200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction analog method

Jörg Franke · J. Fidel González-Rouco ·
David Frank · Nicholas E. Graham

Abstract Spatially resolved climate reconstructions are commonly derived from long instrumental series and proxy data via linear regression based approaches that use the main modes of the climate system. Such reconstructions have been shown to underestimate climate variability and are based upon the assumption that the main modes of climate variability are stationary back in time. Climate models simulate physically consistent climate fields but cannot be taken to represent the real past climate trajectory because of their necessarily simplified scope and chaotic internal variability. Here, we present sensitivity tests of, and a 200-year temperature reconstruction from, the PSR (Proxy Surrogate Reconstruction) method. This method simultaneously capitalizes on the individual strengths of instrumental/proxy data based reconstructions and model simulations by selecting the model states (analogs) that are most similar with proxy/instrumental data available at specific places and specific moments of time. Sensitivity experiments reveal an optimal PSR configuration and indicate that 6,500 simulation years of existing climate models provide a sufficient pool of possible analogs to skillfully reconstruct monthly European temperature fields during the past 200 years. Reconstruction verification based upon only seven instrumental stations indicates potential for extensions back in time using sparse proxy data. Additionally the PSR method allows evaluation of single time series, in this case the homogeneity of instrumental series, by identifying inconsistencies with the reconstructed climate field. We present an updated European temperature reconstruction including newly homogenized instrumental records performed with the computationally efficient PSR method that proves to capture the total variance of the target.

Keywords Climate reconstruction · Europe ·
Analog method · Paleoclimate · Proxy data ·
Instrumental data

J. Franke (✉)
Swiss Federal Research Institute WSL, Zürcherstrasse 111,
8903 Birmensdorf, Switzerland
e-mail: joerg.franke@wsl.ch

J. F. González-Rouco
Universidad Complutense, Madrid, Spain

D. Frank
Swiss Federal Research Institute WSL,
Birmensdorf, Switzerland

N. E. Graham
Hydrologic Research Center, San Diego, CA, USA

N. E. Graham
Scripps Institution of Oceanography, La Jolla, CA, USA

1 Introduction

The range of natural climate variability during the past centuries to millennia is highly debated in the context of on-going global warming. Therefore, many efforts have focused on the reconstruction of climate, mainly temperature, in the pre-instrumental era from local to hemispheric or even global scales (e.g., Esper et al. 2002; Jones and Mann 2004; Hegerl et al. 2006; Juckes et al. 2007; Mann et al. 2008; Frank et al. 2010). Such reconstructions are based on long-instrumental measurements, documentary archives and proxy data from tree rings, ice cores, stalagmites, corals, pollen, boreholes, lake and ocean sediments (Jones et al. 2009, and references therein). As these archives are of a different nature and have undergone

Springer

various analytical and statistical treatments, resulting reconstructions may reflect a limited part of the climate history often conditioned to specific seasons, regions or emphasizing particular frequencies to which the proxy is sensitive or that the reconstruction approach is able to recover (Jones et al. 2009).

Reconstructions of sub-continental to global scales can be divided into two different methodological groups by either reconstructing a single time-series representative for some spatial domain (e.g., Composite Plus Scaling; CPS), or by simultaneously reconstructing numerous time-series that subdivide a spatial domain into a grid (e.g., Climate Field Reconstruction; CFR). CFR become increasingly relevant because they allow changes in the modes of climate variability to be identified (Mann et al. 2009). Most CFR approaches are based on representing both the instrumental and proxy fields in terms of their dominant modes of variability (i.e., the principal components) that are related to each other via multiple regressions that may be of the direct (e.g., Luterbacher et al. 2004) or inverse formulation (e.g., Mann et al. 1998). In some cases an stepwise procedure to iteratively estimate the covariance matrix is used to make a more exhaustive use of the available information as in the RegEM method (e.g., Mann et al. 2007, 2008, 2009). Applications at the hemispheric and global scales have targeted mostly temperature reconstructions, though precipitation, drought and pressure reconstructions have also been performed at continental scales (e.g., Luterbacher et al. 2002; Cook et al. 2004; Zhang et al. 2004; Pauling et al. 2006; Küttel et al. 2010).

Due to the complexity of some of the CFR methods, performance assessment has recently been made using so-called pseudoproxies (Mann and Rutherford 2002), from which it has been suggested that many of these regression based methods may underestimate low-frequency variability (Zorita et al. 2003; von Storch et al. 2004; Bürger and Cubasch 2005; Bürger et al. 2006; Raible et al. 2006; Smerdon and Kaplan 2007; Smerdon et al. 2008; Lee et al. 2008; Riedwyl et al. 2009; Christiansen et al. 2009; Tingley and Huybers 2010a, b). Additionally, the principal component regression methods are broadly based upon assumptions that the patterns identified during the calibration period are stationary back in time. This concern is largely stated as a theoretical limitation, but is generally assumed to be of minor significance during the Holocene (e.g., Mann et al. 2005). However, Cook et al. (2002) detailed strong biases that resulted in an earlier reconstruction of the North Atlantic Oscillation due to the changing signatures of this mode over Scandinavia.

While reconstructions attempt to quantify to true past coarse of climate variability General Circulation Models (GCM) are a tool permitting the basic properties and relationships within the climate system to be investigated.

Driving a GCM by estimates of past changes in natural (i.e., solar, volcanic) and anthropogenic forcings (i.e., land-use, aerosols, greenhouse gases) provides a modeled estimate for past climate variability (e.g., González-Rouco et al. 2003; Stendel et al. 2006; Tett et al. 2007; Ammann et al. 2007). Although these simulations are comprehensive and physically consistent they cannot be regarded as a representation of the real past climate trajectory in the same sense as climate reconstructions. Unknown initial states of the oceans, insufficient resolution to capture important processes such as convective rainfall, and uncertainties in both the forcing series and necessary simplifications of the basic climate system and processes are among the factors that limit the ability of models to accurately represent past changes in the climate system. In particular high frequency climate variability is difficult to simulate because natural internal (annual to decadal) variability dominates over the imposed forcing (Collins et al. 2001).

Proxy data have been compared with model simulations in many recent articles (e.g., Jansen et al. 2007; Franke et al. 2008; González-Rouco et al. 2009) to assess the degree of agreement, but there have been few attempts to combine model and proxy data to achieve improved paleoclimatic reconstructions. One approach are data-assimilation techniques in which atmospheric states of climate models are controlled by observations. Only a few studies have focused on the adaption of this technique, developed in the field of weather forecasting, for paleoclimatic purposes using sparse proxy networks and data with an annual or lower temporal resolution (von Storch et al. 2000; van der Schrier and Barkmeijer 2005; Goosse et al. 2006; Crespin et al. 2009; Widmann et al. 2009; Goosse et al. 2010). Data-assimilation approaches have great potential as they simulate multi-variate climate fields that are consistent with proxy reconstructions. However, they are computationally too intensive to be run with state of the art GCM's and applied models of intermediate complexity do not produce sufficient high frequency climate variability to resolve modes of climate variability such as El Nino Southern Oscillation (Petoukhov et al. 2005).

Methods that search for analogs in existing databases are much less cost-intensive and thus more prevalent in many scientific disciplines. In climatology they have been successfully used for instance in weather forcasting (Lorenz 1969; Kruizinga and Murphy 1983), short-term climate prediction (Barnett and Preisendorfer 1978; van den Dool 1994) or statistical downscaling (Zorita et al. 1995; Cubasch et al. 1996; Zorita and von Storch 1999). In form of model based climate reconstructions this approach was introduced by Graham et al (2007): the so-called Proxy Surrogate Reconstruction (PSR) is an analog method in which climate model simulations are reordered along the

time axis to obtain temporal agreement with actual proxy data (Graham et al. 2007). The application of PSR in the context of paleoclimate reconstructions brings several beneficial aspects: (1) the method has been shown not to underestimate variability in the context of a downscaling approach (Zorita and von Storch 1999); (2) the method can draw upon the pool of model simulated fields already produced for the last Millennium (González-Rouco et al. 2003; Tett et al. 2007; Ammann et al. 2007; Jungclaus et al. 2010) to find analogs for past climate states; (3) this application enjoys benefits of blending model and instrumental/proxy data in a computationally efficient manner. PSR has so far been employed to infer large-scale climatic changes (Graham et al. 2007; Trouet et al. 2009) and to build composites of European mean climate in extreme years (Luterbacher et al. 2010, accepted) for specific periods in the past. While these studies validated the basic PSR method, many basic performance assessments and methodological subtleties remain only superficially explored. In this manuscript, we (1) provide the first systematical exploration the basic PSR approach, and (2) develop a reconstruction of temperatures across Europe.

The remaining climatological and methodological challenges include (1) the arguable breakup of continuity which occurs when the model output is temporally re-shuffled (see Sect. 4.6) and (2) the difficulty in finding an appropriate analog given the large number of degrees of freedom in the climate system (van den Dool 1994) is discussed in Sect. 4.1. After performing these sensitivity tests and methodological optimization, we utilize long-instrumental records from across Europe to constrain modeled fields and thus provide a spatially resolved reconstruction of monthly temperatures across Europe for the past 200-years. This period and region is interesting because (1) numerous long high-quality instrumental data are available, (2) this represents the transition from the pre-industrial to industrial era, and (3) on-going efforts are producing new high-quality proxy data for this region and make a potential extension appealing. As the number of instrumental and proxy data decreases back in time, another aim of this study is to explore the possibility to conduct a reconstruction using a limited set of time series. By testing a number of series in the order of magnitude of available temperature proxies for the last Millennium, the exercise establishes the groundwork for longer reconstructions in the future.

2 Data

For the application of the PSR method, station temperature data serve as a target to match and select modeled fields of temperature across Europe. The instrumental data are provided by Parker et al. (1992), Peterson et al. (1998), Moberg et al. (2002), Jones and Moberg (2003), Auer et al. (2007) and Böhm et al. (2010) and cover roughly the past 200 years although some time series have gaps, especially in the early period. Most of these stations have been used to produce gridded instrumental data sets such as the CRU TS 2.1 (Mitchell and Jones 2005), but in the Greater Alpine Region we include data that have been newly homogenized to correct biases related to poor radiation sheltering in the early instrumental period (Frank et al. 2007; Böhm et al. 2010). In total up to 27 time series are used in this study (Table 1). However, the majority of experiments are conducted with a seven station subset to provide results for situations that would be comparable to that of proxy reconstructions in which the number of predictors is comparatively low (see Sect. 4.4). These seven time series are presented in Fig. 1.

In order to validate the reconstruction, results are compared to the ERA-40 reanalysis data of 2.5° resolution (ECMWF 2006), the 5° gridded instrumental dataset HADCRUT3v (Brohan et al. 2006), and the 0.5° gridded temperature reconstruction by Luterbacher et al. (2004, updated version based on instrumental data of Mitchell and Jones 2005). After 1901 the Luterbacher dataset is identical to CRU TS 2.1 (Mitchell and Jones 2005). Hereafter the entire dataset is referred to as LUT.

Model simulations of the last millennium serve as a pool of possible analog climate states (see Table 2). The pool consists of three 1,000 year-long simulations of the ECHO-G model: one control run and two simulations (called Erik1 and Erik2) externally forced by changes of solar irradiance, greenhouse gases and the radiative effect of volcanic eruptions (González-Rouco et al. 2006; Zorita et al. 2007); three 1,000 year-long simulations of the COSMOS model (Jungclaus et al. 2010): one control run and two simulations similarly forced as ECHO-G but using updated forcings for total solar irradiance (Krivova et al. 2007, updated), greenhouse gases (CO_2 calculated in the model, CH_4 and N_2O, MacFarling Meure et al. (2006)) and volcanic activity (Crowley, unpubl.), natural and anthropogenic aerosols (Tanré et al. 1984; Lefohn et al. 1999) and additionally including land-use changes (Pongratz et al. 2008). Finally one forced simulation (solar irradiance, greenhouse gases, volcanic aerosols, anthropogenic aerosols and land-use/cover changes) from the HADCM3 model for the past 500 years completes the pool of analogs (Tett et al. 2007).

All gridded datasets and simulations are interpolated to the T30 model resolution (ca. 3.75° in longitude and latitude) of the ECHO-G and COSMOS model. The simulations reproduce various ranges of low frequency variability in response to different climate sensitivities, forcing specifications, and initial conditions (Osborn and Briffa 2006; Jansen et al. 2007; Zorita et al. 2007). For

Table 1 Stations with long temperature series which have been used in the experiments with 27, 7, 3 stations spread over Scandinavia and central Europe (3S), and 3 stations concentrated in central Europe (3C)

Location	Lon	Lat	Period	27	7	3S	3C
Archangel	40.53	64.55	1813–2001	x	x		
Bergen	05.40	60.37	1816–1999	x			
Berlin	13.40	52.47	1701–2008	x			
Budapest	19.27	47.43	1780–2008	x	x	x	x
De Bilt	05.18	52.10	1706–2008	(x)			
Central England	−0.60	51.50	1659–2008	x	x		
Geneva	06.15	46.19	1760–2008	x	x	x	x
Kiev	30.52	50.45	1812–2001	x			
Copenhagen	12.55	55.68	1768–1991	x			
Kremsmünster	14.13	48.05	1767–2008	x			
Moscow	37.60	55.75	1779–2001	x			
Munich	11.55	48.17	1781–2008	x			
Olso	10.75	59.95	1816–2001	x			
Padova	11.88	45.40	1774–2008	x			x
Paris	02.33	48.87	1757–1995	x			
Petersburg	30.33	59.93	1743–2008	x			
Riga	24.13	56.97	1795–1989	x			
Rome	12.50	41.90	1811–1996	x			
San Fernando	−6.20	36.47	1787–1996	x	x		
Stockholm	18.07	59.35	1757–1994	x	x		
Stuttgart	09.20	48.83	1792–2008	x			
Tallin	24.75	59.43	1806–2001	x			
Trondheim	10.50	63.40	1761–1981	x	x	x	
Uppsala	13.63	59.85	1722–2000	x			
Vilnius	25.10	54.63	1777–2008	x			
Warsaw	20.97	52.17	1779–2008	x			
Vienna	16.36	48.25	1775–2008	x			

Fig. 1 Annual mean temperature anomalies (reference period 1961−1990) of the instrumental station measurements in *red* and the corresponding HADCRUT3v grid cell in *black*

Table 2 Model simulations that serve as a pool for the analog search

Model	Control runs	Forced runs	Forcing	Length
ECHO-G	1	2	s, g, v	1,000 years
COSMOS	1	2	s, g, v, n, a, l	1,000 years
HADCM3	0	1	s, g, v, a, l	500 years

s solar irradiance, *g* greenhouse gases, *v* volcanic aerosols, *n* other natural aerosols, *a* anthropogenic aerosols, *l* land-use/cover changes

PSR, this spread is beneficial since it increases the pool of possible analog states for extreme warm or cold periods.

3 Proxy surrogate reconstruction method

In PSR, model climate states are selected that are most similar with the proxy or instrumental data available for specific places at specific moments of time. The proxy or instrumental data thus constitute the reconstruction target, for which the model climate fields serve as analogs. In this study, PSR is applied to reconstruct monthly mean instrumental temperature during the 1800−2000 AD period. All of the monthly mean temperature fields simulated in the model runs for the same meteorological season serve as a pool for possible analogs. The division into meteorological seasons assures that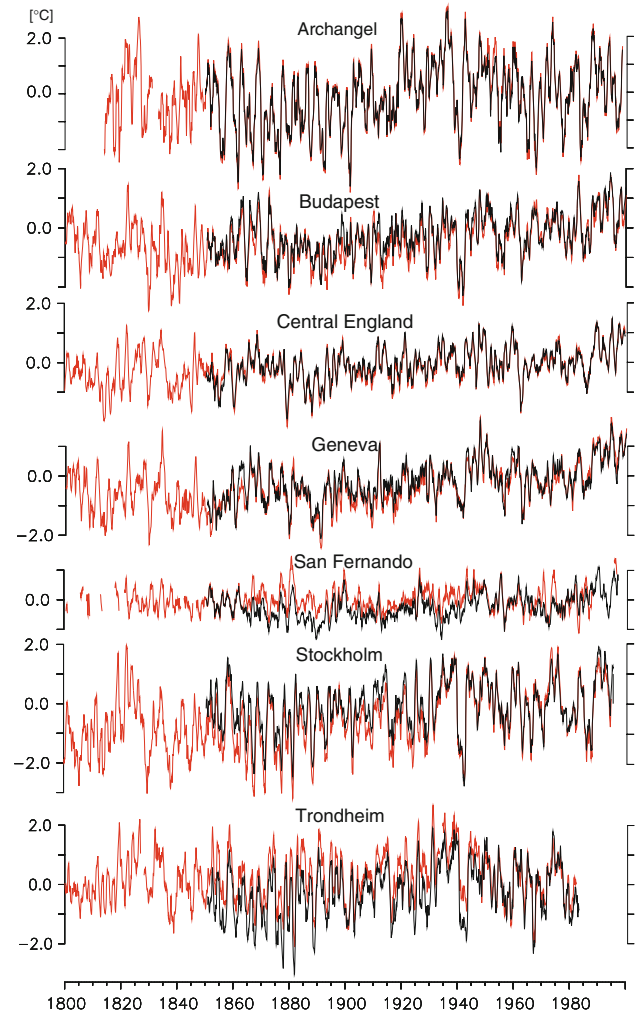 the identified analogs are not influenced by different climate dynamics operating in different seasons. For the 600 months of each season between 1800 and 2000 AD there are always 19,500 possible analogs from the 6,500 total simulation years (Table 1). As the search aims at finding a the best physically consistent spatial pattern that has the smallest distance to the instrumental/proxy data, each analog can be selected from the pool multiple times.

Given that all model simulations and verification data-sets are interpolated to a common resolution, a grid of the same size is generated from the instrumental data, where each grid box contains the time series inside the area. If there is no station inside the grid box a missing value flag is assigned.

The study is based on anomalies calculated with respect to the mean annual cycle during the period 1961−1990 in order to avoid having this deterministic component inflate correlation statistics.

The actual analog search is based on the Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} w_i(x_i - y_i)^2}, \tag{1}$$

where x is the instrumental temperature anomaly and y is the temperature anomaly in the model simulation, both at the same location i. The Euclidian distance is calculated as a sum over all raster cells that include time series.

We applied weighting (w) depending on the local monthly temperature variance. Without weighting, the sites with higher variability would dominate the Euclidean distance in Eq. 1 because they have larger anomalies. This is the case, for more continental locations (e.g., Archangel, Fig. 1) that tend to show more temporal variability than places closer to the ocean (e.g., San Fernando, Fig. 1). To avoid this the differences between instrumental and model temperature are weighted inversely proportional to the standard deviation of the instrumental time series at every location.

Other distance measures such as the squared-chord distance (see overview in Gavin et al. 2003) are not discussed here because they minimally affect the selection of the best analogs but rather the distance metric among the analogs. As the more distant analog climate states from the pool are never considered, only a minor influence is expected due to the choice of the distance measure.

After the calculation of the Euclidian distance, all members of the pool of possible analogs are sorted according to their distance to the target. In the next step a distance-weighted composite is generated from the closest analogs:

$$c(x,y) = \frac{1}{\sum wd_i} \sum_{i=1}^{n} wd_i \cdot a_i(x,y) \tag{2}$$

where c is the composite field, wd is the weight determined by the Euclidian distance (d in Eq. 1), n the number of analogs used and a the analog fields. i.e.: if the composite would be made using just the two closest analogs and the distance of the closest analog to the target would be half of the distance of the second nearest analog, then the best analog would contribute with 66.6% and the second best

with 33.3% to the composite. The number of analogs that are averaged is discussed in Sect. 4.2.

To test the methods and results, we employ the correlation coefficient, the Brier skill score, and the ratio of the variances between the composite analog and the ERA-40 dataset calculated for each grid box and averaged over Europe because the skill changes similarly in the entire spatial domain. The correlation coefficient is a measure of the strength and direction of the linear relationship between predictions and observations. The Brier score is a measure of the explained variance, conventionally described as a relative probability score compared with the probability score of a reference forecast (von Storch and Zwiers 2002; Xoplaki et al. 2003). It is defined as:

$$\beta = 1 - [S_{FP}^2]/[S_P^2] \tag{3}$$

where S_{FP}^2 represents the variance of the error of the forecasts F to the reference predictands P and S_P^2 stands for the variance of the predictand variable P. Thus, for predictions with errors which variance ranges on the order of the variance of the predictand, β is close to 0 or negative and for predictions with a small amount of error, β tends to 1 (von Storch and Zwiers 2002).

4 PSR exploration and verification

In the standard setup for the following sensitivity tests, we use only seven European temperature time series in view of the sparse proxy network available for most of the past millennium (Fig. 1). The seven instrumental series are selected to be well distributed over continental Europe (red squares in Fig. 2) from the records that generally extend
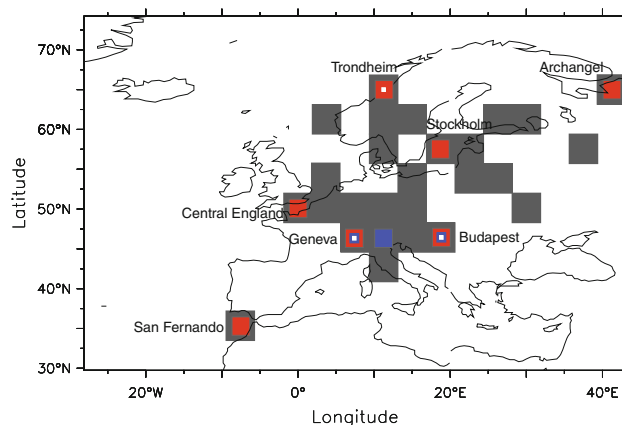

Fig. 2 Grid cell containing instrumental temperature time series of the past 200 years; 27 stations considered in this study in *grey*, standard setup of most experiments of 7 stations in *red*, 3 stations in central Europe in *blue* and 3 stations in central and northern Europe in *white*

back to 1800. A comparison of the single time series with the gridded HADCRUT3v dataset (available back to 1850) indicates that the station data are in general representative for the entire grid cell (Fig. 1). In individual cases or during particular periods some stations show an offset or have missing data. The offsets are explained by the fact that the HADCRUT3v dataset takes all instrumental data available for a grid box into account and not just a single station. Additionally, different versions of instrumental series, due to for example, homogenization or how daily and monthly means are calculated, may also contribute to offsets.

Series with a small amount of missing data are not excluded here, due to the few long instrumental series available. As missing data affects the Euclidian distance, comparisons of Euclidian distances are limited to the 1825–1980 AD period. Composites are made of the ten closest analogs if not stated differently (Sect. 4.2). All analyses are done on a monthly basis using monthly mean values. However, for presentation, monthly results are grouped into meteorological seasons (DJF, MAM, JJA, SON).

4.1 Impact of pool size

The ability to find close analogs intuitively depends not only upon the target, in our case instrumental station temperature data, but also the characteristics of the pool from which potential analogs are drawn. For instance it is interesting to consider to what degree the pool size affects reconstruction quality, whether it makes any difference to use simulations with constant or variable external forcing, or whether certain GCMs intrinsically yield a better fit with

observed climate fields. To assess the influence of the data in the pool, Fig. 3 presents the Euclidian distance between the best analog and the corresponding target for different simulations.

Largest distances are generally observed with the 40 year long reanalysis dataset, decreasing with the 500 year HADCM3 simulation, and even lower distances are obtained with all 1,000 year ECHO-G and COSMOS runs. The lowest distances are obtained using all simulations simultaneously (6,500 years in total, ERA-40 was not included). Thus, the size of the pool is clearly a crucial factor for PSR, but which size is sufficient? The ERA-40 dataset can help answer this question. ERA-40 can be considered as the simulation closest to reality due to the assimilation of observational data for the period 1958–2002. Therefore, it can be viewed as an interesting benchmark since it enables appraisal of the range of distances that can be obtained using quasi-perfect data. A step decrease to smaller distances is found after 1958 because the best analog found for the target month is in fact in most the cases the identical month in the ERA-40 reanalysis dataset. However the distances do not approach zero because the instrumental measurements at a station are not identical to the grid cell of the ERA-40 simulation. Comparing distances for the various simulations with the ERA-40 distances during the 1958-2002 period, we find that the 500 year HADCM simulation (as well as 500-years selected from a COSMOS simulation; not shown) generally have larger distances. In contrast, the 1,000 year long simulations show distances close to those of ERA-40 after 1958 during the entire 200 year period . These results lead us to the general rule of thumb that for the European domain, 1,000 simulation years represents a sufficient pool size for the application of PSR.

In addition to assessment of sufficient pool size, the distances for multiple simulations and models allow other interesting features of both the PSR method and model output to be inferred. For example, The two forced ECHO-G runs (Erik 1 and 2) show broadly similar distances during the nineteenth century, but during the twentieth century Erik 1 fields yield somewhat better matches. These differences result from stochastic processes as the two runs had identical forcings. Furthermore, the two forced ECHO-G runs tend to show larger distances in the twentieth century than the other 1,000-year simulations. This is most probably a consequence of land-cover and anthropogenic aerosol forcings not being included in the ECHO-G, which were identified to reduce the European temperature trend in the HADCM simulations (Tett et al. 2007). If Erik1 and Erik2 anomalies are taken with respect to 1961–1990, a period that is too warm in the model simulation, the majority and earlier part of the simulation is biased towards too cold climate states. Consequently the Erik1 and Erik2
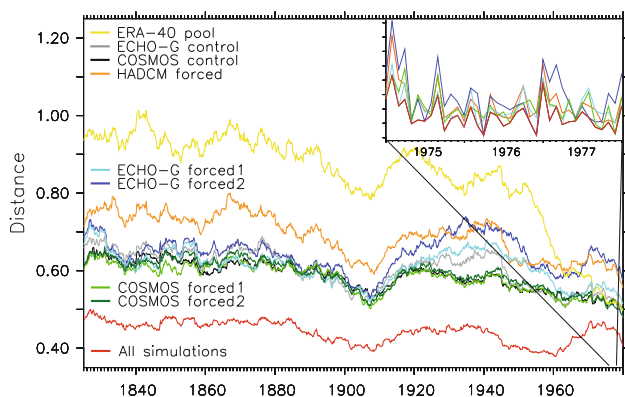

Fig. 3 10-year running mean distance between target and closest analog when different single simulations or all simulations together build the pool for the analog search (*main plot*). Unsmoothed distances (*upper right plot*) highlight that different simulations serve best analogs over time. This results in smallest distances if all simulation are included in the pool simultaneously

simulations effectively have a reduced pool size as many potential analogs are simply too cold to yield a good match with the instrumental targets during the recent warm decades across Europe. Such subtleties related to model characteristics have implications for known or unknown biases in certain model runs related to (1) mean temperature levels, (2) intra-seasonal to inter-annual variance, and (3) spatial autocorrelation to reduce the effective pool size. Interestingly, in this regard, we find that forced simulations do not generally yield smaller distances than unforced simulations, as evidently the internal variability is sufficiently high within this European domain (see also Sect. 4.3). As a final point related to the distances between the model runs and the observational data over Europe, the correlation of the time series in Fig. 3 highlights the peculiarity and the difficulty of finding analogs for each temperature state. If the series were uncorrelated it would mean that the distances depend heavily on the model used. As this is not the case, the relative changes in the distance are due to the fact that in some periods/months it is seemingly more difficult to find analogs than in others (see Sect. 4.5. A slight trend towards increasing distances back in time, may suggest increasing noise in the instrumental data related to changes in intrinsic variance (Della-Marta et al. 2007) or other inhomogeneitites (Frank et al. 2007; Auer et al. 2007).

4.2 Number of analogs in the composite

The identified analog with the smallest distance does not necessarily have to be the best analog for the true climate state as we just subsample the field at a few locations and there might be different patterns matching the same target. Furthermore the station series might include errors of unknown size. To minimize risk associated with selecting a single analog, we generate a composite incorporating the best analogs weighted by their Euclidian distance to the target. This compositing procedure represents a balance between increasing the probability that the composite is representative of the true past climate field versus loosing variance as more analogs are averaged together. To systematically test this, the number of analogs averaged into the composite is stepwise increased as shown in Table 3. The correlation and Brier skill score slowly increase with the size of the composite but stagnate when the number averaged analogs is around 10–20. In contrast, the variance slowly decreases with the number of analogs averaged. Overestimation of the annual mean variance approaches zero when ca. 50 analogs are averaged in the composite. However, this ratio of the variances strongly depends on the season: in summer more than two analogs can already lead to an underestimation of the variance because the climate states in summer have more degrees of freedom

Table 3 Influence of the number of analogs averaged for the composite

Number of analogs	Correlation coefficient	Brier score	Observed ratio of variances		
			ann	JJA	DJF
1	0.82	0.53	1.17	1.00	1.25
2	0.85	0.61	1.13	1.02	1.18
5	0.86	0.65	1.10	0.91	1.25
10	0.87	0.67	1.08	0.91	1.24
20	0.88	0.68	1.06	0.87	1.22
50	0.87	0.68	1.02	0.84	1.2

All values are calculated for the period 1800–2000 and comparing European means of the estimated analog with the ERA-40 reanalysis dataset

than in winter. Thus, more than 50 analogs may not lead to an underestimation of the European mean temperature variance in winter. Consequently, the best compromise depends on the season that should be reconstructed.

As a compromise between seasonal considerations and keeping with a general methodology, we have arrived at ten analogs as a reasonable number to improve the reconstruction skill, compared to a single best analog, without severely underestimating the variance in any season.

4.3 Which simulation serves better analogs?

Each monthly mean temperature pattern of the same meteorological season from each model simulation can be selected as a best analog in each analog search. It is thus possible that individual simulations (or months or time period there from) might be drawn upon more often as a source than others. We performed a search in order to reveal the most selected analogs for the 2,400 monthly target patterns between 1800 and 2000 AD. The results are: no pattern is chosen more than 4 times from the pool, just a single pattern is selected 4 times, 10 patterns are chosen 3 times and 130 patterns are selected twice from the pool as best analogs. Consequently the absolute majority of roughly 88% of the target patterns have a unique best analog with the pool.

Figure 4 represents the number of best analogs selected from each simulation and season. The results are divided into blocks of centuries to identify temporal patterns which might be related to differences in the forcing of the simulations. In spring and autumn all simulations provide a nearly equal number of best analogs. In contrast, there are differences in summer and winter. In summer the ECHO-G control run serves roughly half of the analogs than any other simulations. In winter all ECHO-G simulations and COSMOS 2 provide fewer analogs than the COSMOS control, COSMOS 1 and HADCM. There is no clear

Fig. 4 Histograms presenting the number of best analogs served by specific centuries of all the model simulations which are included in the pool for the analog search (10 bars for 10 centuries in case of millennial length simulations and 5 bars for 5 centuries in case of HADCM3). The results are presented by seasons
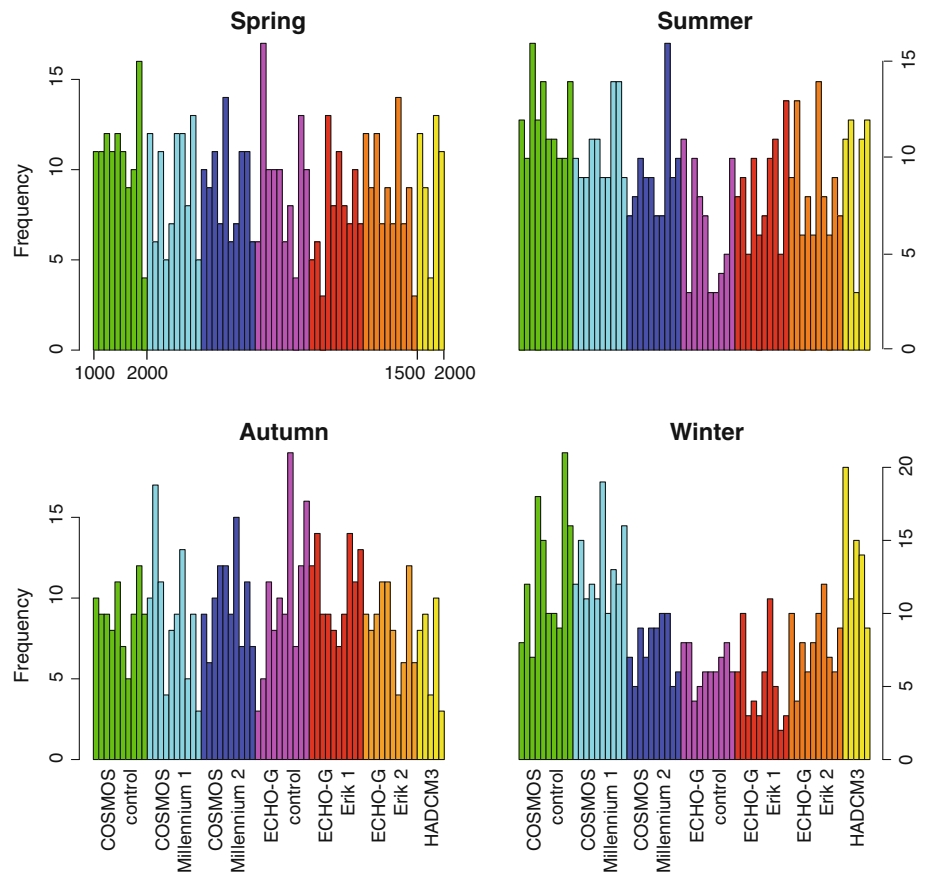
explanation for these differences except for natural internal variability because they cannot be related to the forcings as COSMOS No. 2 and No. 1 have an identical forcing and contribute differently in winter, while the forced and unforced ECHO-G runs contribute a similar amount of analogs (see also Sect. 4.1).

The timing of the control runs is arbitrary because no time dependent forcings were applied. It is therefore not surprising that the sampling for the control runs is distributed relatively evenly through the simulation. The forced simulations, however, simulate past climate variations on decadal to centennial timescales. Hence, it is an interesting result that the best analogs for the twentieth century are not preferentially obtained by forced simulations and in particular not from the recent time period, considering the on-going warming trend. In the unforced control runs of the ECHO-G and COSMOS models the European mean temperature variability exceeds the real temperature variability observed over the past 150 years in the instrumental HADCRUT3v dataset (not shown). This highlights the high internal variability of the models and likely the climate system itself which is large enough on the European scale to provide equally sufficient analog pools from forced and unforced simulations for the purposes of PSR.

4.4 Influence of the number of time series used

At present there is a dense network of meteorological stations measuring temperature in Europe and globally. Further back in time this network shrinks rapidly and in the pre-instrumental period there is even less information from sparsely distributed proxies. How does the analog method react to different amounts and distributions of available temperature information? It is impossible to test all theoretical possibilities but some representative scenarios can be illustrated. This section will focus on four cases: the maximum of 27 European time series for the past 200 years (see grey squares in Fig. 2), 7 stations used previously to simulate a sparser but spatially well distributed proxy network (red squares in Fig. 2), and 3 stations in Scandinavia and central Europe more closely representing available high-resolution proxy data for the last Millennium (white squares in Fig. 2) and 3 stations concentrated in central Europe as an example of sparse and unevenly distributed data (blue squares in Fig. 2).

In order to evaluate the PSR skill for the different scenarios of available station data, correlation coefficients and Brier skill scores between the time series of the analog composite and ERA-40 (1958–2002) at each grid box are calculated as a measure of their covariance. The results

indicate that both the number of time series and the location of the stations influence the correlations and Brier score (Fig. 5). Indeed, skill decreases with a reduction of stations used, although there is only a minor decrease from 27 to 7 stations which is most probably due to the high spatial autocorrelation of temperature (e.g., Büntgen et al. 2010). For other proxies with less spatial covariance such as precipitation the situation might differ. Moving from 7 to 3 stations the decrease in both measures of skill becomes clearly obvious. However, besides the number of stations their locations are of fundamental importance. Using only 3

stations from central Europe results in relative high (>0.6) correlation coefficients in the grid boxes surrounding the stations but very low coefficients in the rest of the reconstruction domain. Instead, incorporating three more evenly distributed stations, correlation coefficients above 0.5 are obtained for nearly all Europe and raster cells with coefficients above 0.7 increase twofold. This indicates that interpolation is possible using the PSR method but extrapolation might only work to a small degree.

Including a larger amount of stations can improve reconstructions slightly, but does not influence results and

Fig. 5 Correlation (*left*) and Brier skill score (*right*) between the estimated analogs for each month and ERA-40 using 27 stations, 7 stations, 3 stations spread over Europe (Budapest, Geneva, Trondheim) and 3 stations in central Europe (Budapest, Geneva, Padova)
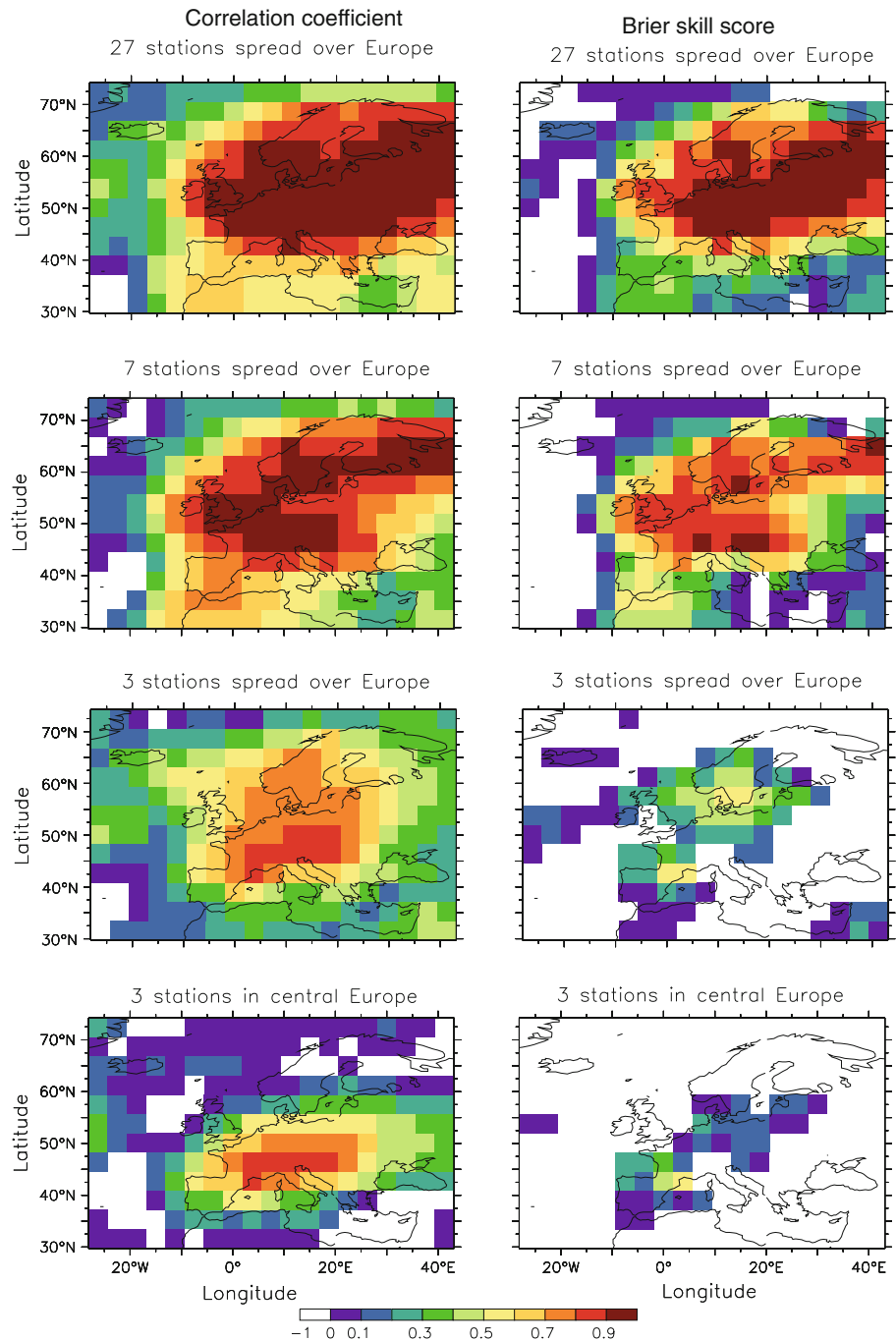
conclusions derived in the preceding sections. Euclidian distances as demonstrated in Fig. 3 and the sources of the best analogs presented in Fig. 4 retain their characteristics independent of the fact that 7 or 27 stations are used. As a general conclusion from these tests, we find that seven well distributed temperature stations are sufficient to represent temperature conditions across continental Europe.

4.5 Climate fields that are difficult to reconstruct

Some targets appear to be more difficult to reconstruct as Euclidian distances are large (Fig. 3). In order to analyze spatial patterns associated with a large distance, we present the temperature field for February 1956 as the worst case having the largest distance between target and best analog (Fig. 6). The entire HADCRUT3v data are plotted for the target climate field to allow for a better visual comparison although the analog search is based on the 7 stations of the standard setup.

It can be observed in Fig. 6 that even the worst-case example has an analog that can reproduce the broad temperature pattern but some features might have a slightly different spatial extension or center. A second observation derived from the comparison with larger distances is that they tend to be related to anomalous temperature patterns. These patterns probably induce large distances for two reasons. Firstly, there are likely fewer analogs in the pool that fit these anomalous states of climate. Secondly, the absolute value of the anomaly is conducive to large distances. If target and analog disagree by a certain percentage, the disagreement in absolute values is larger in case of an anomalous climate pattern. Thus, anomalously warm or cold months generally coincide with comparatively large Euclidian distances. Finally, we observed that it is more difficult to find suitable analogs for temperature fields that exhibit sharp changes within short spatial distances. For example, the southern edge of the cold anomaly (45°N) in Fig. 6 is much more abrupt in the CRU TS 2.1 dataset than the analog composite field.

4.6 Discussion of the temporal consistency

One of the potential weaknesses of PSR is the lack of temporal consistency of the reconstruction due to reshuffling time slices selected from the model. This leads to the question if the reconstructed fields could have existed in this chronological order. On the other hand we focus on the reconstruction of atmospheric variables over a continent that is not as strongly affected by the long-term memory of the climate system caused by the ocean (Trenberth 1995). Thus, lower autocorrelation is expected over the continent. However, the surrounding water masses will still influence European climate and may preclude random selection if
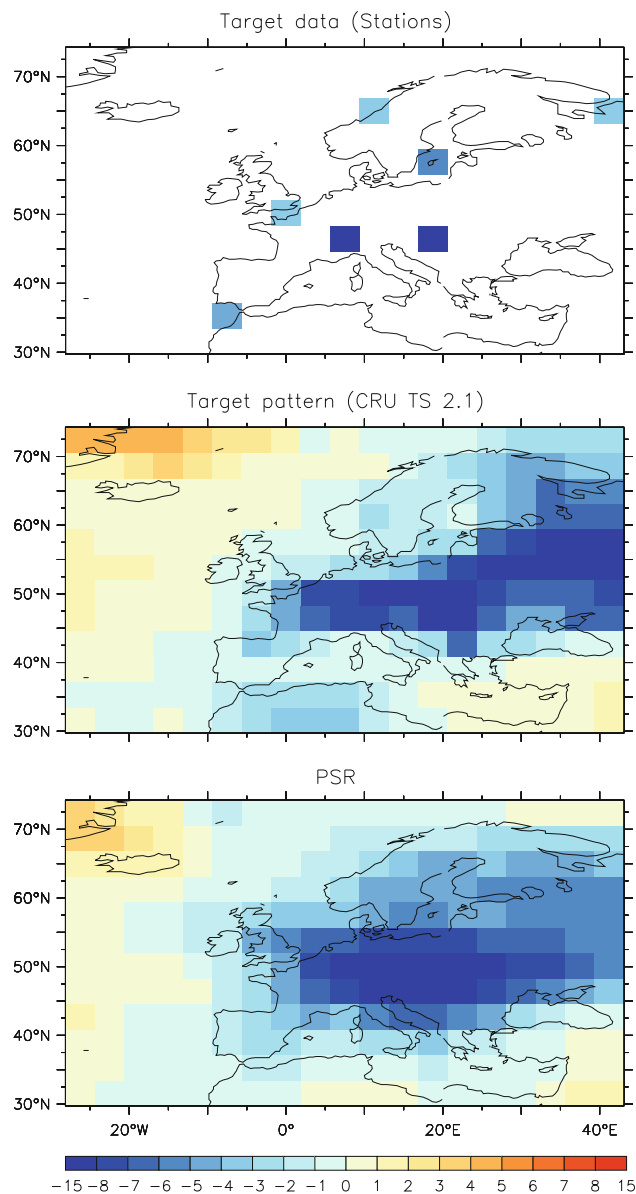

Fig. 6 Worst case PSR reconstruction example (temperature anomalies [°C]) of the month with largest distance between target and best analog (February 1956, *lower panel*), based on the data of the 7 stations (*upper panel*). For better visual comparison the gridded CRU TS 2.1 data are also presented as a target climate pattern (*middle*)

temporal consistency is to be maintained. But at the same time the PSR method cannot select randomly any analog from the pool, because the instrumental stations used to find the analogs force the selected analogs to be temporally as consistent as the instrumental series, at least at the locations where measurements exist. Moreover, in grid cells surrounding the stations no temporal inconsistencies are expected due to the high spatial correlation of temperatures. This latter point is supported by the fact that the correlation and Brier score maps (Fig. 5) indicate skill of the method over the entire European domain.
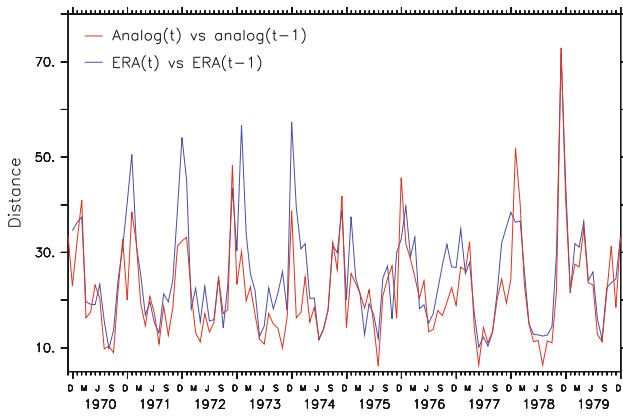
Fig. 7 Euclidian distance between temperature fields of subsequent months in the ERA-40 dataset (*blue*) and in the reconstructed analogs (*red*)

To test this potential limitation, the Euclidian distance between consecutive temperature fields is calculated as a measure of similarity. We considered the distance between subsequent months in the ERA-40 dataset (blue curve in Fig. 7) to represent the true baseline. These distances vary with the annual cycle, in which summer months are rather similar while consecutive winter months differ more substantially (Peixoto and Oort 1992). Distances between subsequent months in the reconstructed analog composites (red curve in Fig. 7) not only show the same annual cycle but also distances of a similar amplitude, and thereby suggesting that potential limitations in model continuity are small.

In this monthly reconstruction of instrumental records, the temporal consistency is particularly an important issue

to address. In applications using proxy data with annual or even lower temporal resolution, temporal continuity is likely even less relevant, especially if the focus is on a region dominated by land and not by ocean, because there is usually very little autocorrelation in the sequence of atmospheric surface temperature patterns over the continents in consecutive years.

4.7 Relevance of single stations

Besides the reconstruction itself, PSR offers a possibility to evaluate single time series. In order to check a record, it is removed from the reconstruction and the method predicts the values for this location using all other series. Comparing the removed record with the values reconstructed/predicted by PSR may allow to find possible inconsistencies in the instrumental (or in extensions, proxy) records, to be identified. An example of this procedure is demonstrated for the De Bilt station (Fig. 8). Here the 10-year low-pass filtered instrumental temperature series of De Bilt is compared to the PSR reconstruction for this location (where the De Bilt record was not included). In autumn and winter the series show high correlation as well as a similar level and amplitude of temperature variations. In spring and summer the agreement is as good as in the other seasons before 1850 and after 1950 but between 1850 and 1950 the station data has a clear offset of +1 to +2°C. This step function to higher temperatures over 100 years indicates possible inconsistencies in the instrumental data. Indeed, this offset only occurs in the long but raw series of De Bilt (Peterson et al. 1998). The GHCN (Global Historical Climatology

Fig. 8 10-year running mean instrumental temperatures anomalies (1961–1990) from De Bilt (*red*) and the corresponding grid cell from the PSR reconstruction (*black*) in which De Bilt was not included
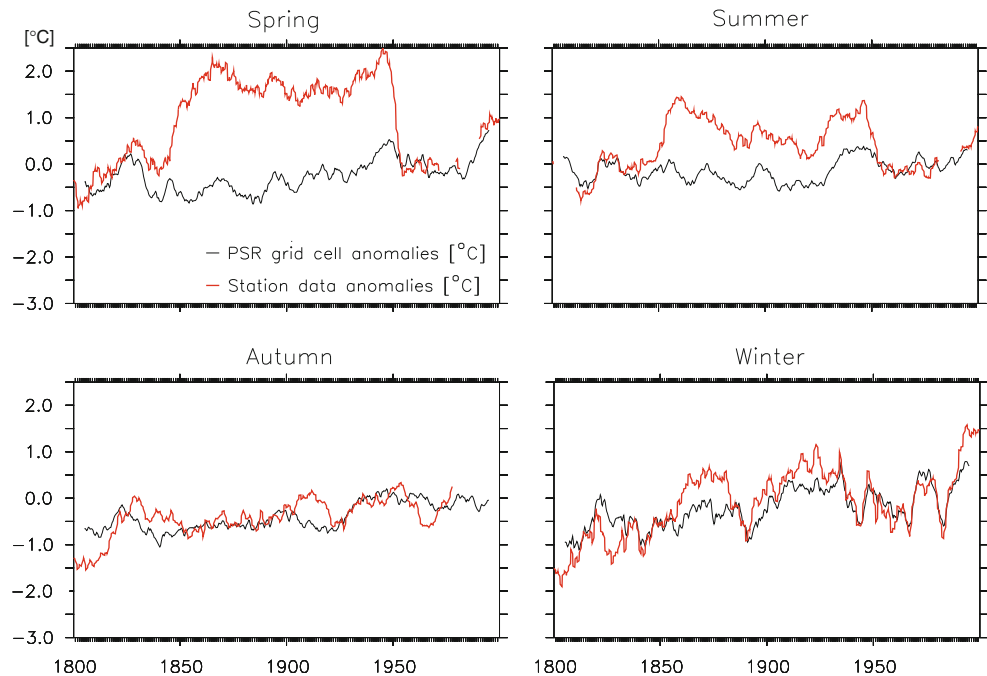
Network) adjusted series has been cut in the year 1850 and the offset between 1850 and 1950 has been removed.

A screening of the other 26 series did not reveal obvious inhomogeneities. Hence, we derived the 200 year European temperature reconstruction based on these instrumental data without De Bilt. However, it cannot be excluded that smaller inhomogeneity issues still affect some European temperature records, especially outside the Greater Alpine Region (Frank et al. 2007; Böhm et al. 2010).

5 200 year temperature reconstruction from instrumental measurements

Having performed methodological tests and developments as described in the previous section, we now turn to applying these methods to present a spatial field reconstruction for Europe using PSR and all 26 instrumental temperature series available (Table 1), excluding De Bilt (see Sect. 4.7). We consider the full pool of 6,500 years of model data and include a weighted mean of the 10 best analogs. Reconstruction skill and properties are evaluated via comparison to: ERA-40, which covers the time period 1957–2002, HADCRUT3v that extends back to the year

1850, and finally the LUT temperature reconstruction covering the full length of this study. After 1901, LUT is the CRU TS 2.1 temperature dataset (see Sect. 4.2). All datasets are interpolated to a 3.75° resolution and for the comparison of the European mean temperature the ocean is masked as no oceanic temperature time series is included in this reconstruction and because LUT only covers land regions. The decreasing area represented by grid boxes of higher latitudes is accounted for in the calculation of the European mean temperature. Some of the instrumental data of the 26 stations used here have also been included in ERA-40, HADCRUT3v and LUT. However, for Budapest, Geneve, Kremsmünster, Munich, Padova, Stuttgart, and Vienna we include newly homogenized series that attempt to correct early inhomogenieties due to insufficient radiation sheltering (Böhm et al. 2010; Frank et al. 2007). In the following discussion, we concentrate on the 1825–1980 period to avoid misinterpretation caused by the missing data.

First, the European average temperatures of the different datasets/reconstructions are compared to investigate the skill of PSR to predict a spatial mean temperature and to focus on the temporal variability of European temperatures. The reconstructed European mean temperature (red lines in Fig. 9) highly correlate with ERA-40, HADCRUT3v and
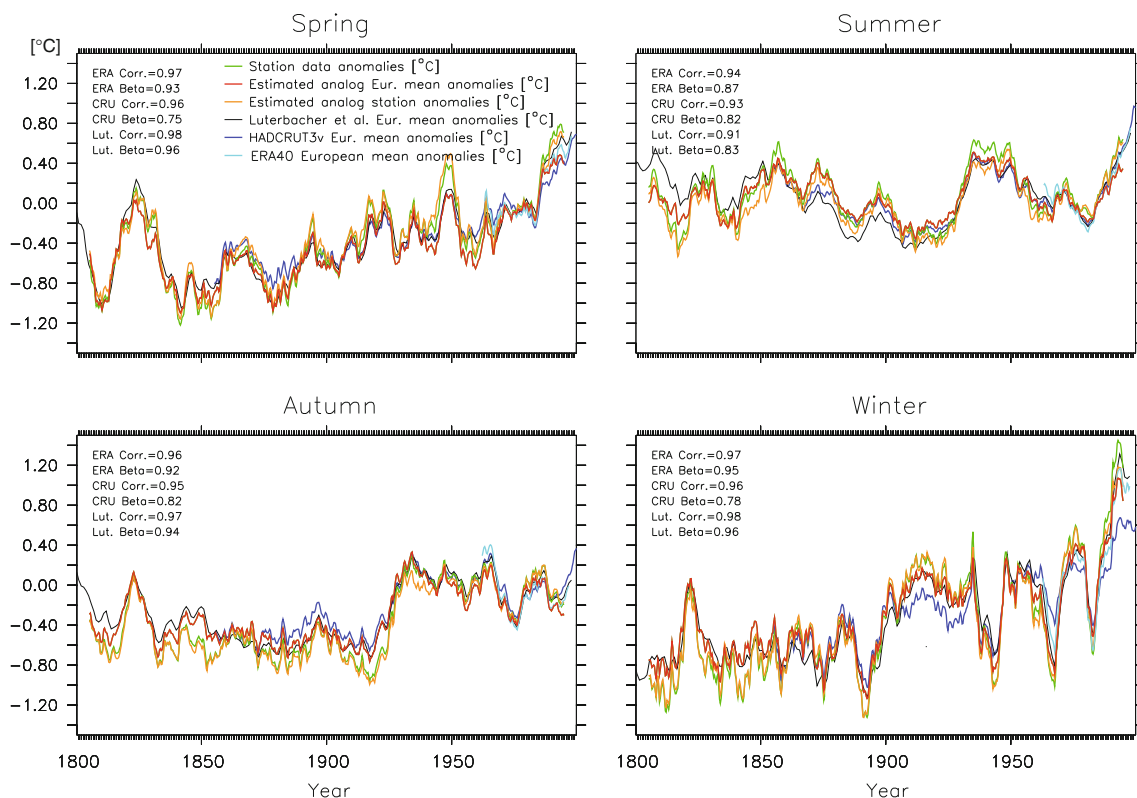

Fig. 9 European monthly mean temperature (10 year running mean) of the estimated analog field (*red*), LUT (*black*), HADCRUT3v (*darkblue*), ERA-40 (*lightblue*) as well the mean temperature of the 26 stations measurements (*green*) and the average of the grid boxes in the estimated analog at these stations (*orange*)

LUT and thus also with CRU TS 2.1 in the twentieth century. This is supported by the high correlations of 0.91–0.98 and Brier scores of 0.75–0.96 depending on the season and dataset.

Between LUT/CRU TS 2.1 and HADCRUT3v slight differences of up to 0.2°C in European mean temperature exist, for instance between 1880 and 1920 in summer or 1910 and 1940 in winter. This might be caused by different data treatment and input data (Brohan et al. 2006; Mitchell and Jones 2005) and spatial resolution and has to be treated as uncertainty in the instrumental data, at least back to the year 1900 where CRU TS 2.1 starts. The PSR reconstruction (red series in Fig. 9) deviates only up to 0.2°C from HADCRUT3v and LUT. Thus, the error of reconstructed European mean temperature does not exceed the uncertainty in the instrumental data. The small offsets between warmer LUT and colder PSR reconstructed temperatures in early nineteenth century summers might be caused by the use of newly homogenized instrumental series in Central Europe (Böhm et al 2010) which were not available at the time when HADCRUT3v and LUT were produced.

The reconstructed mean temperature at the grid cells where the stations are located (orange curve in Fig. 9) is very similar to the mean of the instrumental series (green curve in Fig. 9) as this distance was minimized in the analog search procedure. However, an interesting feature of the PSR reconstruction is that it even predicts a correct mean of the entire European grid if the average of the 26 stations does not agree with the European average (e.g., 1900–1930 in autumn). This appears to reflect the allocation of spatial information in PSR that is not considered in a simple mean of regional station data.

Given that the PSR method appears to satisfactorily capture the temporal changes in European mean temperatures, it is also relevant to evaluate the skill of the PSR method in the spatial dimension via the analysis of reconstructed fields and specific moments in time. In Fig. 10, we present reconstructed fields for an exceptionally cold winter month (February 1929) and a month from the so-called "year without summer" (June 1816). Reconstruction uncertainty is estimated by the standard deviations between analogs that are averaged in the composite. Where all analogs of the composite agree, the
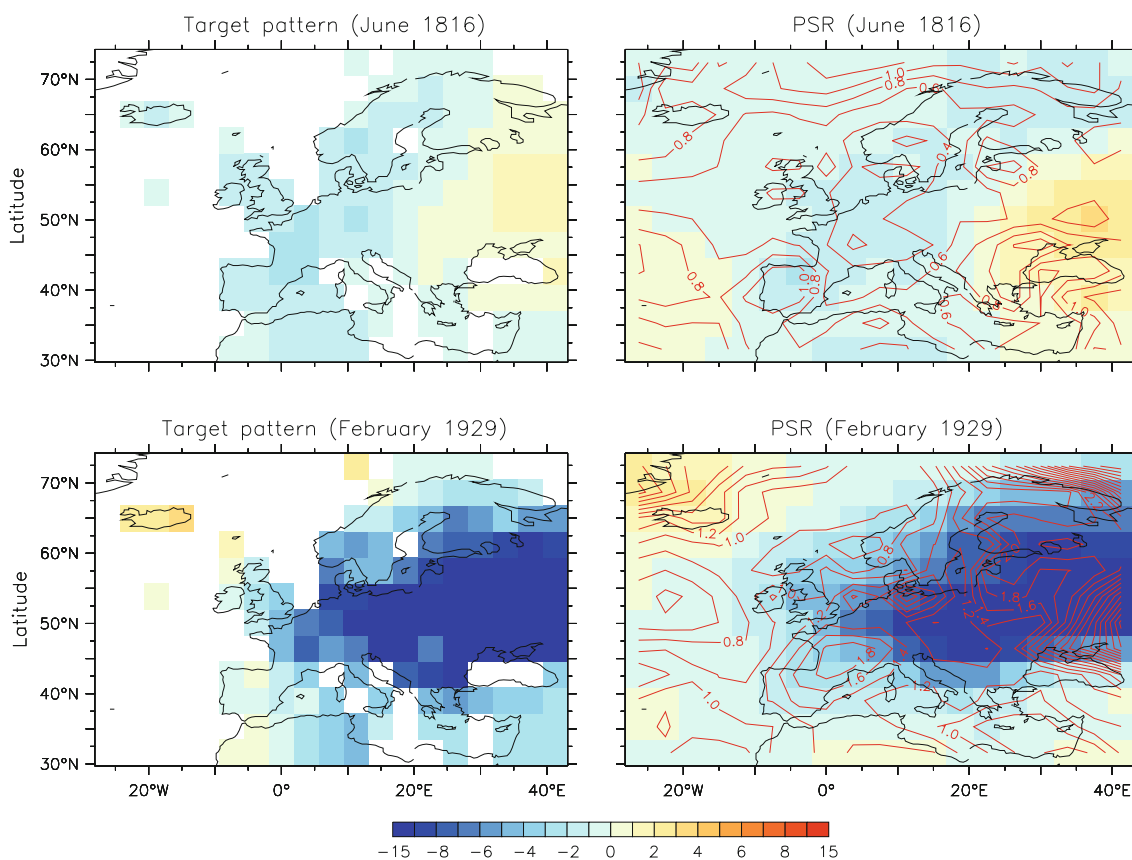

Fig. 10 European monthly mean temperature anomaly field (with respect to 1961–1990) for a month in "the year without summer" 1816 (*upper*) and one of the coldest winter month (*lower*) between 1800 and 2000. The target field from the LUT/CRU TS 2.1 is on the *left* and the PSR composite based on 26 instrumental station records on the *right*. *Red* contours represent the standard deviation between the 10 analog fields the are averaged in the composite and indicate uncertainties in the reconstruction

standard deviation is low and high skill can be expected. In regions with high standard deviations in contrast, single analogs disagree probably because instrumental stations are too far away. At these locations the reconstruction is more uncertain.

The generally good agreement between the target (Fig. 10) and reconstructed climate field confirms the skill of the method to reconstruct spatial patterns of European temperature. Due to the high spatial autocorrelation of temperature, even regions distant from any of the stations can be reconstructed, albeit with decreasing skill. At these distant locations, anomalies generally are underestimated because of the the increased noise averaged into the composite. During the cold winter of 1929, the analog composite correctly suggests a cold anomaly around the Black Sea but not as severe as in the target. However, the strength of the anomaly matches better for instance in the alpine region next to instrumental stations. In places very distant from instrumental stations even the sign of the estimated analog can be misleading. This is seen in the region of southern Turkey and Syria in the example of the summer 1816. This reduced skill is also highlighted by the high standard deviations between the composited analog fields, and emphasizes the limited possibilities to use PSR for extrapolation.

Most existing climate field reconstructions, including some available for European temperatures (Luterbacher et al. 2004) are based upon principal component regression. This method reduces the information of the instrumental data to the dominant modes of the climate system. Reconstructions based on a number of principal spatial patterns (eigenvectors of the empirical orthogonal function analysis) may thus have more limited centers of action in comparison to the real climate system. In Fig. 11 we test for the spatial distribution of climatic extremes in the instrumental data, our analog reconstruction, and a reconstruction based upon principal components regression. Specifically we show counts of the number of times that a grid box represents the location of a monthly minima in summer (JJA) temperature for twentieth century instrumental data, and nineteenth century reconstructions based upon the analog method applied in this study and results from Luterbacher et al. (2004). In the CRU TS 2.1 dataset minima are distributed quasi-evenly over the spatial domain (Table 4), with 83 locations having served as minima during one century. In the PSR reconstruction, the distribution is very similar to the CRU TS 2.1 case (84 vs. 83 cells with minima). In contrast, in LUT only around half as much grid cells (39 vs. 83) are location of minima and raster cells are more clustered with a few cells containing the minimum anomaly in a large number of month (Table 4). Examples of other seasons as well as maxima instead of minima yield similar results. Thus, it appears
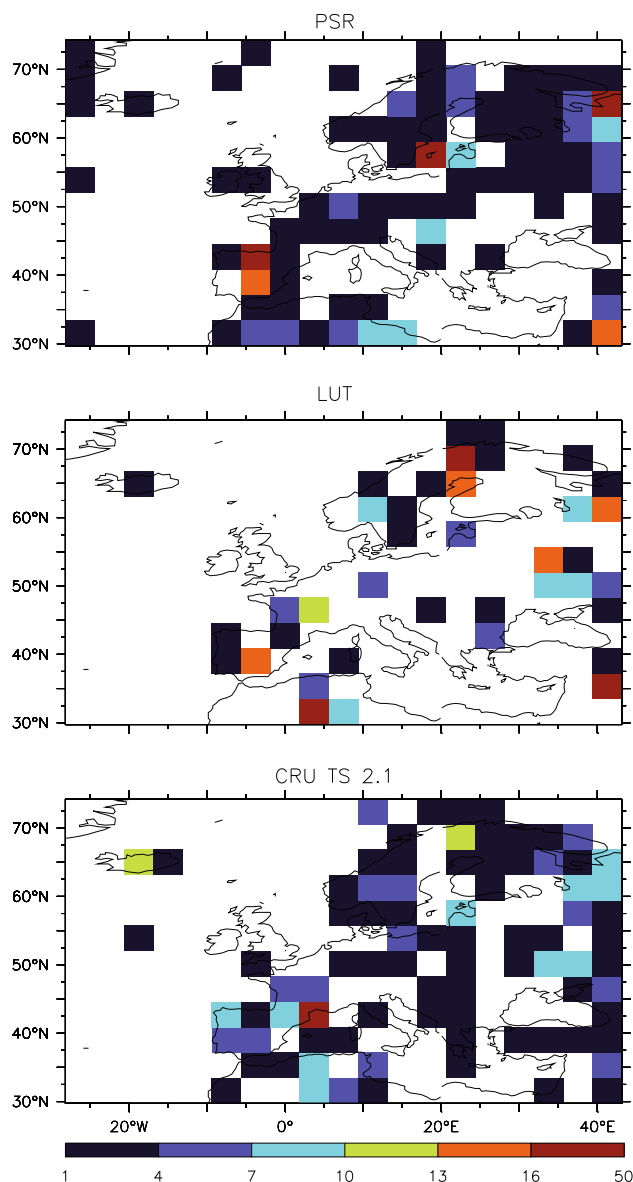

Fig. 11 Number of monthly temperature anomaly minima at each grid box for summers (JJA) in a period of 100 years, 1800–1900 AD in PSR and LUT reconstructions and 1900–2000 AD in CRU TS 2.1

Table 4 Number of grid cells being locations of temperature minima

PSR	LUT	CRU TS 2.1
Grid cells with 1 or more minima		
84	39	83
Grid cells with more than 10 minima		
5	9	3

that PSR can accurately reproduce the characteristics of the natural spatial variability of climate. This characteristic, as well as the freedom from the assumption that the principal patterns identified during the calibration period in the

twentieth century are stationary back in time appear to be a strength of the PSR method. Possible reconstruction biases due to changing climate modes as discussed for the North Atlantic Oscillation (Cook et al. 2002) therefore seem to not apply to PSR.

6 Summary and conclusions

We presented the first systematic examination of the PSR method. In addition this is the first attempt to use PSR as a tool for a completely model-based climate field reconstruction of monthly resolution, constrained by a number of local temperature time series. Our approach has similar objectives and concepts as data assimilation techniques in climate modeling, but is considerably less computationally expensive. A major conceptual strength of data assimilation approaches is their temporal continuity, but we have shown that even at monthly resolution, the lack of temporal continuity in PSR poses minimal, if any, limitations on reconstruction quality. However, as long as PSR is solely constrained by temperature we can only achieve a temperature reconstruction. Other variables such as the atmospheric circulation related to the temperature analogs found, are not temporally consistent.

In comparison with widely applied principal components regression methods, PSR avoids their stationarity assumptions and also appears to be able to overcome loss of variance that has haunted regression based methodologies. The issue of variance loss/amplification still needs to be resolved in future work, but seems to depend at least upon the reconstruction season and likely also the particular characteristics of the model data. As we have limited analysis to the less noisy instrumental period, theoretical advantages related to regression methodology may become more apparent in applications relying only upon proxy data.

Methodological tests reveal that an analog pool of 1000 simulation years can be sufficient if the model simulation does not include unrealistic trends. However, increasing the pool size by a combination of multiple existing simulations of the last millennium, leads to a systematic decrease in distances. We cannot find any advantage of using forced simulations because the internal variability produced in control runs appears to provide sufficient analogs although real climate has undergone a clear temperature trend in the last 200 years. The above results and conclusions likely will need to be revisited when PSR is applied for significantly larger spatial domains.

A reduction of the number of European stations from 27 to 7 used to search for analogs only leads to a small loss of skill. The ability to reproduce European patterns well based upon limited data, suggests extension into the past using proxy information will be possible. Nevertheless, the distribution of the stations in space is very important. Using few, poorly distributed stations, leads to reconstruction skill becoming limited to the area immediately surrounding the stations. This indicates that PSR has limited abilities to extrapolate beyond the spatial range of available data, and thus suggests a significant theoretical advantage for approaches that use covariance information. Besides the reconstruction itself, PSR serves as a consistent framework to identify errors in the time series. This verification ability has been demonstrated on instrumental inhomogeneities, but may be even more interesting to assess proxy data.

Reconstructed fields and their temporal evolution broadly agree with reanalysis data, instrumental datasets, and previous reconstructions. By including newly homogenized series that are available for the Greater Alpine Region that correct biases due to improper early shelters, our reconstruction is based upon improved data compared to existing approaches. However, it is likely that some of these same inhomogeneity issues still affect other long European temperature series (Frank et al. 2007, and references therein) which we utilize. So even for the past 200 years a definitive European temperature history must still be awaited.

For future studies the PSR method offers, among others, the possibility to integrate process-based proxy models (e.g., modeled tree-ring widths vs observed ones). However, the greatest potential appears to be multivariate reconstructions in which proxies for variables such as temperature and precipitation can be incorporated simultaneously. These reconstructions will be consistent at least among all variables used to constrain the climate field and possibly also allow to reconstruct variables which have not been used to constrain the analogs. Another important step and challenge will be the extension of the reconstruction back in time including proxy records with a low signal to noise ratio instead of high-quality instrumental measurements. The skill of PSR when including noisy proxy data can be explored performing proxy based reconstructions in the instrumental period, where the target is known, or more systematically by performing pseudo-proxy experiments where model time series are disturbed by simulated proxy noise. In case the current version of PSR cannot produce skillful reconstructions based on noisy proxy data, methodological improvements such as the use of empirical orthogonal functions have to be explored.

References

Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA (2007) Solar influence on climate during the past millennium: results from transient simulations with the NCAR Climate System Model. Proc Natl Acad Sci 104(10):3713–3718. doi:10.1073/pnas.0605064103

Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones P, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin JM, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovic Z, Nieplova E (2007) Histalp–historical instrumental climatological surface time series of the Greater Alpine Region. Int J Climatol 27(1):17–46. doi:10.1002/joc.1377

Barnett T, Preisendorfer R (1978) Multifield analog prediction of short-term climate fluctuations using a climate state vector. J Atmos Sci 35(10):1771–1787

Böhm R, Jones PD, Hiebl J, Frank D, Brunetti M, Maugeri M (2010) The early instrumental bias: a solution for long central European temperature series 1760–2007. Clim Change. doi:10.1007/s10584-009-9649-4 (in press)

Brohan P, Kennedy J, Harris I, Tett S, Jones P (2006) Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. J Geophys Res Atmos 111(D12): D12106. doi:10.1029/2005JD006548

Büntgen U, Wilson R, Franke J, Frank D, Esper J (2010) Proxy number and location to reconstruct European climate variability. Clim Res (in press)

Bürger G, Cubasch U (2005) Are multiproxy climate reconstructions robust. Geophys Res Lett 58:227–235

Bürger G, Fast I, Cubasch U (2006) Climate reconstruction by regression-32 variations on a theme. Tellus A 32:L23–71. doi:10.1029/2005GL024155

Christiansen B, Schmith T, Thejll P (2009) A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. J Clim 22(4):951–976

Collins M, Tett S, Cooper C (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 17(1):61–81

Cook E, D'Arrigo R, Mann M (2002) A well-verified, multiproxy reconstruction of the winter North Atlantic Oscillation index since AD 1400. J Clim 15(13):1754–1764

Cook E, Woodhouse C, Eakin C (2004) Long-term aridity changes in the western United States. Science 306:1015–1018

Crespin E, Goosse H, Fichefet T, Mann ME (2009) The 15th century Arctic warming in coupled model simulations with data assimilation. Clim Past 5:389–401

Cubasch U, von Storch H, Waszkewitz J, Zorita E (1996) Estimates of climate change in Southern Europe derived from dynamical climate model output. Clim Res 7:129–149

Della-Marta PM, Luterbacher J, von Weissenfluh H, Xoplaki E, Brunet M, Wanner H (2007) Summer heat waves over western Europe 1880 2003, their relationship to large-scale forcings and predictability. Clim Dyn 29:251–275. doi:10.1007/s00382-007-0233-1

ECMWF (2006) European Centre for Medium-Range Weather Forecasts ERA-40 Re-Analysis data. http://badc.nerc.ac.uk/data/ecmwf-e40, British Atmospheric Data Centre. June 2009

Esper J, Cook E, Schweingruber F (2002) Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. Science 295(5563):2250–2253

Frank D, Büntgen U, Böhm R, Maugeri M, Esper J (2007) Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quat Sci Rev 26(25-28):3298–3310

Frank DC, Esper J, Raible CC, Buentgen U, Trouet V, Stocker B, Joos F (2010) Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. Nature 463(7280):527–530. doi:10.1038/nature08769

Franke J, Paul A, Schulz M (2008) Modeling variations of marine reservoir ages during the last 45,000 years. Clim Past 4:125–136

Gavin D, Oswald W, Wahl E, Williams J (2003) A statistical approach to evaluating distance metrics and analog assignments for pollen records. Quat Res 60(3):356–367

González-Rouco J, Zorita E, Cubasch U, von Storch H, Fisher-Bruns I, Valero F (2003) Simulating the climate since 1000 AD with the AOGCM ECHO-G. In: Wilson A (ed) Solar variability as an input to the Earth's environment, International Solar Cycle Studies (ISCS) Symposium, Noordwijk: ESA Publications Division, Tatranská Lomnica, Slovak Republic, pp 329–338, ISBN 92-9092-845-X

González-Rouco JF, Beltrami H, Zorita E, von Storch H (2006) Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. Geophys Res Lett 33(1):L01–703. doi:10.1029/2005GL024693

González-Rouco JF, Beltrami H, Zorita E, Stevens MB (2009) Borehole climatology: a discussion based on contributions from climate modeling. Clim Past 5:97–127

Goosse H, Renssen H, Timmermann A, Bradley RS, Mann ME (2006) Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. Clim Dyn 27(2-3):165–184. doi:10.1007/s00382-006-0128-6

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

Graham NE, Hughes MK, Ammann CM, Cobb KM, Hoerling MP, Kennett DJ, Kennett JP, Rein B, Stott L, Wigand PE, Xu T (2007) Tropical pacific–mid-latitude teleconnections in medieval times. Clim Change 83(1-2):241–285. doi:10.1007/s10584-007-9239-2

Hegerl GC, Crowley TJ, Hyde WT, Frame DJ (2006) Climate sensitivity constrained by temperature reconstructions over the past seven centuries. Nature 440(7087):1029–1032. doi:10.1038/nature04679

Jansen E, J Overpeck KRB, Duplessy JC, Joos F, Masson-Delmotte V, Olago D, Otto-Bliesner B, Peltier WR, Rahmstorf S, Ramesh R, Raynaud D, Rind D, Solomina O, Villalba R, Zhang D (2007) Palaeoclimate. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate change, Cambridge University Press, Cambridge, UK and New York, NY, USA

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

Jones P, Moberg A (2003) Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. J Clim 16(2):206–223

Jones P, Briffa K, Osborn T, Lough J, van Ommen T, Vinther B, Luterbacher J, Wahl E, Zwiers F, Mann M (2009) High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. The Holocene 19(1):3–49

Juckes M, Allen M, Briffa K, Esper J, Hegerl G, Moberg A, Osborn T, Weber S, Zorita E (2007) Millennial temperature reconstruction intercomparison and evaluation. Clim Past 3(4):591–609

Jungclaus JH, Lorenz SJ, Timmreck C, Reick CH, Brovkin V, Six K, Segschneider J, Giorgetta MA, Crowley TJ, Pongratz J, Krivova NA, Vieira LE, Solanki SK, Klocke D, Botzet M, Esch M, Gayler V, Haak H, Raddatz TJ, Roeckner E, Schnur R, Widmann H, Claussen M, Stevens B, Marotzke J (2010) Climate and carbon-cycle variability over the Last Millennium. Proc Natl Acad Sci (submitted)

Krivova N, Balmaceda L, Solanki S (2007) Reconstruction of solar total irradiance since 1700 from the surface magnetic flux. Astron Astrophys 467(1):335–346

Kruizinga S, Murphy A (1983) Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. Mon Weather Rev 111(11):2244–2254

Küttel M, Xoplaki E, Gallego D, Luterbacher J, García-Herrera R, Allan R, Barriendos M, Jones P, Wheeler D, Wanner H (2010) The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750. Clim Dyn (in press). doi:10.1007/s00382-009-0577-9

Lee TCK, Zwiers FW, Tsao M (2008) Evaluation of proxy-based millennial reconstruction methods. Clim Dyn 31(2-3):263–281. doi:10.1007/s00382-007-0351-9

Lefohn A, Husar J, Husar R (1999) Estimating historical anthropogenic global sulfur emission patterns for the period 1850–1990. Atmos Environ 33:3435–3444

Lorenz E (1969) Studies of atmospheric predictability. Tech. rep., Massachusetts Institute of Technology, project No. 8604

Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. Clim Dyn 18:545–561

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H (2004) European seasonal and annual temperature variability, trends, and extremes since 1500. Science 303(5663):1499–1503

Luterbacher J, König SJ, Franke J, van der Schrier G, Della-Marta PM, Jacobeit J, Kttel M, González-Rouco FJ, Zorita E, Xoplaki E, Stssel M, Rutishauser T, Wanner H, Pfister C, Brázdil R, Dobrovolny P, Camuffo D, Bertolin C, Moberg A, Leijonhufvud L, Soderberg J, Allan R, Wilson R, Wheeler D, Barriendos M, Glaser R, Riemann D, Nordli O, Limanwka D, van Engelen A, Zerefos CS (2010) Circulation dynamics and its influence on European and Mediterranean january-april climate over the past half millennium: results and insights from instrumental data, documentary proxy evidence and coupled climate models. Clim Change. doi:10.1007/s10584-009-9782-0 (in press)

MacFarling Meure C, Etheridge D, Trudinger C (2006) Law Dome CO_2CH_4 and N_2O ice core records extended to 2000 years BP. Geophys Res Lett 33:L14810. doi:10.1029/2006GL026152

Mann M, Rutherford S (2002) Climate reconstruction using 'Pseudoproxies'. Geophys Res Lett 29(10):1501. doi:10.1029/2001GL014554

Mann M, Bradley R, Hughes M (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392(6678):779–787

Mann M, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. J Clim 18(20):4097–4107

Mann M, Zhang Z, Hughes M, Bradley R, Miller S, Rutherford S, Ni F (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. Proc Natl Acad Sci 105(36):13252–13257

Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. J Geophys Res 112(D12):D12109. doi:10.1029/2006JD008272

Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F (2009) Global signatures and dynamical origins of the little ice age and medieval climate anomaly. Science 326(1256):1256–1260. doi:10.1126/science.1177303

Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25(6):693–712. doi:10.1002/joc.1181

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

Osborn TJ, Briffa KR (2006) The spatial extent of 20th-century warmth in the context of the past 1200 years. Science 311(5762):841–844. doi:10.1126/science.1120514

Parker D, Legg T, Folland C (1992) A new daily central England temperature series, 1772-1991. Int J Climatol 12(4):317–342

Pauling A, Luterbacher J, Casty C, Wanner H (2006) Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. Clim Dyn 26(4):387–405

Peixoto JP, Oort AH (1992) Physics of climate. Springer, New York

Peterson T, Vose R, Schmoyer R (1998) Global Historical Climatology Network (GHCN) quality control of monthly temperature data. Int J Climatol 18:1169–1179

Petoukhov V, Claussen M, Berger A, Crucifix M, Eby M, Eliseev A, Fichefet T, Ganopolski A, Goosse H, Kamenkovich I, Mokhov I, Montoya M, Mysak L, Sokolov A, Stone P, Wang Z, Weaver A (2005) EMIC Intercomparison Project (EMIP-CO_2): comparative analysis of EMIC simulations of climate, and of equilibrium and transient responses to atmospheric CO_2 doubling. Clim Dyn 25(4):363–385. doi:10.1007/s00382-005-0042-3

Pongratz J, Reick C, Raddatz T, Claussen M (2008) A reconstruction of global agricultural areas and land cover for the last millennium. Global Biogeochem Cycles 22(3):GB3018. doi:10.1029/2007GB003153

Raible CC, Casty C, Luterbacher J, Pauling A, Esper J, Frank DC, Büntgen U, Roesch AC, Tschuck P, Wild M, Vidale PL, Schär C, Wanner H (2006) Climate variability-observations, reconstructions, and model simulations for the Atlantic-European and Alpine Region from 1500–2100 AD. Clim Change 79(1–2):9–29. doi:10.1007/s10584-006-9061-2

Riedwyl N, Küttel M, Luterbacher J, Wanner H (2009) Comparison of climate field reconstruction techniques: application to Europe. Clim Dyn 32(2–3):381–395. doi:10.1007/s00382-008-0395-5

Smerdon JE, Kaplan A (2007) Comments on Testing the fidelity of methods used in proxy-based reconstructions of past climate: the role of the standardization interval. J Clim 20(22):5666–5670. doi:10.1175/2007JCLI1794.1

Smerdon JE, Kaplan A, Chang D (2008) On the origin of the standardization sensitivity in RegEM climate field reconstructions. J Clim 21(24):6710–6723. doi:10.1175/2008JCLI2182.1

Stendel M, Mogensen I, Christensen J (2006) Influence of various forcings on global climate in historical times using a coupled atmosphere ocean general circulation model. Clim Dyn 26:1–15. doi:10.1007/s00382-005-0041-4

Tanré D, Geleyn JF, Slingo J (1984) First results of the introduction of an advanced aerosolradiation interaction in ECMWF low resolution global model. In: Gerber HE, Deepak A (eds) Aerosols and their climatic effects. Deepak Publishing, Hampton, VA, pp 133–177

Tett S, Betts R, Crowley T, Gregory J, Johns T, Jones A, Osborn T, Öström E, Roberts D, Woodage M (2007) The impact of natural and anthropogenic forcings on climate and hydrology since 1550. Clim Dyn 28(1):3–34

Tingley M, Huybers P (2010a) A bayesian algorithm for reconstructing spatially arrayed temperatures. Part 1: development and applications to paleoclimate reconstruction problems. J Clim (in press)

Tingley M, Huybers P (2010b) A bayesian algorithm for reconstructing spatially arrayed temperatures. Part 2: comparison with the regularized expectation-maximization algorithm. J Clim (in press)

Trenberth K (1995) Atmospheric circulation climate changes. Clim Change 31(2):427–453

Trouet V, Esper J, Graham NE, Baker A, Scourse JD, Frank DC (2009) Persistent positive north Atlantic oscillation mode dominated the medieval climate anomaly. Science 324(5923): 78–80. doi:10.1126/science.1166349

van den Dool H (1994) Searching for analogues, how long must we wait? Tellus A 46:314–324

van der Schrier G, Barkmeijer J (2005) Bjerknes' hypothesis on the coldness during AD 1790–1820 revisited. Clim Dyn 25(5):537–553. doi:10.1007/s00382-005-0053-0

von Storch H, Zwiers FW (2002) Statistical analysis in climate research. Cambridge University Press, Cambridge, UK and New York, NY, USA

von Storch H, Cubasch U, González-Rouco J, Jones JM, Voss R, Widmann M, Zorita E (2000) Combining paleoclimatic evidence and GCMs by means of Data Assimilation Through Upscaling and Nudging(DATUN). 11th Symposium on Global Change Studies Long Beach, CA

von Storch H, Zorita E, Jones J, Dimitriev Y, Gonzalez-Rouco F, Tett S (2004) Reconstructing past climate from noisy data. Science 306(5696):679–682

Widmann M, Goosse H, van der Schrier G, Schnur R, Barkmeijer J (2009) Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. Clim Past Discuss 5:2115–2156

Xoplaki E, González-Rouco J, Gyalistras D, Luterbacher J, Rickli R, Wanner H (2003) Interannual summer air temperature variability over Greece and its connection to the large-scale atmospheric circulation and Mediterranean SSTs 1950–1999. Clim Dyn 20:537–554

Zhang Z, Mann M, Cook E (2004) Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to AD1700 from tree-ring data. The Holocene 14(4):502–516. doi:10.1191/0959683604hl727rp

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12(8):2474–2489

Zorita E, Hughes J, Lettemaier D, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8(5):1023–1042

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

Zorita E, González-Rouco F, von Storch H (2007) Comments on "Testing the fidelity of methods used in proxy-based reconstructions of past climate". J Clim 20(14):3693–3698