**Climate
of the Past**

# An ensemble-based approach to climate reconstructions

**J. Bhend[1,*], J. Franke[2], D. Folini[1], M. Wild[1], and S. Brönnimann[2]**

[1]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland
[2]Oeschger Centre and Institute of Geography, University of Bern, Bern, Switzerland
[*]now at: CSIRO Marine and Atmospheric Research, Aspendale, Australia

*Correspondence to:* J. Bhend (jonas.bhend@csiro.au)

**Abstract.** Data assimilation is a promising approach to obtain climate reconstructions that are both consistent with observations of the past and with our understanding of the physics of the climate system as represented in the climate model used. Here, we investigate the use of ensemble square root filtering (EnSRF) – a technique used in weather forecasting – for climate reconstructions. We constrain an ensemble of 29 simulations from an atmosphere-only general circulation model (GCM) with 37 pseudo-proxy temperature time series. Assimilating spatially sparse information with low temporal resolution (semi-annual) improves the representation of not only temperature, but also other surface properties, such as precipitation and even upper air features such as the intensity of the northern stratospheric polar vortex or the strength of the northern subtropical jet. Given the sparsity of the assimilated information and the limited size of the ensemble used, a localisation procedure is crucial to reduce "overcorrection" of climate variables far away from the assimilated information.

## 1 Introduction

Compared to conventional reconstruction methods, data assimilation represents a novel approach to increase our understanding of past climate. In this paper, we explore in an idealised setup if assimilation of sparse and indirect observations of past climate states, as recorded in climate proxies, provides sufficient constraints to skilfully update existing model simulations.

Two distinct approaches have often been used when reconstructing past climate: empirical methods relate the changes in climate proxies – such as tree-ring widths or $\delta^{18}O$

concentrations in ice cores – to changes in climate variables during past decades (see Jansen et al., 2007, and Jones et al., 2009, for an overview of recent advances). This relationship is then extended backwards, allowing for the reconstruction of said climate variables for times when no direct observations of the climate system are available. Empirical methods rely on the stationarity of the relationship between climate and proxy record. In addition, the specifics of high-resolution proxy archives make it hard to quantify low-frequency variability (Moberg et al., 2005). Dynamical methods, on the other hand, use reconstructed external forcings (e.g. changes in solar irradiance, land cover, atmospheric aerosol and greenhouse gas concentrations) to constrain simulations of past climate states (e.g. Jungclaus et al., 2010; Wanner et al., 2008; Ammann et al., 2007). In contrast to empirical approaches, dynamical methods allow us to also reconstruct climate variables, which are only loosely correlated to climate proxies. Ensembles of climate model simulations, however, are often not well constrained, as a large part of the variability is generated in the climate system itself and is thus independent of external forcings.

To overcome the relative weaknesses of these two approaches, it has been proposed to directly assimilate proxy data into climate model simulations (Goosse et al., 2009; Hughes and Ammann, 2009; Widmann et al., 2010). In the following, we provide a short introduction to the concept of data assimilation. The goal of data assimilation is to provide a best estimate of the true state of the system. This estimate is called the *analysis*. A simple way to produce an analysis is to interpolate observations of the true state of the system. In a paleoclimatological context in particular, however, the observations are indirect and sparse in space and time and the problem of producing an analysis is underdetermined.

Additional constraints are therefore needed to solve the problem. These additional constraints originate from the physics and dynamics of the system as formulated in a climate model. The model provides a first guess of the true state of the system that is consistent with the boundary conditions. This *prior* estimate of the true state is updated with observations to produce the analysis. The model is then used to propagate the analysis to provide a first guess for the next analysis cycle. This recursive procedure to accumulate observed information in the model is referred to as data assimilation.

First attempts to assimilate climate proxy information into models include the pioneering work of von Storch et al. (2000), Hargreaves and Annan (2002), van der Schrier and Barkmeijer (2005), Goosse et al. (2006, 2010), and Franke et al. (2010). The proposed approaches can be roughly separated into three groups: the methods of von Storch et al. and van der Schrier and Barkmeijer seek to push a model simulation towards a large-scale target state through nudging (von Storch et al., 2000) or using *singular forcing vectors* (van der Schrier and Barkmeijer, 2005). The methods by Goosse et al. (2006) and Franke et al. (2010) select optimal matches with the available proxy information among a set of model states and combine these to "pseudo-simulations". Recently, Goosse et al. (2010) modified their assimilation method to produce dynamically consistent past climate states, based on simulations with an Earth system model of intermediate complexity. All of the approaches discussed so far do not generically provide confidence intervals together with their best estimate, a shortcoming that is overcome by the approach proposed by Hargreaves and Annan. In contrast, their fully probabilistic approach is not tractable with a complex and computationally expensive model. Therefore, we propose a new approach that both allows us to assimilate proxy data into a high-resolution general circulation model (GCM) and that provides a generic quantification of the uncertainties.

Data assimilation has long been used in numerical weather forecasting to estimate optimal initial conditions for weather predictions (Kalnay, 2003). The variational data assimilation techniques developed for weather forecasting, however, are not suitable for reconstruction of past climate with a much smaller number of observations or climate proxies. A much simpler to implement and computationally less expensive method to assimilate data into climate model simulations is represented by the class of square root filters.

We use the ensemble square root filter (EnSRF) – a variant of the ensemble Kalman filter (EnKF; see Evensen, 2003, and references therein) – as introduced by Whitaker and Hamill (2002) to update the ensemble of model simulations with information from climate proxies. The EnSRF has successfully been used to produce a reanalysis for the period from 1870 to present using sea level pressure measurements (Compo et al., 2006, 2011). Here, we investigate whether EnSRF can also be used with spatially sparse observations with low temporal resolution.

Our main goal is to learn to what extent and under what conditions proxy data provide sufficient constraints for a data assimilation approach to climate reconstructions using the EnSRF algorithm. For this introductory analysis, we use a perfect model framework to explore the potential benefits of an ensemble-based approach to climate reconstructions. One simulation of the ensemble describes the true climate from which we generate pseudo-proxies, and the remaining simulations are used to estimate the true climate.

To be able to experiment with the details of the setup and properly explore the potential of an ensemble-based approach to reconstructions at a reasonable computational cost, we want to be able to run the assimilation off-line. Thus, we use an atmosphere-only GCM to describe past climates. In this setup, the proxy information has a temporal resolution (semi-annual in our case) that is far greater than the deterministic predictability of most atmospheric processes (Lorenz, 1969; Kalnay, 2003). This discrepancy in time scales between the memory of the system (a few days) and the observation interval (semi-annual) has profound implications for the proposed assimilation approach.

In a conventional assimilation, the effect of constraining the model with past observations gets propagated by the model and determines to a large extent the current first guess (unconstrained simulation). This is not the case in our idealised setup. Due to the chaotic nature of the atmosphere, the effect of the previous update (leading to the new analysis) is lost long before the end of the simulation cycle. That is, a forward integration of a constrained simulation and an unconstrained simulation are indistinguishable after six months on average. Thus, observed information does not accumulate over time, but only current observed information constrains the model. Therefore, we can assimilate the data non-recursively; that is, we do not need to feed back the corrected states (the analysis) as new initial conditions for the next simulation cycle. Consequently, we should not refer to the approach as a filter. Instead, we suggest to refer to the method as ensemble square root fitting.

Ultimately, we aim at assimilating climate proxy data into a coupled atmosphere-ocean GCM. In such a coupled system, there is far more long-term memory, and thus we will need to revert to the conventional recursive assimilation procedure. In the light of the final goal and to highlight the origins of the approach, we do not resolve the ambiguity in the abbreviation and we keep referring to our simplified approach as EnSRF.

In the following section, the model data and analysis scheme are introduced. In Sect. 3, we introduce the original EnSRF algorithm along with our modifications. In Sect. 4, we present the results from the validation assessment of the analysis versus the unconstrained ensemble of model simulations. We discuss the strengths and limitations of data assimilation using EnSRF for climate reconstructions in Sect. 5 and finish with conclusions in the final section of the manuscript.

## 2 Materials and analysis metrics
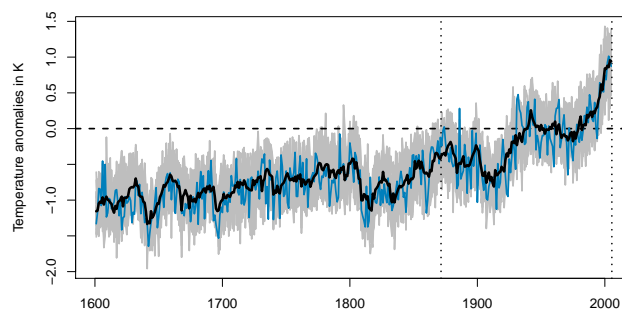
### 2.1 Model simulations

For the assessment of EnSRF for climate reconstructions, we used an initial condition ensemble of 30 simulations with the general circulation model (GCM), ECHAM5.4 (Roeckner et al., 2003, 2004). The model was run in T63L31 resolution, corresponding to an approximate horizontal resolution of 1.875° with 31 vertical levels from the surface to 10 hPa.

ECHAM5.4 was forced with reconstructed sea surface temperatures (SST, reconstruction by Mann et al., 2009), augmented with ENSO-dependent intra-annual variability according to the reconstructed NINO3.4 index of Cook et al. (2008) and climatological sea-ice according to the HadISST climatology (Rayner et al., 2003). We further used reconstructed solar irradiance (Lean, 2000) and land surface parameters derived from the land-use reconstructions of Pongratz et al. (2008). Additionally, the model was forced with reconstructions of volcanic activity by Crowley et al. (2008) and concentrations of long-lived greenhouse gases as used in Yoshimori et al. (2010, and references therein). Finally, transient sulphate concentrations were prescribed according to the reconstructed aerosol loads of Koch et al. (1999); before 1850, tropospheric sulphate aerosol concentrations were set to their 1850 values.

The solar irradiance reconstruction by Lean (2000) exhibits an increase in irradiance of approximately 2.5 W m$^{-2}$ since the Maunder Minimum (MM). Recent reconstructions, however, show less of a change in solar irradiance between the MM and present conditions (Wang et al., 2005; Krivova et al., 2007). Nevertheless, we chose a strong solar forcing, as the recent study by Jungclaus et al. (2010) has shown that this leads to a slightly more realistic climate response over the past 1000 yr in ECHAM5.4.

The individual simulations were branched off from a control run reflecting conditions around 1600 AD. The boundary conditions for the atmospheric model were prescribed and identical across the thirty different ensemble members. The spread of the initial condition ensemble thus reflects internal variability of the atmosphere alone. The time series of Northern Hemispheric annual average temperatures north of 20° N (in Fig. 1) illustrate the relative importance of external forcings and internal variability. The forced response of 2 °C warming from 1600 to 2000 in the extratropical Northern Hemisphere is slightly stronger than previous coupled AOGCM simulations (González-Rouco et al., 2003; Ammann et al., 2007; Tett et al., 2007; Jungclaus et al., 2010). Relative to the forced response, the annual internal variability is pronounced even in this atmosphere-only simulation, indicating the potential for constraining the ensemble with additional information.

We analyse simulated near-surface temperature and precipitation over land and several derived indices characterising atmospheric circulation according to Brönnimann et al.



**Fig. 1.** Annual (November to October) average temperature north of 20° N in the ensemble of ECHAM5.4 simulations as anomalies of the 1961–1990 mean. The grey area denotes the range of the first 29 individual simulations; the thick black line is the corresponding ensemble average and the thin blue line denotes the 30th simulation (the reference).

(2009). The data are aggregated for boreal winter (November to April) and summer (May to October), reflecting the approximate temporal resolution of climate proxies. In order to keep computations tractable, we thinn out the initial model grid by selecting grid boxes only at every third longitude and latitude. The state vector used in the EnSRF approach thus consists of semi-annual temperature and precipitation at 694 locations over land plus four derived indices. These indices include the strength of the northern subtropical jet (SJ), defined as the maximum zonal-mean zonal wind at 200 hPa between the Equator and 50° N, the strength of the Hadley cell (HC), defined as the maximum of the zonal mean meridional stream function at 500 hPa between the Equator and 30° N, the strength of the northern stratospheric polar vortex (z100), defined as the difference in geopotential height at 100 hPa between 75–90° N and 40–55° N, and the dynamic Indian monsoon index (DIMI), defined as the difference in average zonal winds at 850 hPa in the boxes 5–15° N, 40–80° E and 20–30° N, 70–90° E. For further discussion of these indices, please refer to Brönnimann et al. (2009).

From the 405-yr period of simulations from 1601 to 2005, we select the segment of 135 yr from 1871 to 2005 for the analysis. For this period, the simulations were constrained with observed rather than reconstructed SSTs and sea-ice variability as boundary conditions. The spatio-temporal variability of reconstructed SSTs is considerably different from variability in observed SSTs. The results, however, are qualitatively robust to results obtained when assimilating proxy data in the early part of the simulation (not shown).

### 2.2 Pseudo-proxy generation

Of the 30-member initial condition ensemble, we select the 30th simulation as the target or reference time series used for validation, and the remaining 29 simulations represent the unconstrained ensemble. From the temperature time series of the reference simulation, we generate pseudo-proxies at 37

different locations (see Fig. 2). The pseudo-proxy locations are chosen to reflect the distribution of temperature-sensitive proxies over land, such as tree-ring series and ice cores (e.g. Mann et al., 2009). Proxy networks such as collections of tree-ring series in North America and Europe are represented by a single pseudo-proxy.

We use a simple approach to fabricate pseudo-proxy time series following earlier work (Mann and Rutherford, 2002; von Storch et al., 2009): at each proxy location, we add normally distributed white noise to the temperature time series of the reference simulation. The noise variance is scaled to produce signal-to-noise ratios of 0.33 or correlations of 0.5 on average. Due to the limited sample size, the actual correlations range from 0.36 to 0.66 and the signal-to-noise ratio from 0.26 to 0.46, as shown in Fig. 2. The pseudo-proxies are also slightly biased compared to the original series, with normally distributed biases centred at zero and ranging from $-0.4$ to $0.5$ K (not shown). The bias in the pseudo-proxy time series reflects a potential estimation error when calibrating real-world proxy time series. Unlike a real-world situation, however, the noise added to the reference time series does not exhibit spatio-temporal coherence.

## 2.3 Metrics of skill and reliability

We analyse the skill in reconstructing different global and continental-scale indicators. Skill is measured using a mean squared error skill score (Murphy and Epstein, 1989), which is also known as the reduction of error (RE, Cook et al., 1994):

$$RE = 1 - \frac{\sum (x_i^a - x_i^{ref})^2}{\sum (x_i^b - x_i^{ref})^2} \quad (1)$$

$x^a$ and $x^b$ denote the analysis and the unconstrained initial condition simulation, respectively; $x^{ref}$ is the reference simulation (the target). The summation is over $i$, and $i$ counts the different time steps. This skill score ranges from 1 to $-\infty$; positive values indicate that the analysis is closer to the reference simulation in mean square error terms than the unconstrained simulation. As we constrain the full set of simulations, we investigate both the skill score for the ensemble mean and the individual simulations. In doing so, we compare the ensemble mean analysis $\bar{x}^a$ with the unconstrained ensemble mean $\bar{x}^b$, and each individual analysis simulation with its unconstrained counterpart.

Furthermore, we also analyse the change in correlation from the correlation of the unconstrained simulations with the reference simulation to the correlation of the analysis with the reference simulation.

Our method for reconstructing past climate states produces not only a best estimate of the past climate state, but also an associated uncertainty estimate. To assess whether the uncertainty is correctly reflected in the ensemble-based, probabilistic prediction (the analysis), the concept of reliability is
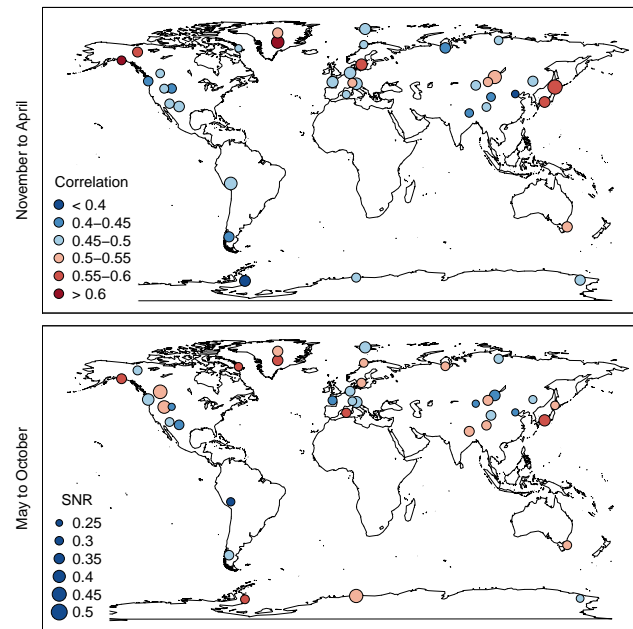


**Fig. 2.** Correlation and signal-to-noise ratio of pseudo-proxies with reference time series in boreal winter (November to April, upper panel) and summer (May to October, lower panel). The colour in both panels reflects the correlation according to the legend in the upper panel; the size of the circles in both panels reflects the signal-to-noise ratio according to the legend in the lower panel.

employed. A probabilistic prediction system is deemed reliable if the frequency of occurrence matches the predicted probability over a large set of predictions. For ensemble-based systems, the analysis rank histogram allows us to assess the reliability of the prediction system (Anderson, 1996).

The rank histogram is produced by computing the rank of the observations of the true climate state (here the reference simulation) compared with the sorted ensemble of predictions (the analysis) for each individual grid box and time step. In a perfectly reliable ensemble, we would expect the observations to fall in each of the $n_{ens} + 1$ classes with equal probability, thus resulting in a flat rank histogram. A U-shaped histogram, in contrast, indicates a negative bias in the analysis variance, i.e. the ensemble is overconfident and the true climate state often lies outside the range of values predicted by the analysis. Correspondingly, a dome-shaped rank histogram denotes an analysis ensemble that overemphasises uncertainty.

To assess the significance of deviations from a flat rank histogram, we provide guidance based on the unconstrained ensemble. Using each of the simulations in turn as reference and the remaining 29 ensemble members to compute the rank histogram, we quantify the effect of sampling uncertainty in a perfectly reliable ensemble.

## 3 Ensemble Square Root Fitting

We use a variant of ensemble Kalman filtering (EnKF, see Evensen, 2003, and references therein) to update model simulations with measurements of the climate system – here pseudo-proxy time series derived from one model simulation. For readers not familiar with EnKF, we provide a short introduction before discussing the modifications and specifics of the approach used in this study. For a more comprehensive introduction, please refer to the above reference.

Data assimilation seeks to provide a best estimate of the true climate state – the analysis. As the problem is not well constrained by the often sparse observations, additional constraints from mathematical models are used to estimate the analysis. Specifically, the models are used to provide a first guess of the true climate state consistent with the boundary conditions. This *prior* estimate of the background climate is then updated with the observations to form the *posterior* distribution of the analysis. The data assimilation can thus be expressed as a Bayesian update problem with

$$P(x^{\mathrm{a}}|y) \sim P(y|x^{\mathrm{b}}) \, P(x^{\mathrm{b}}). \tag{2}$$

Where $P(*|*)$ denotes a conditional probability, $x$ is the climate state in the model and $y$ contains the observations of the climate state. For clarification, the background or prior distribution of the climate state is indicated with the superscript "b" as in $x^{\mathrm{b}}$, and the analysis or posterior is denoted with superscript "a". In addition to providing a first guess of the true climate state, the mathematical model is also used to propagate the analysis for providing a first guess for the next analysis cycle if the data assimilation procedure is cycled. Thereby, the analysis is consistent with all previous and current observations. As discussed earlier, we do not use a cycling of the procedure and therefore, the analysis is constrained by current observations only. In the linear Gaussian case, the above Bayesian update is identical with the Kalman filter update.

In the Gaussian case, we can characterise the background climate at a given time through its mean $\bar{x}^{\mathrm{b}}$ and its covariance matrix $\mathbf{P}^{\mathrm{b}}$, where $\bar{x}^{\mathrm{b}}$ is a vector of length $n$, the dimension of the model state, and $\mathbf{P}^{\mathrm{b}}$ is a matrix of dimension $n \times n$. The observations $y$ of the true state are collected at $m$ distinct points, with $m \ll n$ in general. We account for the fact that $y$ consists of indirect observations of the true state and assume the observation error to follow a zero-mean Gaussian process with covariance $\mathbf{R}$, the observation error covariance. To link model and observation space, we define the operator $\mathbf{H}$ of dimension $m \times n$ that extracts the observations from the model space. $\mathbf{H}$ can be non-trivial in a paleoclimatology context, as this operator reflects the complex dependence of climate proxies, such as tree rings on climate. In our idealised study, however, $\mathbf{H}$ only extracts temperature at proxy locations from the model state $\bar{x}^{\mathrm{b}}$. The development of proxy forward models and corresponding observation operators will be dealt with elsewhere.

The analysis in Kalman filtering is again a multivariate Gaussian, with mean $\bar{x}^{\mathrm{a}}$ and covariance $\mathbf{P}^{\mathrm{a}}$. Following the notation in Whitaker and Hamill (2002), the traditional Kalman filter update equations for the mean and covariance are

$$\bar{x}^{\mathrm{a}} = \bar{x}^{\mathrm{b}} + \mathbf{K}(y - \mathbf{H}\bar{x}^{\mathrm{b}}) \tag{3}$$

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\,\mathbf{H})\,\mathbf{P}^{\mathrm{b}} \tag{4}$$

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{T} + \mathbf{R})^{-1} \tag{5}$$

$\mathbf{K}$ denotes the Kalman gain, a matrix of dimension $n \times m$, and $\mathbf{I}$ is the $n \times n$ identity matrix.

In the ensemble Kalman filter (EnKF), the mean states and covariances are approximated by the sample mean and covariance based on a finite number of simulations. Thus, the background mean in EnKF is $\bar{x}^{\mathrm{b}} = 1/n_{\mathrm{ens}} \sum_k x_k^{\mathrm{b}}$, where $k$ counts the $n_{\mathrm{ens}}$ ensemble members used. Correspondingly, the background covariance is $\mathbf{P}^{\mathrm{b}} = 1/(n_{\mathrm{ens}} - 1) \sum_k (x_k^{\mathrm{b}} - \bar{x}^{\mathrm{b}})^2$. Given that the true mean states and covariances are never known, but estimated, the ensemble approximation is an intuitive and more practically relevant representation of the problem than the general case. Hereafter, mean states $\bar{x}$, covariances $\mathbf{P}$, and the corresponding Kalman gain $\mathbf{K}$ denote the sample-based EnKF representation of these quantities.

In addition to being of more practical relevance than the Kalman filter, the EnKF also allows us to express the filtering problem in a computationally more efficient way. Instead of operating on the full $n \times n$ covariance matrix, the ensemble members are updated individually without explicitly updating the covariances. To reflect the observation error distribution, the observations used to update each individual simulation have to be randomly perturbed according to the observation error covariance. Consequently, EnKF is biased due to sampling uncertainty in both the background covariance $\mathbf{P}^{\mathrm{b}}$ estimated from the ensemble of model simulations and the observation perturbations. Due to the nonlinear dependence of the analysis covariance $\mathbf{P}^{\mathrm{a}}$ on the background covariance $\mathbf{P}^{\mathrm{b}}$, $\mathbf{P}^{\mathrm{a}}$ will be biased low and therefore underestimate ensemble mean errors on average. This underestimate of $\mathbf{P}^{\mathrm{a}}$ can lead to filter divergence and will result in an overly confident analysis in general. The perturbation of observations also increases sampling error and leads to the analysis-error covariance estimate $\mathbf{P}^{\mathrm{a}}$ being less accurate on average. To overcome the above limitations, Whitaker and Hamill (2002) propose a novel approach that does not rely on the perturbation of observations; this approach is referred to as the ensemble square root filter (EnSRF).

Using EnSRF, the update is separated into an ensemble mean update (Eq. 6), which is identical to the EnKF update and an update of the anomalies about the ensemble mean (Eq. 7; see Whitaker and Hamill 2002). Thus, we decompose the background state $x^{\mathrm{b}}$ into the ensemble mean background state $\bar{x}^{\mathrm{b}}$ and the deviation from the ensemble mean $x'^{\mathrm{b}}$ and express the update equations as follows:

$$\bar{x}^{a} = \bar{x}^{b} + \mathbf{K}(\bar{y} - \mathbf{H}\bar{x}^{b}) \tag{6}$$

$$x'^{a} = x'^{b} + \tilde{\mathbf{K}}(y' - \mathbf{H}x'^{b}) = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})x'^{b} \text{ with: } y' = 0 \tag{7}$$

The Kalman gain matrix $\mathbf{K}$ is identical to the gain matrix in the classical EnKF approach as shown in Eq. 5. The gain matrix for the ensemble anomalies, $\tilde{\mathbf{K}}$, is expressed as follows:

$$\tilde{\mathbf{K}} = \mathbf{P}^{b}\mathbf{H}^{T}\left[\left(\sqrt{\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}}\right)^{-1}\right]^{T} \times$$
$$\left(\sqrt{\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}} + \sqrt{\mathbf{R}}\right)^{-1} \tag{8}$$

In our implementation, the background state $x^{b}$ is a vector of length $n = 1392$, consisting of semi-annual temperature and precipitation over land at 694 grid boxes and 4 derived indices. We assimilate pseudo-proxies at $m = 37$ locations. As the pseudo-proxies are generated from the true climate state (the reference simulation) by adding white noise, the observation error variances are known and the covariance matrix $\mathbf{R}$ is diagonal. Therefore, we can update the ensemble serially by including one observation at a time. This greatly enhances the computational tractability of the problem.

### 3.1 Ensemble covariance localisation

Simulating large ensembles of high-resolution GCMs is expensive. Consequently, we have to estimate the background error covariance from a rather limited set of simulations. Here, we estimate the $1392 \times 1392$ dimensional background error covariance matrix $\mathbf{P}^{b}$ from an ensemble of only 29 simulations. The use of a finite ensemble to approximate the background error covariance leads to spurious correlations off the diagonal in $\mathbf{P}^{b}$. These spurious correlations result in small unphysical updates and reduce the analysis variance. In the recursive implementation, this effect will lead to filter divergence. In our non-recursive implementation, the sampling uncertainty leads to an overly confident reconstruction.

Various approaches exist to correct for spurious correlations in the background error covariance and to thereby avoid filter divergence. Covariance inflation (Anderson and Anderson, 1999) and covariance localisation (Houtekamer and Mitchell, 2001) are two commonly used strategies. The former compensates for filter divergence by inflating the background error covariance previous to the update step. The latter is based on the assumption that the correlation between variables decreases with distance between the variables. In this study, we apply a simple covariance localisation to deal with spurious correlations. The localisation reflects our belief that the analysis at each grid box depends more strongly on observed information that is close by than on very distant observations.

To ensure that the localised background error covariance is positive-definite, we use an element-wise product of the sample covariance and a correlation function with local support (see Gaspari and Cohn, 1999, for correlation functions). We redefine $\mathbf{P}^{b}$ used in the EnSRF algorithm above to account for spurious correlations according to Eq. 9:

$$p_{i,j}^{b} = \frac{1}{n_{ens}-1}\sum_{k=1}^{n_{ens}} x'^{b}_{i,k}\, x'^{b}_{j,k}\, \exp\left(-\frac{|d_i - d_j|^2}{2L^2}\right) \tag{9}$$

$p_{i,j}^{b}$ denotes row number $i$ and column number $j$ of $\mathbf{P}^{b}$; $k$ indexes the $n_{ens}$ different ensemble members. $|d_i - d_j|$ is the distance in km between grid box $i$ and grid box $j$, and $L$ is the cutoff distance at which the sample covariance decreases by 39 %. For the aggregated indices in the state vector, $|d_i - d_j|$ is set to the minimum distance between the respective grid box $i$ and the points in the box used to compute the index corresponding to $j$ (or vice versa). The influence of the choice of $L$ on the analysis is discussed in the following.

## 4 Results

First, we analyse the effect of the covariance localisation to deal with spurious covariances. Figure 3 illustrates the benefits of localisation of $\mathbf{P}^{b}$. Without localisation, skill – measured in terms of mean squared error of the ensemble mean compared with the reference time series (see Eq. 1) – is confined to the regions where proxy data are assimilated; elsewhere, we find negative skill. That is, without localisation, assimilation of pseudo-proxies leads to an "overcorrection" of the ensemble in regions far away from where information is assimilated (Fig. 3a and b). With localisation, skill is less confined to the regions where we assimilate data (Fig. 3c and d), as the closest proxies are given more weight in the data assimilation procedure. For example, we find positive skill almost throughout North America with localisation, whereas without localisation, positive skill is confined to western North America, where proxy information is assimilated. In regions far away from proxy locations, such as Africa or the Amazon Basin, the above mentioned overcorrection disappears, resulting in zero skill.

Covariance localisation does not only help to avoid overcorrection in regions far from the assimilated information. In addition, localisation also helps to avoid filter divergence or, in the case of a non-recursive procedure, overly confident analyses. To assess the dependence of the reliability of the analysis on the cutoff for covariance localisation, we make use of the rank histogram, as introduced in Sect. 2.3. The rank histogram for the analysis without covariance localisation is strongly U-shaped, indicating that this analysis is overly confident in many regions (Fig. 4a). This is consistent with the occurrence of negative skill in areas far from the assimilated information, as shown in Fig. 3. With decreasing cutoff length, the negative bias in the variance of the analysis vanishes and the rank histogram gets flat. Compared to

November to April        May to October
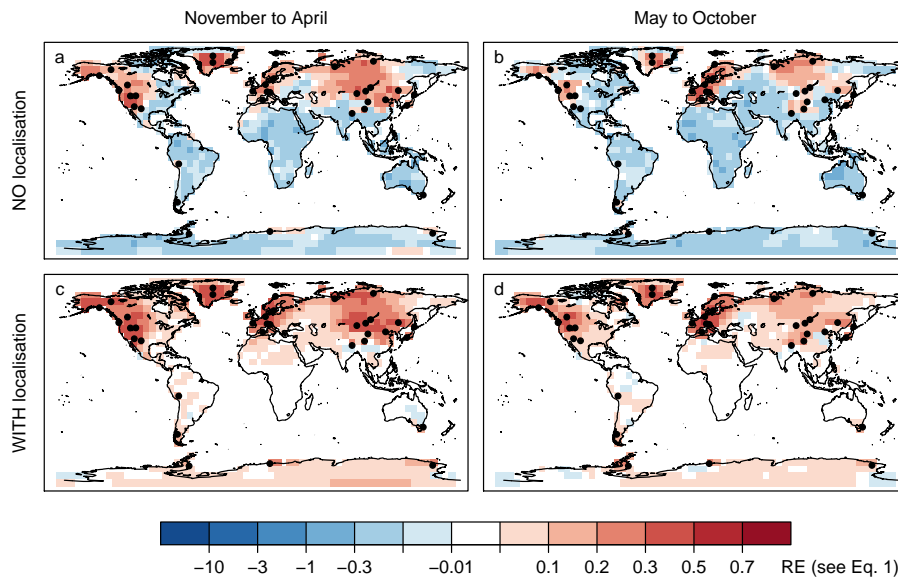


**Fig. 3.** Mean square error skill score (RE) for near-surface temperature of the analysis ensemble mean compared to the unconstrained ensemble mean without (**a** and **b**) and with localisation (**c** and **d**). Results for boreal winter (November to April, **a** and **c**) and for boreal summer (May to October, **b** and **d**). Black dots indicate locations at which pseudo-proxies are assimilated.

the cross-validation with the unconstrained ensemble (black lines in Fig. 4), deviations from a flat rank histogram are indistinguishable from the rank histogram of a perfectly reliable ensemble for cutoff lengths of 2000 km and less. Therefore, we set the cutoff length at 2000 km for all further analyses.

While covariance localisation reduces the negative impact of spurious correlations on the analysis variance, it also affects the sharpness of the analysis. The spread of the ensemble – here expressed as the intra-ensemble standard deviation – indicates the sharpness of the hindcast. In the case of the unconstrained hindcast, the spread represents the uncertainty due to internal variability. In the case of the analysis, we hope to make use of the information about the state of internal variability of the reference, and thus we expect to reduce the hindcast uncertainty, and thereby reduce ensemble spread. The influence of the data assimilation on the ensemble spread for temperature is shown in Fig. 5. The ensemble spread, and thus the uncertainty in the hindcast, is significantly reduced in regions close to the assimilated information (Europe, Central Asia and western North America). As a consequence of the localisation, the spread is not reduced in regions far from the assimilated information (e.g. Sub-Saharan Africa). In addition, data assimilation leads to more wide-spread and larger reductions in ensemble spread in boreal winter than in boreal summer.

In the following figures, the skill scores (Eq. 1) for the respective ensemble means are displayed as arrowheads and the individual simulations as box plots (see Fig. 6). The boxes indicate the interquartile range of the 29 simulations in the analysis; the thick horizontal line indicates the

median simulation, and the whiskers denote the range of the simulations.

In Fig. 6, Northern Hemispheric and northern European land temperature, northern European precipitation, and various circulation indices are analysed in detail. These aggregated indices have been chosen to illustrate the advantages and limitations of the method. We analyse both the mean square error skill score (Eq. 1, Fig. 6a and b) and changes in correlation (Fig. 6c and d) between the unconstrained ensemble and the analysis. The skill score is generally slightly more positive for the individual simulations (boxes in Fig. 6, a and b) than for the ensemble average (arrowheads in Fig. 6). This is due to the fact that the unconstrained ensemble average is – due to its low variance and small bias – an a priori good guess for an additional simulation in mean square error terms. Correlation of the ensemble mean with the reference simulation, however, is generally greatly increased when information is assimilated (see Fig. 6c and d).

We find positive skill scores for most indicators in boreal winter (Fig. 6a). Not surprisingly, skill is strongest in regions that are close to the assimilated information (e.g. northern European temperature and precipitation). However, we find positive skill also for the strength of the northern subtropical jet (SJ) and the stratospheric polar vortex (z100). In boreal summer, skill is slightly reduced but still positive for most of the indicators (shown in Fig. 6). For indicators far from the assimilated information, such as the strength of the northern Hadley cell (HC) or the dynamic Indian monsoon index (DIMI), skill is close to zero as a consequence of the localisation procedure.
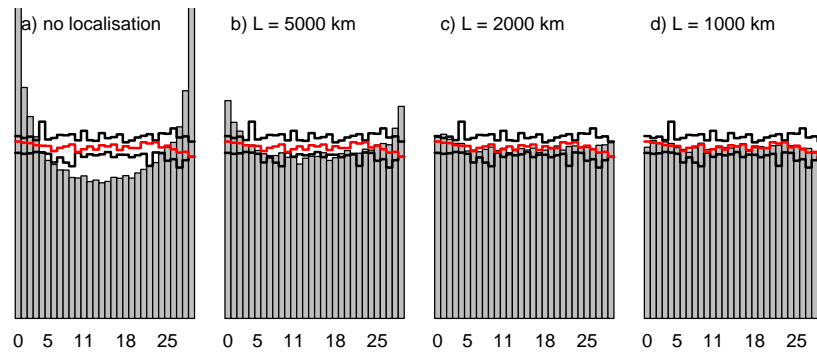
**Fig. 4.** Rank histogram of the reference simulation compared to the analysis using different lengths for the covariance localisation. Without covariance localisation, the spread of the ensemble is negatively biased, leading to a strongly U-shaped rank histogram. The black lines are empirical confidence intervals from a cross-validation with the unconstrained ensemble, using each of the 30 simulations as observations in turn (see text for further discussion). The red line indicates the rank histogram of the reference simulation against the unconstrained ensemble. The scale on the y-axis is the same for all plots, and the frequency of occurrence of ranks 0 and 29 are clipped in **(a)**.
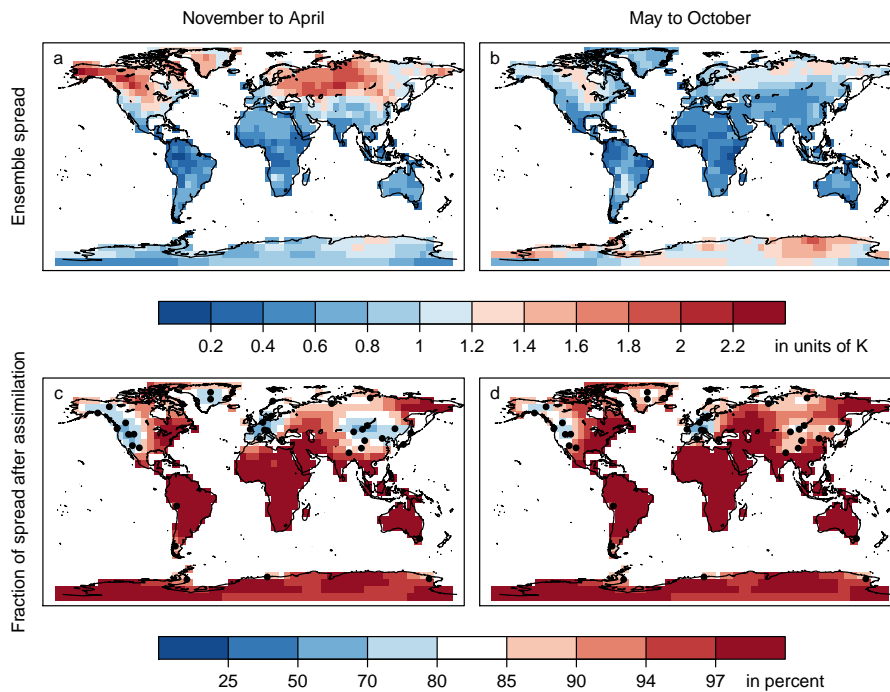


**Fig. 5.** Average intra-ensemble standard deviation (spread) for temperature of the ECHAM ensemble in winter (**a**, November to April) and summer (**b**, May to October). Percentage of the intra-ensemble standard deviation in the analysis ensemble with respect to the unconstrained ensemble for the EnSRF analysis with pseudo-proxies and localisation in (**c**) and (**d**).

Correlation increases considerably with data assimilation for all indicators except HC and DIMI (Fig. 6c). For northern European temperature over land (NEUt2m), correlation of most individual simulations (boxes) increases from close to zero to around 0.5 after assimilation. As with skill, the benefits of data assimilation decrease with increasing distance from the assimilated information. In boreal summer, in contrast, increases in correlation after data assimilation are much more moderate (Fig. 6d).

To test the robustness of the results to the choice of reference simulation, we performed a cross-validation using each individual simulation as reference in turn. Although the choice of reference simulation leads to slight differences in the results, the findings presented here are qualitatively robust to the choice of reference simulation (not shown).

Finally, we investigate the effect of the ensemble size on data assimilation. In EnSRF, the model physics are represented through the error covariance matrix $\mathbf{P}^{b}$, which is
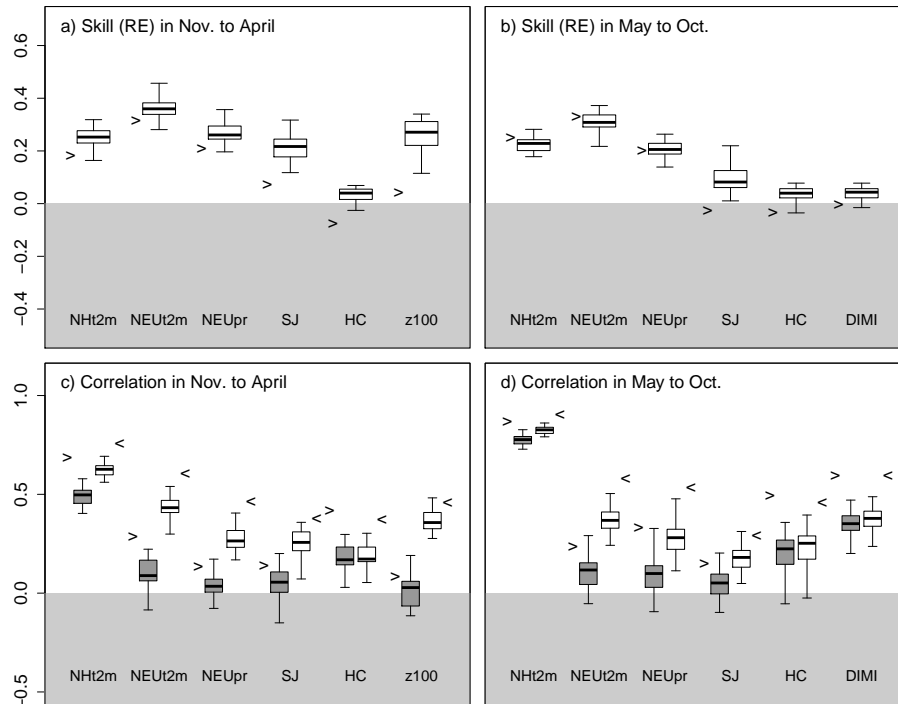
**Fig. 6.** Mean square error skill score (see Eq. 1) for boreal winter and summer in (**a**) and (**b**), and correlation for boreal winter and summer in (**c**) and (**d**) respectively for seven large-scale indicators. The indicators are Northern Hemispheric near-surface temperature over land (NHt2m), northern European temperature (NEUt2m) and precipitation (NEUpr) over land, the strength of the northern subtropical jet (SJ), the northern Hadley cell (HC), the stratospheric polar vortex (z100, boreal winter only), and the dynamic Indian monsoon index (DIMI, boreal summer only). Boxes indicate the interquartile range of skill scores (correlation) for the individual simulations, and the whiskers indicate the range of skill scores(correlation); the arrowheads indicate the skill scores (correlation) of the ensemble mean. In (**c**) and (**d**), the grey boxes and right-facing arrowheads indicate correlation between the unconstrained ensemble and the reference simulation; the white boxes and left-facing arrowheads are the correlation between the simulations after data assimilation and the reference simulation.

estimated directly from the ensemble. Thus, increasing ensemble size allows us to capture more details of the interrelation of variables and its spatial features. In addition, estimation errors decrease with increasing ensemble size. Computation of very large ensembles, however, is very costly, and therefore we would like to learn about minimal requirements for climate reconstructions. Therefore, we run the EnSRF approach with randomly selected sets of 5, 10, 15, 20, 25, and 29 ensemble members and compare the results with the reference simulation. In order to reduce sampling issues, we repeated the experiment 10 times for each ensemble size.

Mean square error skill generally increases with ensemble size for the various indicators (shown in Fig. 7). This increase in skill is moderate for indicators close to the assimilated information, such as mean temperature over land in the Northern Hemisphere or northern European total precipitation (Fig. 7a, b, e and f). In contrast, the increase in skill with increasing ensemble size is considerable for indicators further away from the assimilated information, such as the strength of the subtropical jet (SJ, Fig. 7c and g) or the strength of the stratospheric polar vortex (z100, Fig. 7d). For these indicators, we find positive skill for most of the

individual simulations only with ensembles of size 10 or more. We find simulations that perform well even with small ensembles; the positive effect of increasing ensemble size, however, is clearly visible in reducing the number of simulations with negative skill. For indicators with marginal skill, such as the dynamic Indian monsoon index (DIMI, Fig. 7h), increasing ensemble size reduces the spread in the results for the individual simulations without affecting the overall skill. For all indicators, the ensemble mean skill strongly benefits from large ensembles.

## 5 Discussion

This study illustrates the potential of data assimilation using EnSRF for paleoclimatology. Depending on the indicator of interest, we find considerable skill even when assimilating spatially sparse information with low temporal resolution. Positive skill is not only constrained to the climatic parameters that are assimilated, but it extends to other climatic variables as well. Furthermore, we find positive skill constraining upper air quantities, such as the strength of the
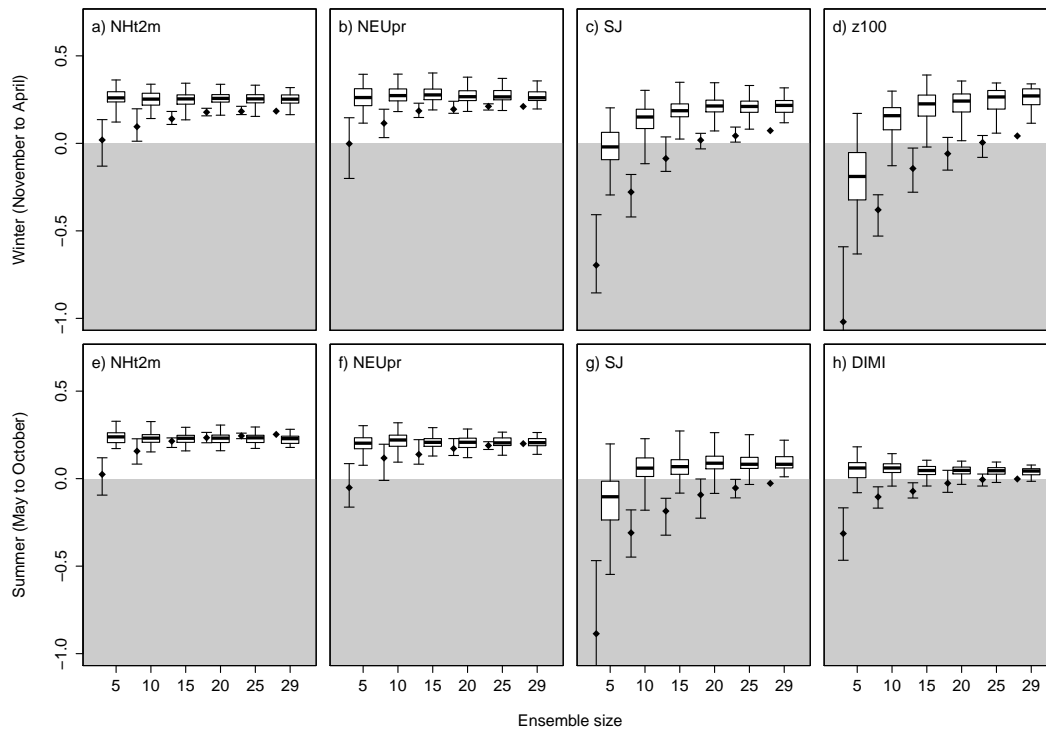
**Fig. 7.** Skill in dependence of ensemble size for different aggregated indicators. The boxes summarise the distribution of the mean squared error skill for individual simulations (10 times the number of ensemble members). The black bars denote the spread of skill for the ensemble mean for the 10 different realisations of varying ensemble size; the black diamond indicates the average ensemble mean skill over these 10 realisations. For 29 ensemble members, there is only one analysis.

northern subtropical jet or the strength of the polar vortex through assimilation of surface quantities (here indirect and thus noisy observations of near-surface temperature).

Skill is generally confined to the Northern Hemisphere. This is a consequence of both the greater number of proxy records and the larger fractional land area in the Northern Hemisphere. As a consequence of the experimental setup (an atmosphere-only GCM), we do not expect large differences over oceans and adjacent land due to the dominant influence of sea surface temperatures (SSTs), which are prescribed in the model simulations. We find strongest positive skill for variables in boreal winter, during which weather in the northern midlatitudes is strongly influenced by large-scale circulation. In boreal summer, when weather is much more dependent on local processes, data assimilation is less beneficial (see Fig. 6). This finding is in line with other studies (Brönnimann and Luterbacher, 2004; Rutherford et al., 2005; Franke et al., 2010; Griesser et al., 2010).

We assimilate semi-annual data and analyse skill both in summer and in winter. The extension of the methodology to be able to assimilate data with higher (monthly) or lower (annual to decadal) temporal resolution is straightforward. Most temperature-sensitive climate proxies, such as tree rings, reflect summer temperatures; however, we assess skill also for the winter half-year in order to explore the potential benefits

of assimilating early instrumental observations and documentary evidence.

The skill metric presented here reflects value added to the initial condition ensemble by the data assimilation. The results are thus not comparable with previous studies making use of pseudo-proxies (Mann and Rutherford, 2002; von Storch et al., 2004; Bürger et al., 2006). In the following, we highlight the most important difference between the study presented here and earlier work involving pseudo-proxies. The crucial element of empirical climate reconstructions is to establish the relationship between proxy records and certain climatic features (e.g. local climate or large-scale patterns) in the calibration period. Pseudo-proxy analyses have been used to investigate how well these relationships can be extrapolated to characterise past climates (see Rutherford et al., 2005; Bürger et al., 2006; Mann et al., 2007; Christiansen et al., 2009, for a discussion of different reconstruction methods). In the data assimilation framework, the proxy-climate relationship is characterised by the observation operator $\mathbf{H}$ and the update of the unconstrained model simulations further depends on the observation error covariance $\mathbf{R}$ and the background error covariance $\mathbf{P}^b$. As we are interested in quantifying the skill emerging from the assimilation of spatially sparse information with low temporal resolution, we do not touch on the issue of how to best estimate $\mathbf{H}$ and

**R**. Instead, we focus on the differences between an unconstrained ensemble and the analysis after data assimilation. Nevertheless, we recognise that correct formulation of forward proxy models and their related observation operators is crucial for real-world applications of the data assimilation procedure for climate reconstruction, and we are currently working on this issue.

Correlations between individual simulations and the reference simulation improve considerably after assimilation of pseudo-proxies (Fig. 6c and d). This indicates that we can indeed use data assimilation to constrain internal variability. It is noteworthy that positive correlations occur also in the unconstrained simulations (grey boxes and right-facing arrows), indicating that the individual ensemble members co-vary with the reference simulation. This is due to the deterministic response to changing boundary conditions, as illustrated for annual average temperature north of $20°$ N in Fig. 1. Due to the strong anthropogenic forcing during the twentieth century, we find positive correlations for most indicators. Co-variability in the unconstrained ensemble is reduced for indicators aloft such as z100 and SJ, but also for non-thermal indicators such as northern European precipitation (NEUpr). This illustrates that the deterministic response to varying boundary conditions is weak compared to internal variability for these indicators. The dominance of internal variability in turn highlights the potential benefits of data assimilation approaches.

Our ensemble of analyses also indicates combined model and proxy reconstruction uncertainty. In this idealised setup, the analysis ensemble spread directly measures reconstruction uncertainty. In a real-world application, however, interpretation of reconstruction ensemble spread will be complicated by the fact that the model provides an imperfect representation of the true physics and dynamics of the system.

To ensure that the analysis ensemble reliably captures the uncertainty, we apply a covariance localisation. The localisation uses horizontal distance to artificially reduce correlation and thus suppress the influence of spurious correlation arising from the small ensemble size used to estimate the correlation. This seems to work well for surface quantities (e.g. near-surface temperature and precipitation). Nevertheless, we cannot rule out the possibility that our localisation procedure suppresses real, far-reaching correlations (e.g. teleconnections) and that we thus unintentionally reduce skill in areas far away from the assimilated information. Given the issue of "overcorrection" (Fig. 3) and underestimation of reconstruction uncertainty (Fig. 4) without localisation, we consider the potential reduction in skill due to overly restrictive localisation to be a conservative approach. Several authors developed adaptive approaches to allow for spatially and temporally more complex patterns of influence (see Anderson, 2007; Bishop and Hodyss, 2007; Fertig et al., 2007). While these adaptive approaches are potentially useful to overcome the problem described above, their implementation

is much less straightforward and beyond the scope of this study.

Furthermore, we investigate the effect of ensemble size on our ability to successfully constrain the simulations with the available proxy information (see Fig. 7). Both the ensemble mean and simulations with negative skill benefit most from increasing ensemble size. For indicators close to the assimilated information, small ensembles are sufficient to represent the relationship between proxies and the respective indicator. For indicators that are less directly related to and/or further away from the assimilated information, large ensembles help to better specify the relation between proxies and indicators. We note, however, that the covariance localisation is dependent on the ensemble size, as its goal is to account for increasing sampling errors with decreasing ensemble size. For an infinite ensemble, no covariance localisation is needed. We do not optimise the cutoff length for different ensemble sizes here. We conclude that while EnSRF with ensembles as small as 15 ensemble members leads to considerable skill in regions close to the assimilated information, larger ensembles are needed to reduce uncertainty in areas further away and for variables that are less directly connected to the assimilated proxy information.

Finally, we would like to touch on more general limitations arising from the experimental setup. By using an atmosphere-only GCM, we restrict climate to closely follow reconstructed boundary conditions. These reconstructions, in turn, are themselves uncertain. It would thus be desirable to allow for uncertainties in the boundary conditions as well. We refrain from perturbing boundary conditions, as such an ensemble would not allow us to properly investigate the strengths and limitations of the non-recursive data assimilation approach due to severe sampling issues. Instead, our experimental setup, and the thus resulting ensemble, offers us the opportunity to develop our capabilities in assimilating proxy data (this study) and in formulating proxy forward models (on-going work) and to understand the respective impacts on our ability to reconstruct climate.

It is important to note, however, that the use of an atmosphere-only GCM limits the potential skill of a data assimilation approach rather than overemphasising it. Neglecting additional sources of uncertainty, such as forcing and parameter uncertainty (not explored here), reduces the variability of the unconstrained ensemble. This will generally lead to conservative estimates of the assimilation skill, as the unconstrained ensemble is already close to the reference simulation. In a study with a coupled ocean-atmosphere GCM, data assimilation should lead to more considerable improvements than documented here.

The natural extension of our approach would be to assimilate data in a coupled Earth system model to better quantify our uncertainty about past climates. Such an experimental setup, however, requires on-line data assimilation, as the temporal limit for predictability of slowly varying parts of the Earth system such as the ocean or the land surface exceeds

the temporal resolution of the assimilated information. While such a coupled Earth system model with data assimilation is our final goal, we again stress the importance of developing the capabilities required to set up and run such a model with a simpler and controllable experimental setup.

## 6    Conclusions

Data assimilation provides a third alternative to the traditional empirical methods for climate reconstructions and purely model-based approaches (see Jansen et al., 2007, for a review of recent advances). We conclude that ensemble square root filtering (EnSRF) is a promising way to reconstruct past climates. Previously, the technique has been successfully applied in the twentieth century reanalysis project (Compo et al., 2011). Here, we show that data assimilation through EnSRF is beneficial, even when assimilating much sparser information with low temporal resolution and with considerable measurement errors. This approach extends previous suggestions for data assimilation in paleoclimatology to an ensemble-based approach with a high-resolution GCM.

We assimilated temperature-sensitive pseudo-proxies with semi-annual resolution at 37 locations mainly in the Northern Hemisphere. Thereby, we managed to reduce the spread of the unconstrained ensemble – and thus our uncertainty about past climate – by up to 50 % for near-surface temperature in areas close to the assimilated information. For parameters other than near-surface temperature such as total precipitation, assimilation of temperature proxies is less beneficial, but we still find positive skill. Furthermore, positive skill is not only constrained to near-surface quantities, but we find value added through data assimilation also for indicators of extratropical and subtropical circulation.

A crucial element of the data assimilation procedure is the background error covariance localisation. This reduces "overcorrection" in areas far away from the assimilated information by giving distant information less weight. In addition, localisation also helps to produce a reliable analysis; that is, localisation ensures that the analysis spread captures the actual uncertainty. The effect of the localisation, however, is most obvious in regions far away from the assimilated information where we find negative skill and an overly confident analysis without the localisation. In these areas, the localisation inhibits spurious correlation with distant proxies to incorrectly constrain the climate model simulations.

The use of an ensemble-based approach to reconstructions allows us to express the uncertainty about past climate states in a natural way. Whereas intra-ensemble spread in the initial-condition ensemble indicates how well the past climate state is constrained by the boundary conditions, the change in spread from the unconstrained ensemble to the analysis can be used to assess the value added through the assimilation of climate proxy information.

## References

Ammann, C. M., Joos, F., Schimel, D. S., Otto-Bliesner, B. L., and Tomas, R. A.: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, P. Natl. Acad. Sci. USA, 104, 3713–3718, 2007.

Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, J. Climate, 9, 1518–1530, 1996.

Anderson, J. L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, Physica D, 230, 99–111, 2007.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, 1999.

Bishop, C. H. and Hodyss, D.: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation, Q. J. Roy. Meteor. Soc., 133, 2029–2044, doi:10.1002/qj.169, 2007.

Brönnimann, S. and Luterbacher, J.: Reconstructing northern hemisphere upper-level fields during World War II, Clim. Dynam., 22, 499–510, doi:10.1007/s00382-004-0391-3, 2004.

Brönnimann, S., Stickler, A., Griesser, T., Fischer, A. M., Grant, A., Ewen, T., Zhou, T. J., Schraner, M., Rozanov, E., and Peter, T.: Variability of large-scale atmospheric circulation indices for the northern hemisphere during the past 100 years, Meteorol. Z., 18, 379–396, 2009.

Bürger, G., Fast, I., and Cubasch, U.: Climate reconstruction by regression – 32 variations on a theme, Tellus A, 58, 227–235, doi:10.1111/j.1600-0870.2006.00164.x, 2006.

Christiansen, B., Schmith, T., and Thejll, P.: A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness, J. Climate, 22, 951–976, doi:10.1175/2008JCLI2301.1, 2009.

Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D.: Feasibility of a 100-year reanalysis using only surface pressure data, B. Am. Meteorol. Soc., 87, 175–188, 2006.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. Roy. Meteor. Soc., 137, 1–28, doi:10.1002/qj.776, 2011.

Cook, E. R., Briffa, K. R., and Jones, P. D.: Spatial regression methods in dendroclimatology – A review and comparison of 2 tech-

niques, Int. J. Climatol., 14, 379–402, 1994.

Cook, E. R., D'Arrigo, R., and Anchukaitis, K.: ENSO reconstructions from long tree-ring chronologies: Unifying the differences?, talk presented at a special workshop on "Reconciling ENSO Chronologies for the Past 500 Years", held in Moorea, French Polynesia on 2–3 April, 2008.

Crowley, T., Zielinski, G., Vinther, B., Udisti, R., Kreutz, K., Cole-Dai, J., and Castellano, E.: Volcanism and the Little Ice Age, PAGES newsletter, 22–23, 2008.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean. Dynam., 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Fertig, E. J., Hunt, B. R., Ott, E., and Szunyogh, I.: Assimilating non-local observations with a local ensemble Kalman filter, Tellus A, 59, 719–730, doi:10.1111/j.1600-0870.2007.00260.x, 2007.

Franke, J., González-Rouco, J., Frank, D., and Graham, N.: 200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction analog methods, Clim. Dynam., 1–18, doi:10.1007/s00382-010-0802-6, 2010.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, 1999.

González-Rouco, F., von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, Geophys. Res. Lett., 30, 165–184, doi:10.1029/2003GL018264, 2003.

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, Clim. Dynam., 27, 165–184, 2006.

Goosse, H., Mann, M. E., and Renssen, H.: Climate of the past millennium: combining proxy data and model simulations, Wiley Blackwell, online first: doi:10.1002/9781444300932.ch7, 2009.

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, J. Geophys. Res., 115, D09108, doi:10.1029/2009JD012737, 2010.

Griesser, T., Brönnimann, S., Grant, A., Ewen, T., Stickler, A., and Comeaux, J.: Reconstruction of global monthly upper-level temperature and geopotential height fields back to 1880, J. Climate, 23, 5590–5609, doi:10.1175/2010JCLI3056.1, 2010.

Hargreaves, J. C. and Annan, J. D.: Assimilation of paleo-data in a simple Earth system model, Clim. Dynam., 19, 371–381, doi:10.1007/s00382-002-0241-0, 2002.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.

Hughes, M. K. and Ammann, C. M.: The future of the past-an Earth system framework for high resolution paleoclimatology: editorial essay, Climatic Change, 94, 247–259, 2009.

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.: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, chap. Paleoclimate, 433–498, Cambridge Uni-

versity Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, AMOJ, 58, 233–248, 2009.

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, Clim. Past, 6, 723–737, doi:10.5194/cp-6-723-2010, 2010.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, Cambridge, 2003.

Koch, D., Jacob, D., Tegen, I., Rind, D., and Chin, M.: Tropospheric sulfur simulation and sulfate direct radiative forcing in the Goddard Institute for Space Studies general circulation model, J. Geophys. Res.-Atmos., 104, 799–822, 1999.

Krivova, N. A., Balmaceda, L., and Solanki, S. K.: Reconstruction of solar total irradiance since 1700 from the surface magnetic flux, Astro. Astrophys., 467, 335–346, 2007.

Lean, J.: Evolution of the sun's spectral irradiance since the Maunder Minimum, Geophys. Res. Lett., 27, 2425–2428, 2000.

Lorenz, E. N.: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289–307, 1969.

Mann, M. E. and Rutherford, S.: Climate reconstruction using "pseudoproxies", Geophys. Res. Lett., 29, 139-1–139-4, doi:10.1029/2001GL014554, 2002.

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Robustness of proxy-based climate field reconstruction methods, J. Geophys. Res.-Atmos., 112, D12109, doi:10.1029/2006JD008272, 2007.

Mann, M. E., Woodruff, J. D., Donnelly, J. P., and Zhang, Z. H.: Atlantic hurricanes and climate over the past 1500 years, Nature, 460, 1256–1260, 2009.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlen, W.: Highly variable northern hemisphere temperatures reconstructed from low- and high-resolution proxy data, Nature, 433, 613–617, doi:10.1038/nature03265, 2005.

Murphy, H. A. and Epstein, E. S.: Skill scores and correlation coefficients in model verification, Mon. Weather Rev., 117, 572–581, 1989.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochem. Cy., 22, GB3018, doi:10.1029/2007GB003153, 2008.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyzes of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res.-Atmos., 108, 4407, doi:10.1029/2002JD002670, 2003.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM5 Part I: model description, Tech. rep. 349, Max Planck Institute for Meteorology, 2003.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, A.:

The atmospheric general circulation model ECHAM5 Part II: Sensitivity of simulated climate to horizontal and vertical resolution, Tech. rep. 354, Max Planck Institute for Meteorology, 2004.

Rutherford, S., Mann, M. E., Osborn, T. J., Briffa, K. R., Jones, P., Bradley, R. S., and Hughes, M. K.: Proxy-based northern hemisphere surface temperature reconstructions: sensitivity to method, predictor network, target season, and target domain, J. Climate, 18, 2308–2329, doi:10.1175/JCLI3351.1, 2005.

Tett, S. F. B., Betts, R., Crowley, T. J., Gregory, J., Johns, T. J., Jones, A., Osborn, T. J., Öström, E., Roberts, D. L., and Woodage, M. J.: The impact of natural and anthropogenic forcings on climate and hydrology since 1550, Clim. Dynam., 28, 3–34, doi:10.1007/s00382-006-0165-1, 2007.

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

von Storch, H., Cubasch, U., González-Rouco, J., Jones, J., Voss, R., Widmann, M., and Zorita, E.: Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (DATUN), in: 11th Symposium on Global Change Studies, American Meteorological Society, 2000.

von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F., and Tett, S. F. B.: Reconstructing past climate from noisy data, Science, 306, 679–682, 2004.

von Storch, H., Zorita, E., and González-Rouco, F.: Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, Int. J. Earth. Sci., 98, 67–82, 2009.

Wang, Y.-M., Lean, J. L., and N. R. Sheeley, J.: Modeling the sun's magnetic field and irradiance since 1713, Astrophys. J., 625, 522–538, 2005.

Wanner, H., Beer, J., Bütikofer, J., Crowley, T. J., Cubasch, U., Flückiger, J., Goosse, H., Grosjean, M., Joos, F., Kaplan, J. O., Küttel, M., Müller, S. A., Prentice, I. C., Solomina, O., Stocker, T. F., Tarasov, P., Wagner, M., and Widmann, M.: Mid- to late holocene climate change: an overview, Quaternary Sci. Rev., 27, 1791–1828, 2008.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.

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, Clim. Past, 6, 627–644, doi:10.5194/cp-6-627-2010, 2010.

Yoshimori, M., Raible, C. C., Stocker, T. F., and Renold, M.: Simulated decadal oscillations of the Atlantic meridional overturning circulation in a cold climate state, Clim. Dynam., 34, 101–121, 2010.