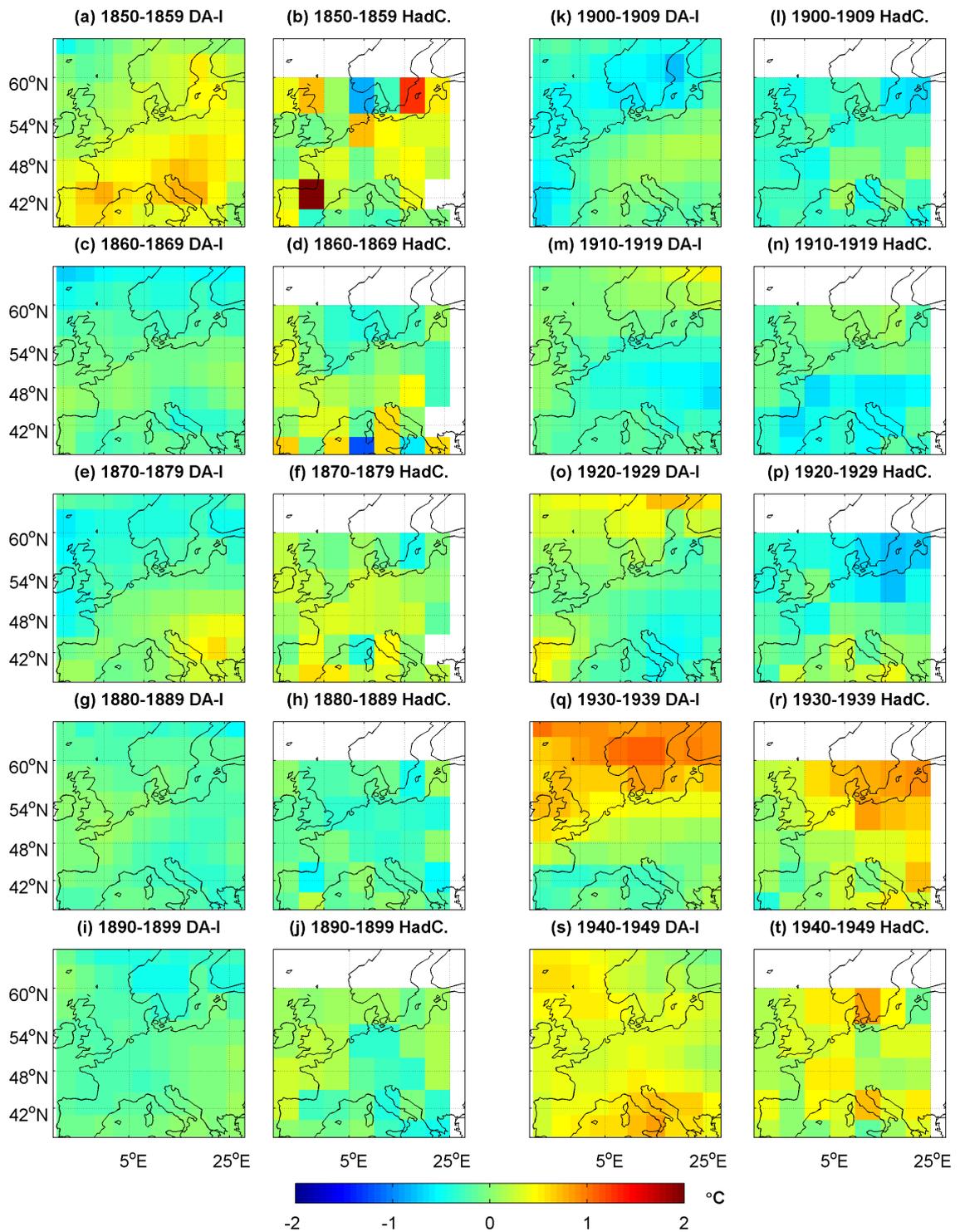
Figure 5.12: European decadal surface air temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the DA-I analysis and the HadCRUT3v reconstructions, for the summers of 1850-1949 AD.
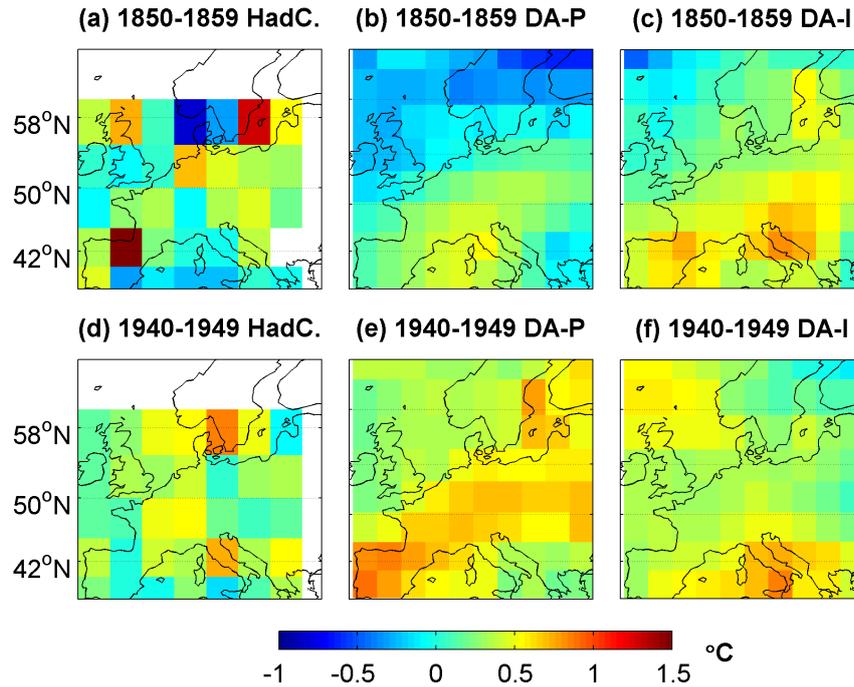
Figure 5.13: European decadal surface air temperatures (anomalies w.r.t. the 1850-1949 AD mean) for the DA-P analysis, the DA-I analysis and the HadCRUT3v reconstructions, for the summers of 1850-1859 AD and 1940-1949 AD.

the errors in the proxy-based reconstructions are not the main source of this lack of skill. The use of the continental scale target temperatures is thus the dominant source of error for the unrealistic representation of the small-scale temperature variability. Other aspects of the methodology, such as the ensemble size, may also have contributions. However, even though larger ensemble sizes might lead to some improvements, the complete lack of skill for our 20-member ensemble indicates that the main reason for the lack of skill is in fact the cost function formulation.

5.4 Summary and discussion

This chapter aimed at analysing two sets of ensemble simulations, which have assimilated continental-scale decadal instrumental temperature observations and

Decade	Europe summer	Europe winter	NH summer	NH winter
1850-59	-0.02	-0.20	0.03	0.02
1860-69	0.32	0.37	0.10	0.16
1870-79	-0.46	0.19	0.13	0.05
1880-89	-0.20	-0.31	0.18	0.18
1890-99	-0.19	-0.10	0.02	0.06
1900-09	0.00	-0.38	-0.07	0.00
1910-19	0.25	-0.59	-0.01	-0.09
1920-29	-0.42	-0.14	0.02	0.05
1930-39	-0.39	0.33	0.18	0.13
1940-49	-0.13	-0.57	-0.08	-0.16
Average	-0.12	-0.14	0.05	0.04

Table 5.4: Spatial correlations between the DA-P analysis and the HadCRUT3v temperature reconstructions for Europe and the NH, in summer and winter, for the decades 1850-59 AD to 1940-49 AD.

Decade	Europe summer	Europe winter	NH summer	NH winter
1850-59	0.17	0.21	-0.07	0.06
1860-69	0.32	-0.31	0.00	-0.03
1870-79	0.21	-0.40	0.08	0.03
1880-89	0.15	-0.05	0.00	-0.07
1890-99	-0.09	0.01	0.13	-0.01
1900-09	0.49	-0.63	0.20	-0.13
1910-19	0.63	0.63	0.12	0.28
1920-29	-0.25	0.10	-0.01	-0.07
1930-39	0.66	0.35	0.22	0.07
1940-49	0.06	-0.43	0.07	0.03
Average	0.24	-0.05	0.07	0.01

Table 5.5: Spatial correlations between the DA-I analysis and the HadCRUT3v temperature reconstructions for Europe and the NH, in summer and winter, for the decades 1850-59 AD to 1940-49 AD.

proxy-based reconstructions respectively. On the continental and hemispheric scale, both DA analyses follow the target variations well. However, although the correlations of the two DA analyses with the target temperatures are in general similar, the RMS errors of the DA-I analysis are much lower than those of the DA-P one. An exception for this is the case of Asia, where the DA-I analysis is inconsistent with the observations due to the limited observational coverage, which leads to errors in the target data. Another measure used for evaluating the two DA schemes is the decadal cost functions. The cost functions of the DA-I scheme

are found to be consistently lower than those of the DA-P scheme, in agreement with the lower RMS errors of the DA-I analysis. Furthermore, the DA-I analysis outperforms the DA-P one in terms of the correlations against the observed climate (HadCRUT3v data), but not substantially.

The overall performance of the two schemes confirms the expectation that the skill of DA is improved by the use of more realistic assimilation data, as these include more realistic teleconnections. In all continents, the DA temperatures are influenced by both the assimilated data and the teleconnections between the NH continents. In cases where the target is unrealistic, such as the Asia mean temperature of the DA-I scheme, the influence of the teleconnections leads to the simulated temperatures in the DA analysis being different to the target data. In addition to assessing the differences in the performance of the two DA schemes, the study serves as a validation of the DA scheme of Chapter 4 (Matsikaris et al., 2015b). In principle, since the DA-P scheme is the same as the Matsikaris et al. (2015b) scheme, which was run for the pre-industrial period (1750-1849 AD), the high consistency of the DA-P analysis with the observations over the instrumental period on the large spatial scales shows that the Matsikaris et al. (2015b) setup yields analyses that are close to the true continental and hemispheric-scale temperatures given realistic targets. The fact that the HadCRUT3v is used as both the validation data and the assimilation target for the DA-I scheme is a limitation of the study.

No skill on representing the regional temperature variability was found for either of the two DA schemes, as the spatial correlations between both DA analyses and the gridded instrumental data for the NH and the European sectors were non-significant. This demonstrates that even target data with only small errors are associated with low skill on small spatial scales. Therefore, methodology is the main source of this lack of skill. Error separation, which was one of the main objectives of the study, could however only be partly achieved, since the observational record is affected

by significant sampling error. The methodology consists of several components, such as the assimilation of continental means, the ensemble size and the specific formulation of the cost function. Given that the combination of the 20-member ensemble with the cost function formulation leads to a skilful representation of the large-scale variations, we believe that the main reason for the lack of added value on small spatial scales is the use of the continental scale for the assimilated data.

The lack of skill in capturing small-scale temperature variability when assimilating continental mean temperature might be related to the spatial scale or to the location of the continental-scale target data. A systematic investigation of the optimal spatial scale and optimal locations for assimilated temperatures is essential for the improvement of DA methods for palaeoclimatic applications. If the spatial scale is too large, which is likely to be the case for continental-scale target series, the temperature gradients within the regions that are caused by the major circulation modes are ignored and the information about the phase of these modes might thus be lost. At the other end of the spectrum is the assimilation of individual local temperature reconstructions. Although this would in principle capture the full information about spatial temperature fields, the local reconstructions might be associated with much larger errors than reconstructions for larger areas, which are based on a larger number of proxy records. While advanced assimilation methods that take into account the errors of the assimilated data can be expected to work well in this situation, simple methods as those used here might work better if spatially averaged temperatures with smaller errors are assimilated.

Chapter 6

Conclusions

6.1 Summary of the main findings

Proxy-based palaeoclimatic reconstructions are spatially incomplete and contain large uncertainties caused by non-climatic noise, reconstruction methods, measurement errors, inadequate understanding of the proxy response to environmental variations and other factors (e.g. Jansen et al., 2007; Jones and Mann, 2004). Standard model simulations cannot follow the real internal climate variability, and include systematic biases and errors in the forcings or in the response to them (Fernandez-Donado et al., 2013). DA combines empirical information from proxy data with numerical simulations to obtain the best climate state estimates, aiming to capture both the forced and internal climate variability. Different palaeoclimate DA approaches have been employed in the recent past. In most cases an ensemble-selection approach has been employed, as pioneered by Goosse et al. (2006). The primary aim of the research presented in this doctoral thesis was to implement and test DA methods using a GCM for the reconstruction of the past climate, and specifically some of the key periods from the last millennium.

As a first step in the study, an off-line and an on-line ensemble-based method were compared (Chapter 3), using as testing period the 17th century, which led into the Maunder Minimum. For this purpose, ten ensemble members were simulated for each DA scheme using a low-resolution version of the MPI-ESM model, and the PAGES 2K NH continental temperature reconstructions were assimilated. In the off-line approach, the ensemble for the entire simulation period was generated first and then the ensemble was used in combination with the empirical information to produce the analysis. In contrast, in the on-line approach, the ensembles were generated sequentially for sub-periods based on the analysis of previous sub-periods. The validation showed that both schemes performed better than the simulations without DA in following the assimilated target data on the continental and hemispheric scales. The on-line scheme would be expected to perform better than the off-line one 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 was based on analysing correlations and differences between the analysis and the proxy-based reconstructions, similar skill for both methods was found 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.

After the selection of the on-line DA scheme as the most appropriate one, 20 ensemble members were run implementing the specific setup for the reconstruction of the climate for 1750–1850 AD (Chapter 4). The performance of the assimilation was evaluated on large and small spatial scales. The assimilation was found to have good skill for large-scale temperatures, but no agreement between the DA analysis and the independent proxy-based reconstructions of Luterbacher et al.

(2004) for small-scale temperature patterns within Europe or with reconstructions for the NAO index (e.g. Luterbacher et al., 2002) was noted. To explain the lack of added value in small spatial scales, a MCA of links between NH temperature and sea level pressure was performed based on a control simulation with the MPI-ESM model. For annual values, winter and spring, the NAM is the pattern found to be most closely linked to the NH continental temperatures, while for summer and autumn it was a wave-like pattern. This link was reproduced in the DA for winter, spring and annual means, providing potential for constraining the NAM/NAO phase and in turn regional temperature variability. It was shown that the lack of actual small-scale skill is likely due to the fact that the link might be too weak, as the NH continental mean temperatures are not the best predictors for large-scale circulation anomalies, or that the PAGES 2K temperatures include noise. Both factors can lead to circulation anomalies in the DA analysis that are substantially different from reality, leading to unrealistic representation of small-scale temperature variability. Moreover, it was shown that even if the true amplitudes of the leading MCA circulation patterns were known, there is still a large amount of unexplained local temperature variance.

The question whether the lack of information propagation to small spatial scales found in Chapter 4 is due to the methodology or to errors in the assimilated reconstructions was investigated in Chapter 5. The aim of the chapter was twofold: (i) to assimilate instrumental data for 1850-1949 AD as a target with no errors for the DA, in an attempt to obtain the aforementioned error separation; and (ii) the use of the instrumental data for an independent evaluation of the DA scheme employed in Chapters 3 and 4. Simulations for the instrumental period allow validating the assimilation method against instrumental records. Error separation is fundamental, as it can lead to improvements in DA methods. The two aims were addressed by performing a new set of simulations, using two different sets of target data; the proxy-based PAGES 2K reconstructions (DA-P scheme), and the HadCRUT3v

instrumental observations (DA-I scheme). Ensemble-member selection on-line DA using the MPI-ESM model was again employed, and continental mean temperatures for the NH were assimilated. Both DA schemes were found to follow the large-scale target and observed climate variations well, but the assimilation of instrumental data improved the performance. This improvement could not be seen for Asia, where the limited instrumental coverage leads to errors in the target data and low skill for the DA-I scheme. No skill on small spatial scales was found for either of the two DA schemes, demonstrating that errors in the assimilated data are not the main reason for the unrealistic representation of the regional temperature variability. It was thus concluded that assimilating continental mean temperatures is not ideal for providing skill on small spatial scales.

6.2   Limitations and scope for further research

It is important to note that the studies presented here are only a first step towards the assimilation of palaeoclimate data, and that the field has only been developed in the recent past. Some of the main limitations of the work undertaken in this thesis and different perspectives for further research have already been identified in previous chapters, and are revisited in the discussion of this section. The limitations must be kept in mind when interpreting the results. Despite the obstacles associated with the studies presented here, the thesis provides valuable perspectives on challenges faced by future studies. The main issues limiting the skill of DA in this work can be grouped into four principal categories; limitations associated with: (i) the type of DA; (ii) the specifications for the particle filter; (iii) the type of information that are assimilated; and (iv) the model and proxies used. Furthermore, the validation setup for the pre-industrial studies was also limited, because validation was only performed against proxy-based reconstructions, which are not the reality. Alternatively, validation against the (unknown) true climate

would require pseudoproxy studies or instrumental data, which are not available for the pre-industrial period.

Starting with the DA type, the particle filter ensemble selection performed in this thesis is just one out of many approaches to DA. In principle, DA for palaeoclimate applications faces various different challenges (Section 1.3), and most of the palaeoclimate DA approaches undertaken so far, including the one employed here, are different than in meteorology, although still formulated in a proper Bayesian framework. The fact that the observations are averages over long periods, the unknown observation operator for the direct assimilation of proxy data, and the requirement for efficiency of the palaeoclimate DA methods for long simulations, lead to various technical problems in some of the methods. Moreover, the degenerate particle filter employed here is a simple DA approach that does not attempt to take into account model and observation errors. Progress of DA methods requires the use of realistic error estimates for both proxies and models. Other DA schemes, such as variational DA and Kalman filters, include error estimates, but are not easy to adapt to palaeoclimate applications because of the assimilation of temporal means rather than individual states. The attempt of employing other DA methods is one of the aims of future work.

The specific choices for the particle filter limit the potential for a broad generalization of the results obtained in this thesis. Additional experiments are required to check whether the conclusions reached in the particular setups of each chapter are valid in other cases. The general skill of the methods might depend on the type of particle filter employed, the variable assimilated, the frequency of assimilation, the selected cost function, the ensemble size, and other factors. In this study, the selection of the best member was always based on the atmospheric temperature state, which cannot assure that the ocean state is also realistic. In addition, the sample size for ensemble-based methods plays an important role for the success

of the methods. The 10- and 20- member ensembles performed in the different studies were a compromise between ensemble size and computational cost. Larger ensembles are expected to give a better chance for finding simulated states that are close to the true climate signal.

The type of information that should be assimilated in DA schemes is another important aspect that needs to be further examined. The lack of small spatial-scale skill found in Chapters 4 and 5 might, for instance, be related to the spatial scale or the location of the target data that have been assimilated. The link between assimilated temperature reconstructions and large-scale circulation is expected to get stronger when regional rather than continental-scale reconstructions are used. Reconstructed temperatures at smaller spatial scales are however affected by larger errors and are highly variable or even contradictory at nearby locations. Based on the findings of this thesis, we believe that a systematic investigation of optimal assimilation locations and spatial scales in order to provide skill on smaller spatial scales is essential for the improvement of DA. Future work on this can include a statistical analysis of observed and simulated temperature and pressure values, to identify locations across the Northern and Southern Hemisphere that can provide a stronger constraint of regional climates by temperature variations.

Finally, apart from the basic approach to DA itself, major limitations are introduced by the uncertainties of the data that are assimilated and the consequences of the use of imperfect climate models. Since both proxy-based reconstructions and models are associated with substantial uncertainties and errors, palaeoclimate DA will advance through the improvement of these components. The variance underestimation, often introduced by the regression-based calibration methods followed in many proxy-based reconstructions, is an important obstacle. Furthermore, the paucity of proxies, especially in the SH and the tropics, is another constraint to DA. The development of more accurate proxy-based reconstructions, or reliable observation

operators (also known as forward models) for the case of direct assimilation of proxies, will be of great benefit to DA applications. In terms of the climate models, DA cannot produce reliable results if model bias problems are present. In principle, although the use of anomalies for the assimilation subtracts the model bias in the mean, the realistic structure of variability is a prerequisite for good DA performance. In addition, since information propagation on long timescales in the real and in the model world relies on the structure of variability, advances in climate modelling are likely to lead to more skilful predictability and further progress in DA.

Taking into account the main conclusions of this thesis and the principal limitations associated with it, a future direction of the research presented here would be to test different DA methods, or employ the full rather than the degenerate particle filter, i.e. retain more than one ensemble member, which could produce a bigger ensemble spread for the ocean. Another setup that can be attempted is to define the cost function on the basis of one- or thirty-year mean values instead of decadal means, in order to check whether ocean memory on those timescales leads to improvements to the on-line DA approach. More tests could be carried out by enhancing the ensemble size, or by using different proxy datasets of higher resolution and at locations that are the most appropriate for constraining some main modes of atmospheric variability.

Bibliography

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.

Bothe, O., Evans, M., Donado, L. F., Bustamante, E. G., Gergis, J., Gonzalez-Rouco, J. F., Goosse, H., Hegerl, G., Hind, A., Jungclaus, J., Kaufman, D., Lehner, F., McKay, N., Moberg, A., Raible, C. C., Schurer, A., Shi, F., Smerdon, J. E., von Gunten, L., Wagner, S., Warren, E., Widmann, M., Yiou, P., and Zorita, E.: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past, 11, 1673–1699, doi:10.5194/cp-11-1673-2015, 2015.

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., Smith, C., and Wallace, J. M.: An intercomparison of methods for finding coupled patterns in climate data, Journal of Climate, 5, 541–560, 1992.

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.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from

1850, Journal of Geophysical Research-Atmospheres, 111, doi:10.1029/2005jd006548, 2006.

Brohan, P., Allan, R., Freeman, E., Wheeler, D., Wilkinson, C., and Williamson, F.: Constraining the temperature history of the past millennium using early instrumental observations, Climate of the Past, 8, 1551–1563, doi:10.5194/cp-8-1551-2012, 2012.

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.

Cole-Dai, J.: Volcanoes and climate, Wiley Interdisciplinary Reviews-Climate Change, 1, 824–839, doi:10.1002/wcc.76, 2010.

Cole-Dai, J., Ferris, D., Lanciki, A., Savarino, J., Baroni, M., and Thiemens, M. H.: Cold decade (AD 1810-1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption, Geophysical Research Letters, 36, doi:10.1029/2009gl040882, 2009.

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.

Fernandez-Donado, L., Gonzalez-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., Garcia-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P., and Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, Climate of the Past, 9, 393–421, doi: 10.5194/cp-9-393-2013, 2013.

Frank, D. C., Esper, J., Raible, C. C., Buntgen, U., Trouet, V., Stocker, B., and Joos, F.: Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, Nature, 463, 527–U143, doi:10.1038/nature08769, 2010.

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., Bronnimann, 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.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1107, doi:10.1175/2009bams2607.1, 2009a.

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, 2009b.

Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., and Zorita, E.: Detection of human influence on a new, validated 1500-year temperature reconstruction, Journal of Climate, 20, 650–666, doi:10.1175/jcli4011.1, 2007.

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.

Hurrell, J. W.: Decadal trends in the North-Atlantic Oscillation - Regional temperatures and precipitation, Science, 269, 676–679, doi:10.1126/science.269.5224.676, 1995.

Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W., Rahmstorf, S., Ramesh, 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.

Jones, P. D., Jonsson, T., and Wheeler, D.: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland, International Journal of Climatology, 17, 1433–1450, 1997.

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.

Jungclaus, J. H., Lohmann, K., and Zanchettin, D.: Enhanced 20th-century heat transfer to the Arctic simulated in the context of climate variations over the last millennium, Climate of the Past, 10, 2201–2213, doi:10.5194/cp-10-2201-2014, 2014.

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.

Klein, F., Goosse, H., Mairesse, A., and de Vernal, A.: Model-data comparison and data assimilation of mid-Holocene Arctic sea ice concentration, Climate of the Past, 10, 1145–1163, doi:10.5194/cp-10-1145-2014, 2014.

Law, K. J. H. and Stuart, A. M.: Evaluating Data Assimilation Algorithms, Monthly Weather Review, 140, 3757–3782, iSI Document Delivery No.: 036GD, 2012.

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.

Lozier, M. S.: Overturning in the North Atlantic, Annual Review of Marine Science, Vol 4, 4, 291–315, doi:10.1146/annurev-marine-120710-100740, 2012.

Luterbacher, J., Xoplaki, E., Dietrich, D., Jones, P. D., Davies, T. D., Portis, D., Gonzalez-Rouco, J. F., von Storch, H., Gyalistras, D., Casty, C., and Wanner, H.: Extending North Atlantic Oscillation reconstructions back to 1500, Atmospheric Science Letters, 2, 114–124, doi:10.1006/asle.2001.0044, 2002.

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, 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. and Jones, P. D.: Global surface temperatures over the past two millennia, Geophysical Research Letters, 30, doi:10.1029/2003gl017814, 2003.

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.

Maraun, D., Widmann, M., Gutierrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A. I.: VALUE: A framework to validate downscaling approaches for climate change studies, Earths Future, 3, 1–14, doi:10. 1002/2014ef000259, 2015.

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.

Matsikaris, A., Widmann, M., and Jungclaus, J.: On-line and off-line data assimilation in palaeoclimatology: a case study, Climate of the Past, 11, 81–93, doi:10.5194/cp-11-81-2015, 2015a.

Matsikaris, A., Widmann, M., and Jungclaus, J.: Assimilating continental mean temperatures to reconstruct the climate of the late pre-industrial period, Climate Dynamics, doi:10.1007/s00382-015-2785-9, 2015b.

Moberg, A. and Brattstrom, G.: Prediction intervals for climate reconstructions with autocorrelated noise-An analysis of ordinary least squares and measurement error methods, Palaeogeography Palaeoclimatology Palaeoecology, 308, 313–329, doi:10. 1016/j.palaeo.2011.05.035, 2011.

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.

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., Gonzalez-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., Raible, C. C., Steig, E. J., van Ommen, T., Vance, T., Villalba, R., Zinke, J., and Frank, D.: Inter-hemispheric temperature variability over the past millennium, Nature Climate Change, 4, 362–367, doi:10.1038/nclimate2174, 2014.

O'Neill, J. J., Cai, X. M., and Kinnersley, R.: A generalised stochastic backscatter model: large-eddy simulation of the neutral surface layer, Quarterly Journal of the Royal Meteorological Society, 141, 2617–2629, doi:10.1002/qj.2548, 2015.

Oppenheimer, C.: Climatic, environmental and human consequences of the largest known historic eruption: Tambora volcano (Indonesia) 1815, Progress in Physical Geography, 27, 230–259, doi:10.1191/0309133303pp379ra, 2003.

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.

Robock, A.: Volcanic eruptions and climate, Reviews of Geophysics, 38, 191–219, doi:10.1029/1998rg000054, 2000.

Rohde, R., Muller, R., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C.: A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, Geoinfor Geostat: An Overview 2012, 1:1, doi:http://dx.doi.org/10.4172/2327-4581.1000101, 2012.

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.

Schurer, A. P., Tett, S. F. B., and Hegerl, G. C.: Small influence of solar variability on climate over the past millennium, Nature Geoscience, 7, 104–108, doi:10.1038/ngeo2040, 2014.

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.

Sundberg, R., Moberg, A., and Hind, A.: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium - Part 1: Theory, Climate of the Past, 8, 1339–1353, 2012.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, doi:10.1175/bams-d-11-00094.1, 2012.

Trouet, V., Esper, J., Graham, N. E., Baker, A., Scourse, J. D., and Frank, D. C.: Persistent Positive North Atlantic Oscillation Mode Dominated the Medieval Climate Anomaly, Science, 324, 78–80, doi:10.1126/science.1166349, 2009.

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

Van der Schrier, G. and Barkmeijer, J.: Bjerknes' hypothesis on the coldness during AD 1790-1820 revisited, Climate Dynamics, 25, 537–553, doi:10.1007/s00382-005-0053-0, 2005.

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.

Von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., Gonzalez-Rouco, F., and Tett, S. F. B.: Reconstructing past climate from noisy data, Science, 306, 679–682, doi:10.1126/science.1096109, 2004.

Widmann, M.: One-dimensional CCA and SVD, and their relationship to regression maps, Journal of Climate, 18, 2785–2792, doi:10.1175/jcli3424.1, 2005.

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.

Williams, L. D. and Wigley, T. M. L.: A comparison of evidence for late Holocene summer temperature variations for the Northern Hemisphere, Quaternary Research, 20, 286–307, doi:10.1016/0033-5894(83)90014-5, 1983.

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(15), doi:10.1029/2005gl023424, 2005.

Zanchettin, D., Timmreck, C., Graf, H. F., Rubino, A., Lorenz, S., Lohmann, K., Kruger,

K., and Jungclaus, J.: Bi-decadal variability excited in the coupled ocean-atmosphere system by strong tropical volcanic eruptions, Climate Dynamics, 39, 419–444, iSI Document Delivery No.: 965CZ5, 2012.

Zanchettin, D., Timmreck, C., Bothe, O., Lorenz, S. J., Hegerl, G., Graf, H. F., Luterbacher, J., and Jungclaus, J. H.: Delayed winter warming: A robust decadal response to strong tropical volcanic eruptions?, Geophysical Research Letters, 40, 204–209, doi:10.1029/2012gl054403, 2013a.

Zanchettin, D., Bothe, O., Graf, H. F., Lorenz, S. J., Luterbacher, J., Timmreck, C., and Jungclaus, J. H.: Background conditions influence the decadal climate response to strong volcanic eruptions, Journal of Geophysical Research-Atmospheres, 118, 4090–4106, doi:10.1002/jgrd.50229, 2013b.

Zorita, E. and von Storch, H.: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, Journal of Climate, 12, 2474–2489, part 2, 1999.

Publications

Matsikaris, A., Widmann, M., and Jungclaus, J.: On-line and off-line data assimilation in palaeoclimatology: a case study, Climate of the Past, 11, 81-93, doi:10.5194/cp-11-81-2015, 2015a.

Matsikaris, A., Widmann, M., and Jungclaus, J.: Assimilating continental mean temperatures to reconstruct the climate of the late pre-industrial period, Climate Dynamics, doi:10.1007/s00382-015-2785-9, 2015b.

Matsikaris, A., Widmann, M., and Jungclaus, J.: Influence of proxy data uncertainty on data assimilation for the past climate, Clim. Past Discuss., doi:10.5194/cp-2015-192, in press, 2016.