

A regional climate palaeosimulation for Europe in the period 1500–1990. Part 2: Shortcomings and strengths of models and reconstructions

Juan José Gómez-Navarro¹, Oliver Bothe², Sebastian Wagner², Eduardo Zorita², Johannes P. Werner³, Jürg Luterbacher⁴, Christoph C. Raible¹, and Juan Pedro Montávez⁵

¹Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Switzerland

²Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

³Department of Earth Science and Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway

⁴Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus Liebig University of Giessen, Giessen, Germany

⁵Department of Physics, University of Murcia, Murcia, Spain

Correspondence to: J.J. Gómez-Navarro (gomez@climate.unibe.ch)

Abstract. This study compares gridded European seasonal series of surface air temperature (SAT) and precipitation (PRE) reconstructions with a regional climate simulation over the period 1500–1990. The area is analysed separately for nine sub-areas that represent the majority of the climate diversity in the European sector. In their spatial structure, an overall good agreement is found between the reconstructed and simulated climate features across Europe, supporting consistency of both products. Systematic biases between both data sets can be explained by a priori known deficiencies in the simulation. Simulations and reconstructions however largely differ in the temporal evolution of past climate for European sub-regions. In particular, the simulated anomalies during the Maunder and Dalton minima show stronger response to changes in the external forcings than recorded in the reconstructions. Although this disagreement is at some extent expected given the prominent role of internal variability in the evolution of regional temperature and precipitation, certain agreement is a priori expected in variables directly affected by external forcings. In this sense, the inability of the model to reproduce a warm period similar to that recorded for the winters during the first decades of the 18th century in the reconstructions is indicative of fundamental limitations in the simulation that preclude reproducing exceptionally anomalous conditions. Despite these limitations, the simulated climate is a physically consistent data set, which can be used as a benchmark to analyse the consistency and limitations of gridded reconstructions of different variables. Comparison of the leading modes of SAT and PRE variability indicates that reconstructions are too simplistic, especially for precipitation, which in turn is associated to the linear statistical techniques used to generate the

reconstructions. The analysis of the co-variability between Sea Level Pressure (SLP) and SAT and PRE in the simulation resemble the canonical co-variability recorded in the observations for the 20th century. However, the same analysis for reconstructions exhibits anomalous low correlations, which points towards a lack of dynamical consistency between independent reconstructions.

1 Introduction

Confidence in projections of future climate change is supported by a better understanding of current and past climate changes and by the assessment of the skill of climate models in simulating past and present climate variations (Schmidt et al., 2014). In turn, evidence about the climate in pre-industrial times stems from various sources such as instrumental observations, documentary evidence, environmental proxy-archives or climate simulations. Given this variety, gaining reliable insight in past climate variability requires climatological, statistical and dynamical consistency across these different sources, especially between reconstructions and simulations. However, numerous uncertainties affect the assessment of past climate variability.

Disagreements between simulations and reconstructions may be caused by deficiencies in reconstruction methods (e.g. Tingley et al., 2012), by model limitations that reflect the inadequate spatial resolution and missing and simplified (parameterised) physical processes (Gómez-Navarro et al., 2011, 2013) or both. Beyond these methodological shortcomings, both data sources ultimately rely on inferences from environmental archives, since simulations require to some extent input from reconstructions of past forcing data, for instance related to changes in solar and volcanic activity or land use changes. Environmental proxies (Evans et al., 2013) record influences of various environmental factors and, in turn, palaeo-observations do not necessarily perfectly reflect one particular environmental variable (e.g. Franke et al., 2013). Rather, they usually explain only part of the variability of the variable of interest.

In addition to shortcomings in the data sets, internal variability may become dominant compared to externally forced signals in the variable of interest, especially at regional scale (Gómez-Navarro et al., 2012). This implies that a single model simulation represents only one possible realization, among infinite many, of a possible past climate evolution constrained by initial and boundary conditions and the presence of unforced natural internal climate variability. Thus, a perfect agreement with reconstructions cannot be expected at local scales. A further hint to the comparison relates to the fact that part of the simulated and reconstructed variability is associated to non-climatic effect due to the intrinsic characteristics of the reconstruction methods, i.e. model deficiencies and proxy-specific error terms. At larger scales (continental-to-global) it is assumed that the random internal variability is averaged out. However, a recent comprehensive study indicates that even on continental scales the (global) climate models fail in reproducing specific periods in the historical past, especially over the

southern hemisphere and immediate periods following volcanic eruptions (PAGES2k-PMIP3 group,
55 2015).

In addition internal modes of climate variability may respond to external forcing events like large
tropical volcanic eruptions (Yoshimori et al., 2005; Zanchettin et al., 2012) or variations in compo-
nents of changes in solar activity (Shindell et al., 2001; Vieira et al., 2011). However, especially the
influence of low-frequency solar activity changes on climate and climate variability, is still under
60 discussion (Gómez-Navarro and Zorita, 2013; Anet et al., 2013, 2014; Raible et al., 2014). Environ-
mental archives integrate these internal variations, and while climate simulations cannot be expected
to replicate the exact unforced variations, they ideally should be capable of replicating the forced
variability (if they include the relevant processes). This is particularly the case for Surface Air Tem-
perature (SAT), and to a lesser extent for precipitation (PRE) as both variables are thought to be
65 sensitive to the external forcing variability during the last millennium (Gómez-Navarro et al., 2012).

Attempts to reconcile climate simulations and reconstructions are further hampered by fundamen-
tal differences in the characteristics of the information they provide. Simulations and reconstructions
represent data on different spatial and temporal scales. Simulations provide information with high
temporal resolution, spatially averaged to the grid-cell size. Reconstructions are based on archives
70 affected by local (environmental) climate conditions. Additionally, the specific relation between lo-
cal and large scale environmental factors is only partially constrained (Kim et al., 1984). Various ap-
proaches exist for combining the information obtained from reconstructions and simulations. Among
them are proxy-forward models (Phipps et al., 2013; Evans et al., 2013), data-assimilation (Goosse
et al., 2006, 2012; Widmann et al., 2010) and proxy surrogate reconstructions, i.e. analog methods
75 (Franke et al., 2010; Luterbacher et al., 2010a). In addition to these techniques, dynamical and sta-
tistical down- and upscaling methods are currently introduced (Gómez-Navarro et al., 2011; Wagner
et al., 2012; Gómez-Navarro et al., 2013; Eden et al., 2014).

The basis of dynamical downscaling includes the implementation of a Regional Climate Model
(RCM), driven at its boundaries by a Global Circulation Model (GCM). This allows spatially highly
80 resolved climate simulations over limited areas, consistent with the driving model. This downscaling
approach provides the potential to bridge the spatial scale gap between simulated and reconstructed
estimates of past climate variability. Besides refining the spatial resolution of the model dynamics,
the higher resolved orography of regional simulations also allows an improved representation of the
regional scale boundary conditions. This approach has been successfully applied over the Iberian
85 Peninsula (Gómez-Navarro et al., 2011) or the Baltic Sea (Schimanke et al., 2012). However, the
relatively low number of available regional palaeoclimate simulations is a fundamental restriction.
Recently, Gómez-Navarro et al. (2013) have shown how a high resolution regional climate simulation
with the RCM MM5 is able to improve the performance of its driving GCM when compared to 20th
century observations over Europe for the distributions of precipitation over regions with complex
90 terrain.

Despite the limitations of climate models, a potential benefit relates to their dynamically consistent estimates for different variables, because the evolution of the climate within the model is produced by the application of well-known physical conservation laws. This allows us assessing, through a suitable comparison between reconstructed and simulated climates, to what extent the reconstructions provide dynamically consistent estimates of past climate variability. Likewise it permits to evaluate the consistency of climate reconstructions for different variables, their spatio-temporal distributions and their main variability modes.

Here, we extend the previous assessment of Gómez-Navarro et al. (2013) by evaluating the level of agreement between a regional simulation over Europe for the period 1500–1990 and available reconstructions of seasonal SAT and PRE. We focus our analysis on regions where Gómez-Navarro et al. (2013) found that the regional model provides added value beyond the skillful spatial scales of the global climate model. This way we increase our confidence not only in potential agreement between simulations and reconstructions, but also in the conclusions we can draw from potential disagreements. That is, we do not benchmark the simulation against the reconstruction, instead we jointly analyse both uncertain estimates with the aim of increasing our understanding of past seasonal climate changes in Europe.

The manuscript is organised as follows: In the following section we introduce the observations, simulation and reconstructions used for analysis, including a short overview of the methods. In Section 3 we discuss the past climate evolution in terms of seasonal surface air temperature and precipitation variability present in the data for a number of European sub-regions. We analyse the evolution of probability density functions of precipitation and temperature. In Section 4 we turn our attention from the temporal agreement towards the variability modes; we first compare the dominant reconstructed and simulated variability modes (Section 4.1) for temperature and precipitation. Then, we investigate the consistency between these variables and sea level pressure in terms of canonical correlation. A discussion and subsequent concluding remarks close the study.

2 Data and methods

2.1 Climate simulations

Our analysis uses the output of a high-resolution climate simulation carried out with a RCM over Europe for the period 1500–1990. The RCM consists of a climatic version of the meteorological regional model MM5. This simulation is driven at its boundaries by the Global Circulation Model (GCM) ECHO-G. The horizontal model resolution is 45 km and its domain covers Europe almost entirely (see Fig. 1). This nesting setup is referred hereinafter as MM5-ECHO-G. Both models are driven by identical reconstructions of several external forcings to avoid physical inconsistencies: greenhouse gases, Total Solar Irradiance (TSI) and the radiative effect of tropical volcanic events. This simulation is described in detail by Gómez-Navarro et al. (2013), including a discussion of the

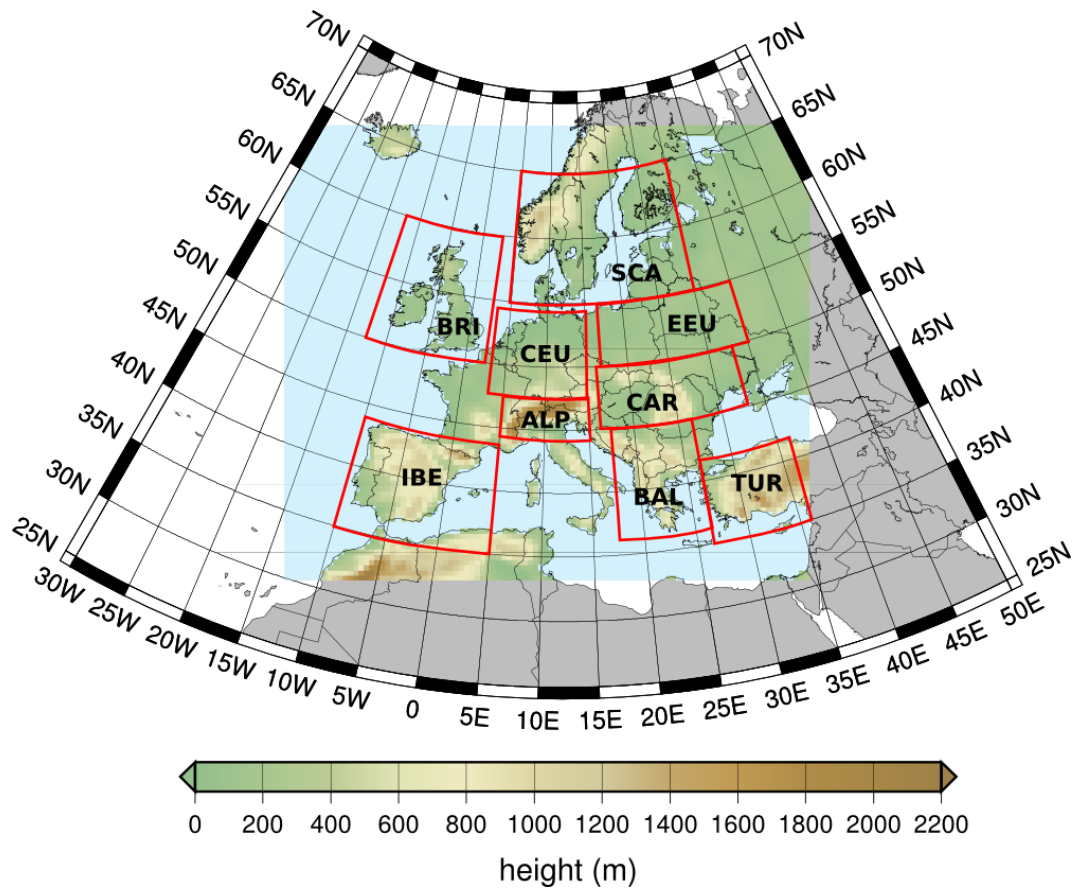


Figure 1. Topography and landmask implemented in the regional simulation, with a horizontal resolution of 45 km. The rectangles show the nine subregions used for more detailed analysis. IBE, Iberian Peninsula; BRI, British Isles; CEU, Central Europe; EEU, Eastern Europe; SCA, Scandinavian Peninsula and Baltex Sea; CAR, Carpathian Region; BAL, Balkan Peninsula; ALP, Alps; TUR, Turkey.

skill of the model MM5-ECHO-G in reproducing the European climate against gridded observational precipitation and temperature data sets. Results of this validation indicate an added value with respect to the driving GCM. However, there are still deviations between the regional simulation and the observations. Prominent problems relate to the divergent 20th century temperature trends. Gómez-Navarro et al. (2013) argued that this could originate from missing anthropogenic aerosol forcing in the simulation, which is an important factor with a potential net cooling effect, especially in the second half of the 20th century (Andreae et al., 2005). Furthermore, the driving simulation with ECHO-G simulates a strong positive trend in the North Atlantic Oscillation (NAO) index under anthropogenic forcing, which is absent in the observations. This leads to a negative trend in winter precipitation in southern Europe and a positive trend in near SAT over Northern Europe.

2.2 Observational data sets

The analysis employs various observational data sets: SAT and precipitation are taken from the monthly data set developed by the Climate Research Unit (CRU) at the University of East Anglia (Harris et al., 2014). This global gridded product includes several climatic variables over land areas with a spatial resolution of $0.5^\circ \times 0.5^\circ$ for the period 1901–2005. In this comparison exercise only temperature and precipitation series up to 1990 are considered, since this is the overlap period between observations and simulation. The data are bi-linearly interpolated onto the MM5 grid to provide a suitable basis for comparison. To keep consistency with reconstructions, only land points are considered for the comparison.

The SLP field consists of monthly means of this variable extracted from the NCEP reanalysis for the period 1948–1990 (Kalnay et al., 1996). This data set has a spatial resolution of $2.5^\circ \times 2.5^\circ$, slightly higher than ECHO-G, and has been used on its original grid without any further spatial interpolation.

2.3 Gridded reconstructions

We use climate reconstructions for three variables, winter and summer SAT, PRE and SLP. In particular we use the gridded data sets by Luterbacher et al. (2004, 2007) for SAT and Pauling et al. (2006) for precipitation. Both data sets consist of seasonal series on a $0.5^\circ \times 0.5^\circ$ regular grid over land areas of Europe. Similar to observations, these data sets were interpolated onto the MM5 grid prior to analysis. These reconstructions are based on a large variety of long instrumental series, indices from historical documentary evidence and natural proxies (see Luterbacher et al., 2004, 2007; Pauling et al., 2006, for details). The basis for the reconstruction is related to the use of linear methods (i.e. principal component regression). Despite the underlying assumptions, e.g. the stationarity of the relationship between the proxy and the climatic variable, the method is able to provide gridded fields for both, temperature and precipitation. Luterbacher et al. (2004, 2007) and Pauling et al. (2006) critically addressed the uncertainties and skills of their reconstructions, especially in the early period of the 16th and 17th century, when less records and only those with lower quality are available. Also Pauling et al. (2006) provide performance maps for their precipitation reconstruction for the reduction of error (RE) of the reconstruction. This allowed a rigorous assessment of the spatial pattern of skill of the reconstruction. An important characteristic of the reconstructed precipitation in contrast to reconstructed temperature relates to the large spatial heterogeneity caused by a considerably shorter spatial de-correlation distance of precipitation. This characteristic becomes critical when attempting to reconstruct hydrological fields from a sparse network of proxy data (Gómez-Navarro et al., 2014).

Additionally, the SLP reconstruction by Küttel et al. (2010) is used, which is based only on station pressure data and ship logbook information, and is thus completely independent from the SAT and

PRE reconstructions. This selection ensures that the dynamic consistency between SLP and SAT and PRE reconstructions can be assessed avoiding circularity (Luterbacher et al., 2010a, b). This data set has a resolution of $5^\circ \times 5^\circ$ and spans the period 1750–1990.

2.4 Framework of the joint analysis of simulated and reconstructed climate

As discussed in the introduction, besides model and reconstruction errors, the presence of internal variability and reconstruction-specific errors a priori prevent perfect agreement between the temporal evolution of the simulated and reconstructed climate variables (Gómez-Navarro et al., 2012). A simple way to partially ameliorate this problem is low-pass filtering the climate series. The underlying argument is that the ratio of forced to internal variability increases at lower frequencies. Since the degree of required filtering is unknown we apply a multi-decadal 31-years running mean using a Hamming window.

In the following we compare the temporal evolution of SAT and PRE as simulated by MM5-ECHO-G with the reconstruction of Luterbacher et al. (2004, 2007) and Pauling et al. (2006), respectively, in nine European sub-domains (Fig. 1). The separation into these nine subregions is a compromise between being able to amalgamate information and taking into account Europe’s climatic complexity. The division is based on the guidelines for coordinated efforts such as the project PRUDENCE (Christensen and Christensen, 2007). We restrict the analysis on the period prior to 1900 to prevent an overlap from the calibration period. As the reconstructions are calibrated using the observational or re-analysis data sets, they should basically agree with the observations used in (Gómez-Navarro et al., 2013) for validation purposes. The authors highlighted the general over-estimation of temperature trends in the simulation during this period, which is strongest for winter in northern Europe. Similarly, precipitation trends of observations and the simulation during the 20th century are often not consistent. We note the contrast between observed wetter conditions and simulated drying in southern Europe in winter. Gómez-Navarro et al. (2013) also found that the regional simulation improved the representation of the observed climatology in the European sub-domains of Scandinavia and the Baltic Sea (SCA), the British Isles (BRI), the Iberian Peninsula (IBE), the Alps (ALP), the Balkan Peninsula (BAL), the Carpathian region (CAR) and Turkey (TUR) relative to the global simulation, whereas the representation did not improve much for Central Europe and Eastern Europe. The reasons mostly pertain to the complex terrain over those regions including a more complex coastline, whereas central and Eastern Europe do in general show less complex topographic characteristics. Therefore we restrict our analysis to those five regions showing an added value over the global simulation.

A simple comparison between the reconstructed and simulated time series might be misleading given the presence of internal variability in the simulation. For this reason, we additionally use Empirical Orthogonal Functions (EOF) analysis to identify the main variability modes of mean seasonal SAT and PRE. These patterns are not critically dependent on the precise temporal evolution within

each data set. Thus, they facilitate the comparison of the climate variability reproduced by the model and the reconstructions. Similarly, Canonical Correlation Analysis (CCA) helps to identify the representation of the spatial co-variability between climate variables in a linear sense, which gives a hint about potential underlying physical mechanisms. Thus, this statistical tool allows us assessing the dynamical consistency among different reconstructions. The two aforementioned techniques are widely used in climate research, therefore we provide only a brief introduction here (the reader is referred to von Storch and Zwiers (1999) for a comprehensive overview).

The basic philosophy of EOF analyses relates to decomposing the spatial (anomaly) fields of the climate variable under consideration into patterns representing most of its variance. An important characteristic of the resulting patterns (denoted as EOFs) and their corresponding time-dependent amplitudes relates to the fact that they are mutually orthogonal in the space and time. From a statistical point of view this characteristic is often of interest, but from more physical point of view the interpretation of the EOF patterns may be complicated because the real world processes and patterns are not necessarily orthogonal. Therefore, the physical interpretation of EOFs has to be performed with caution, especially when consecutive EOFs explain similar amount of variance compared to higher indexed EOFs. To overcome this limitation, several techniques have been proposed to rotate EOFs. They allow to obtain other variability patterns as a result of linear combinations of the original ones. However there is not a unique criterion to perform such rotation, and thus results are affected by a certain degree of subjectivity (von Storch and Zwiers, 1999). Given that in this study we are concerned with the way variance is distributed through the spectrum of EOFs, rather than with obtaining physical meaning from such modes, we restrict the analysis to the standard EOFs.

CCA is a technique related to EOF-analysis. It also decomposes the original variable in a number of components or patterns. However, in this case the aim is to identify pairs of patterns in two variables whose temporal component in the original series exhibits a maximal temporal correlation. Similarly to EOFs, the resulting CCA pairs of time series are ranked according to their mutual correlation, although an important difference with EOF is that in this technique the canonical pairs do not form an orthogonal decomposition of the original space. Instead, the CCA time series corresponding to consecutive pairs are uncorrelated in time. Often the most physical meaningful information is spanned by the leading CCA patterns, although the associated patterns may not explain the largest amount of variance. An advantage of CCA for our purposes is that it helps disentangling the most important (canonical) relationships between climate variables, in the observations, the reconstructions and simulations. Hence, from a physical point of view the leading patterns should show similar characteristics when the mechanisms leading to the relationships between the climate fields are controlled by the same processes. Conversely, deviations from this behaviour are indicative of physical inconsistencies among variables.

3 Temporal agreement of regional series and climatologies

3.1 Regional time series

Figures 2 and 3 depict the evolution of the averaged winter and summer SAT, respectively. It is estimated as the median value within each subregion in the ECHO-G-MM5 model, in the driving GCM and in the Luterbacher et al. (2004, 2007) reconstruction. For the sake of brevity, the figures corresponding to the intermediate seasons are shown in the supplementary material, but the respective main characteristics are also outlined here. As mentioned section 2, the series are low-pass filtered (with a 31-year Hamming low-pass filter) to emphasise the low-frequency variability. Also the evolution of the 25-75 interquartile range is shown to illustrate the heterogeneities within each sub-region.

A first result is the reduction of warm biases in winter through downscaling the GCM output, mainly over areas of strong land-sea contrast near the Mediterranean (Fig. 2 and 3). The width of the interquartile range is similar between the data sets, although the GCM exhibits a larger width of the Probability Density Function (PDF) of winter SAT in the BAL and SCA regions. In summer (Fig. 3) the RCM is not clearly able to reduce biases, and both simulations are generally too cold. It is noteworthy how the RCM increases the width of the PDF compared to the driving GCM, resulting in a better agreement to the reconstructions. Intermediate seasons (see supplementary material) show a more heterogeneous pattern. Absolute biases in autumn are generally smaller: ECHO-G exhibits biases that are positive and negative depending on the season, whereas MM5-ECHO-G is systematically colder. A similar behaviour is found in Spring, where the RCM simulations are slightly but consistently colder than reconstructions. However, the sign of the biases is reversed across areas, and also in different seasons, which precludes drawing a simple picture of the behaviour of biases. The added value of the RCM becomes more clear-cut in the width of the PDF in areas of complex topography such as ALP or IBE, where the GCM produces too small variability (Figs. 2 and 3). These results resemble those described for observations (see Fig. 10 in Gómez-Navarro et al., 2013). This is an indication that the biases between the simulation and the reconstructions are probably associated to model deficiencies (e.g. too zonal simulated atmospheric circulation), rather than to potential errors in the gridded reconstructions. Similarly, variability is larger in winter than in summer in both data sets, as well as in Northeastern areas (note the different scales in Fig. 2 and 3). This agreement is related to the skill of the model setup to reproduce the general climatic features of the European climate (Gómez-Navarro et al., 2013), and the fact that the reconstructions are calibrated with observational records over the 20th century. Hence, this agreement is linked to the consistency of both data sources and their ability to reproduce the observed climate during the 20th century.

Focusing on the temporal evolution, the RCM follows the evolution of SAT of the GCM. Therefore the following discussion is solely based on MM5-ECHO-G. Both the reconstruction and the RCM simulation generally agree better in their low-frequency evolution over northern Europe. Over

southern Europe no clear-cut similarities are found. Regarding the centennial to decadal evolution, the simulation and reconstruction generally agree until 1700. There are anomalous episodes which appear to be synchronised between different regions (Fig. 2 and 3). This can be seen for both, the reconstructions and the simulation and is indicative of prominent anomalies taking place at larger spatial scales. However, these episodes are not synchronised across both data sets, indicating that these decadal variations might be unrelated to variations of external forcings. Since the early 19th century the simulated summer and winter temperatures show a clear warming trend across all regions. The trends of reconstructed temperatures start rising later, are generally lower and/or restricted to one of the two seasons. Thus, regional decadal anomalies of simulated and reconstructed data diverge for most regions over the past approximately 200 years. However, disagreement at decadal scales increases in some regions as early as the beginning of the 18th century. While IBE, BRI, ALP, BAL and TUR reconstructed and simulated series start to diverge in the early or the mid-19th century, CAR and SCA show pronounced anomalies in the 18th century which lead to large simulation-reconstruction deviations. This is also seen in the central and eastern European domains. Overall, there are no statistically significant correlations between the filtered series of reconstructed and simulated SAT, (taking into account the presence of serial autocorrelation in the filtered time series). Also, the temporal agreement does not show any seasonality signal. Considering that SAT is potentially strongly influenced by the external forcings (Gómez-Navarro et al., 2012), the lack of agreement points toward inconsistencies between the smoothed simulated and reconstructed SAT that cannot be explained by internal variability alone.

The time series of seasonal precipitation are shown in Figs. 4 and 5. In contrast to temperature, the RCM improves the seasonal precipitation compared to the driving GCM, which is in agreement to earlier findings (Gómez-Navarro et al., 2013). This is mainly due to the fact that precipitation processes are more notably influenced by orographic features, which are better resolved in the RCM. Similar to SAT, there are noticeable biases that can be explained with model deficiencies. For example, the model tends to overestimate winter precipitation in central and northern Europe in the observational period since 1905 (Gómez-Navarro et al., 2013), which generates a wet bias in SCA (also in CEU and EEU, see supplementary material). It is noteworthy that biases are not as prominent in summer. This is also the case when the model is compared to observations for the 20th century (Gómez-Navarro et al., 2013). Indeed the RCM is able to improve the general underestimation of precipitation of the GCM in summer (Fig. 4 and 5). In Autumn and Spring, biases are generally smaller and do not show any systematic sign, because the systematic biases in the zonal circulation play a minor role in the precipitation during these seasons. Independently from the biases, the agreement between simulation and reconstruction is expected to be lower for this variable due to the larger importance of internal and small-scale variability on precipitation (Gómez-Navarro et al., 2012, 2014).

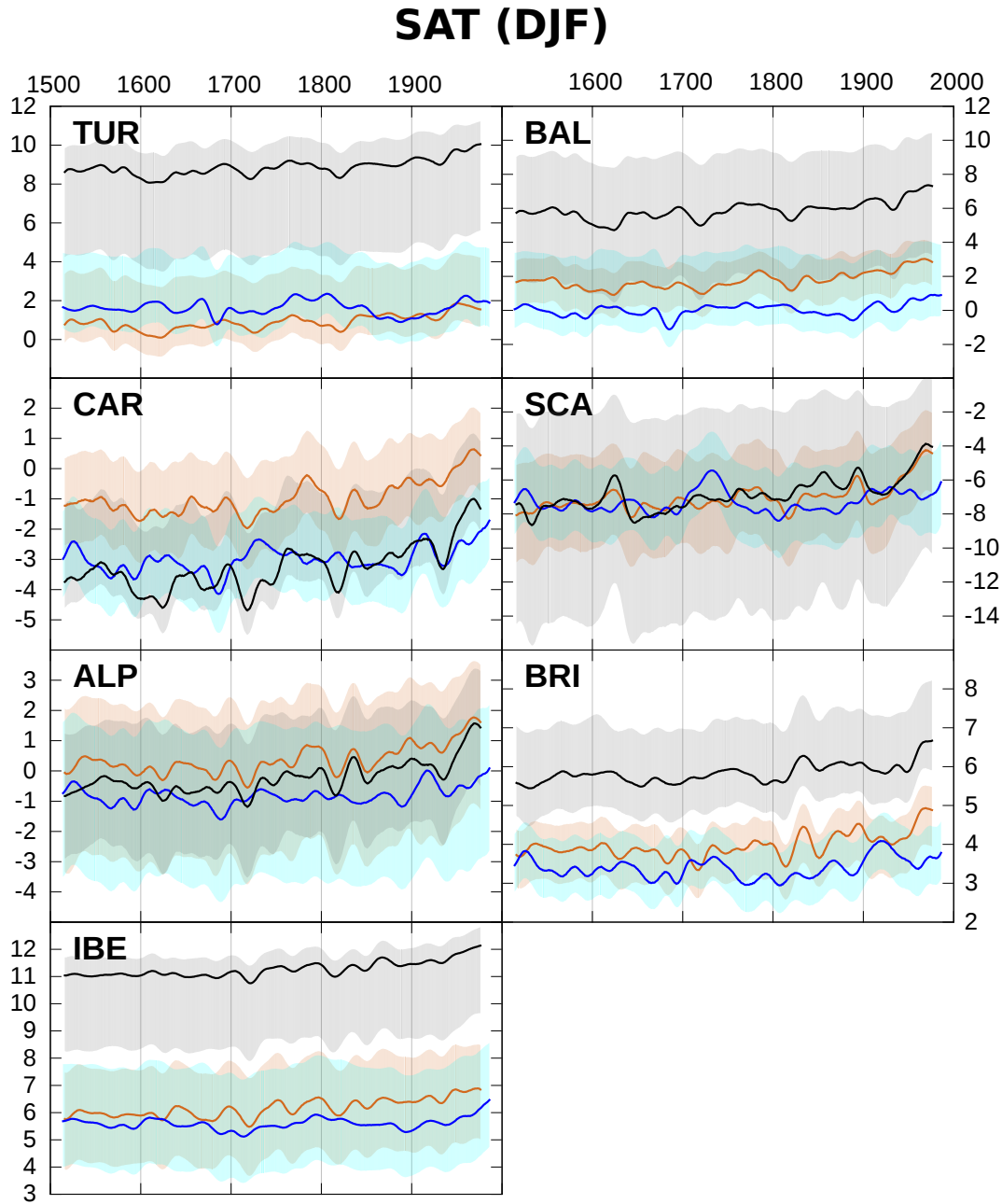


Figure 2. Temporal series of winter SAT in the seven areas indicated in Fig. 1 that exhibit added value in the MM5-ECHO-G simulation compared to the GCM alone according to Gómez-Navarro et al. (2013). The series corresponding to the three different data sets are shown with different colours: driving GCM (i.e. ECHO-G model alone, black), RCM (i.e. MM5-ECHO-G, orange) and gridded reconstruction (i.e. the Luterbacher et al. (2004) reconstruction, blue). Bold lines correspond to the median, whereas the light shading indicates the 25-75 interquartile range to illustrate heterogeneities within each region. After the calculation of the annual values, the series are smoothed through a Hamming window of 31 time steps to emphasise the low-frequency variability. Note the different scale in different panels.

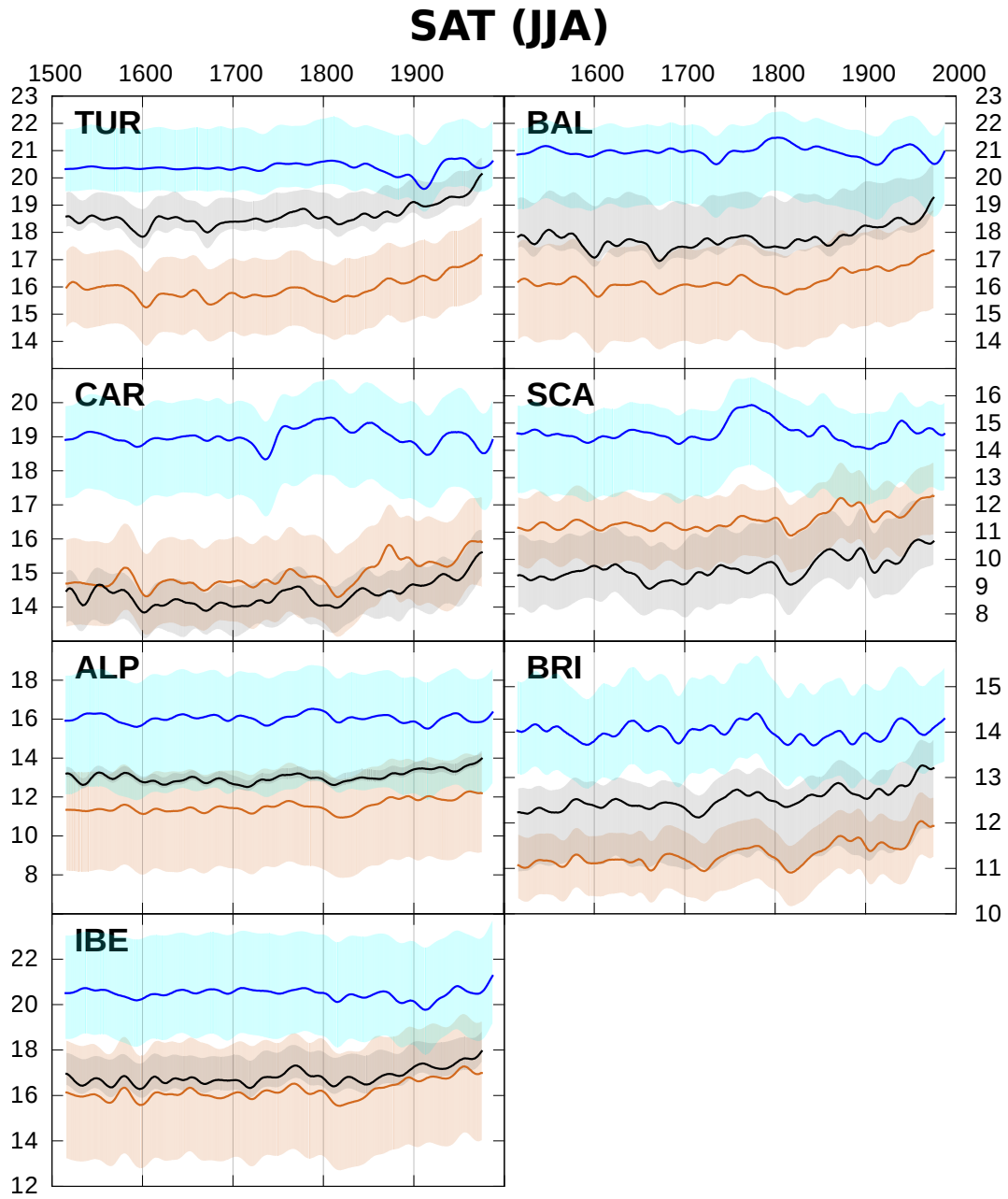


Figure 3. As Fig. 2 but for simulated summer SAT.

A comparison between seasonal reconstructed and simulated precipitation shows less variability in northern than in southern areas. The temporal variability appears to be particularly large in areas of complex orography such as ALP, TUR or IBE. Both data sets show strong low-frequency variations in most regions with pronounced dry and wet episodes over the period 1500–1900. However, these episodes are neither synchronised between both data sets nor for the two seasons (Fig. 4 and 5). Variability also appears to change over time. For instance, simulated winter variability increases in TUR whereas reconstructed summer variability weakens in CAR.

The most prominent features and discrepancies between reconstructions and the simulation are: In the early 16th century, CAR and ALP suggest prominent summer dryness, which is absent in the other series. Reconstructions further show wet winters in BRI in the 16th century. There are hints of coherence between reconstructed and simulated summer ALP precipitation. Reconstructed summer precipitation in the 17th century indicates very wet conditions for CAR, BAL and ALP while BRI summers appear to have been dry. Anomalous dryness is also seen in the early 18th century in summer in CAR, TUR, BAL and ALP reconstructions while summers were wet in BRI and SCA during that period. Winter wetness in the 19th century is prominent in many regions in the simulation (Fig. 4).

A regional peculiarity is a pronounced alternation between drier and wetter conditions with diminishing amplitude and shortening period between 1500 and 1800 in reconstructed CAR summer precipitation. Variations in TUR winter precipitation are very large in the simulation but rather low in the reconstruction. Iberian winter precipitation shows an apparent anti-phase between simulation and reconstruction.

In summary, we do not find clear temporal agreement between the simulation and the reconstructions, especially for PRE. Although forcing leaves an imprint in the simulated SAT, no general congruence between the simulation and reconstructions is found. Pronounced anomalous periods are evident in reconstructed winter temperature in the early 18th century and in reconstructed 17th and 18th summer precipitation which are absent in the simulation. Section 3.3 assesses in more detail the anomalies in some key periods.

3.2 Evolution of climatological probability distributions (PDFs)

The nine regions depicted in Fig. 1, are comparatively large in their spatial extent. Indeed, they often include very different climatic characteristics, where the model produces opposite biases (see Figs. 4 to 8 in Gómez-Navarro et al., 2013). Further, the mean value potentially discards valuable information, such as regional deviations or widening of the distributions of temperature or precipitation within a region in different periods of time. To account for this important part of climatic variability, Figs. 2 to 5 show also the time series of the interquartile range of the spatial distribution of the seasonal means of grid-cell temperature and precipitation within each region. This range is used as a proxy for the actual PDFs, which are not shown to avoid too complex figures. This range provides

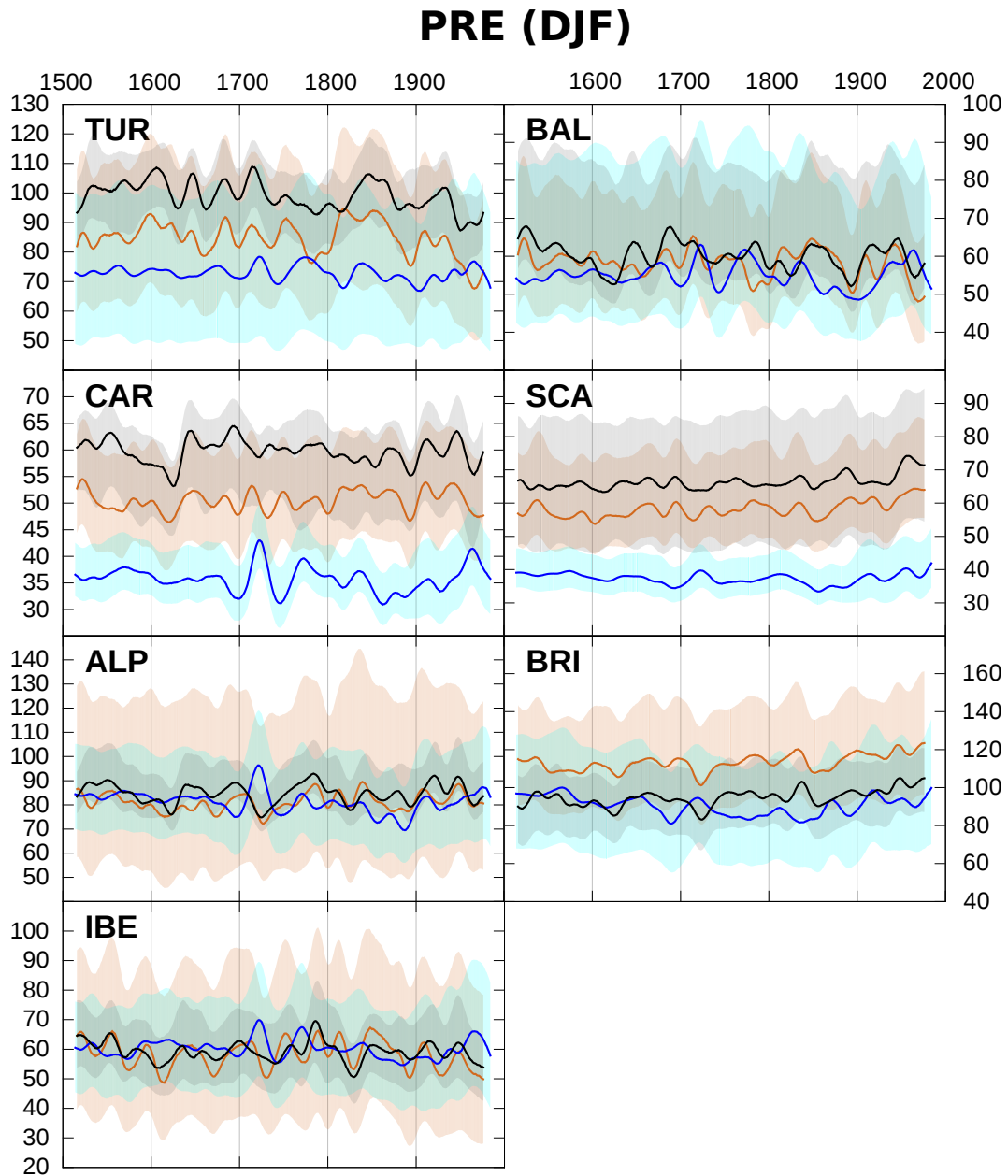


Figure 4. As Fig. 2 but for winter PRE. PRE reconstructions by Pauling et al. (2006)

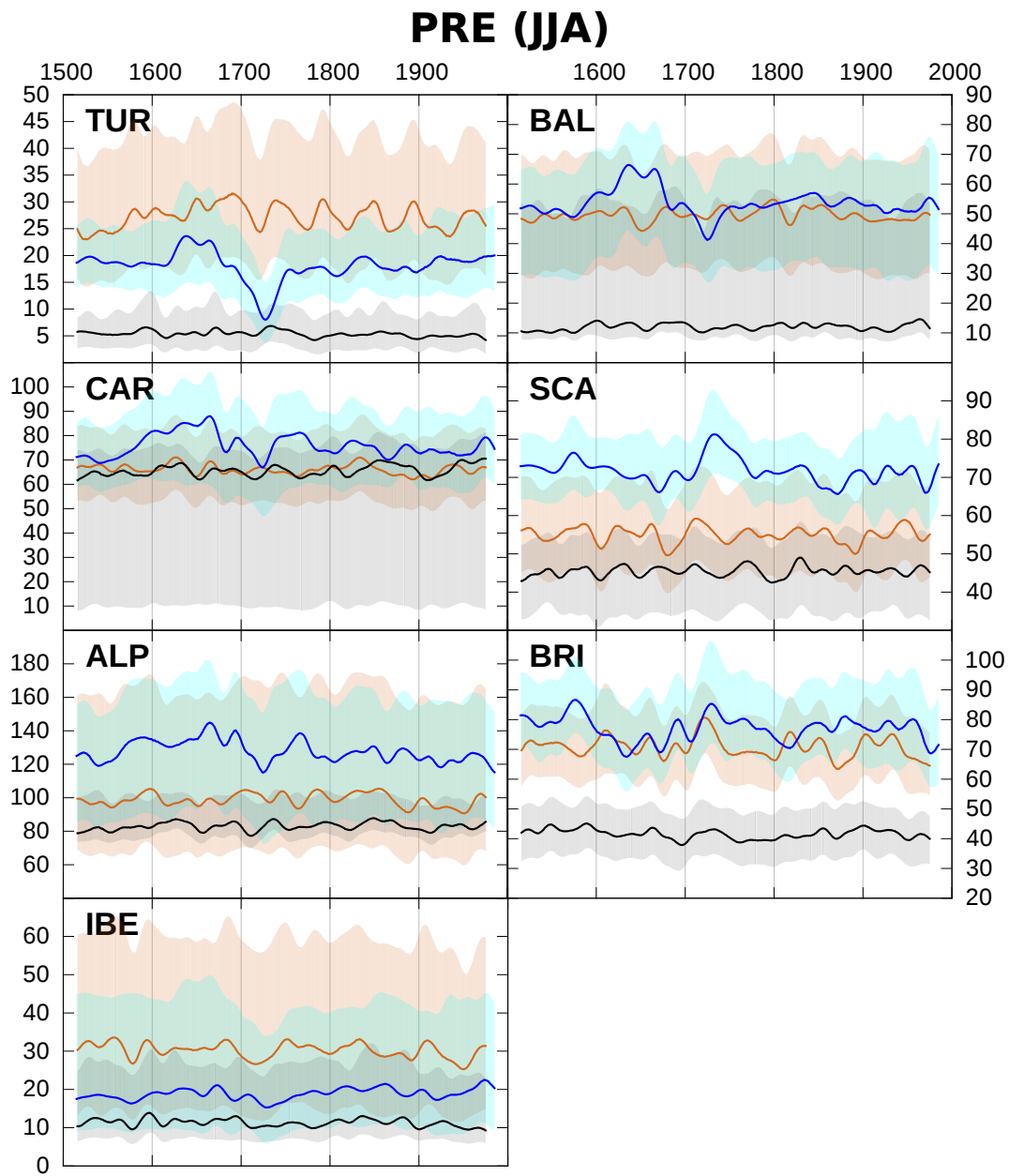


Figure 5. As Fig. 4 but for summer PRE.

information beyond the mean value alone, enabling also the evaluation of the evolution of the spatial PDFs of SAT and PRE within regions, particularly the presence of skewness in the distributions.

Low frequency variability in the median generally translates to variability in the PDF, i.e. the distributions shift in time as a whole, with little changes in their shape. This indicates that the median is a valid indicator for the regional evolution of all percentiles. The relation holds less well for precipitation, especially in summer and to a larger degree in the reconstruction. This is potentially due to the convective and localised character of summer precipitation that leads to non-normal PDFs (Gómez-Navarro et al., 2014).

The median series in Figs. 2 to 5 already suggest that differences between the Late Maunder (1675–1715) and Dalton (1780–1820) minima and the recent 20th century climatology (1961–1990) disagree between the simulation and reconstructions. However, while the percentiles reflect changes in the mean temperature, shifts in the distributions are rather small in the order of 1°C to 2°C colder means and quartiles. Most notable is the cooling for both periods in the winter SCA temperature. Distinct precipitation changes occur only for SCA and only in winter, with low solar forcing periods being drier than the recent climatology (see Fig. 4).

The underlying temperature PDFs generally agree well between the simulation and the reconstruction, in contrast to the evolution of the median time-series. Simulated winter temperature distributions are close or similar for IBE, SCA, BRI, TUR and ALP. Simulated summer temperature distributions are clearly biased towards a colder mean in all regions. Nevertheless, the shape of the distributions is generally similar (not shown).

The simulation and reconstruction disagree more on the PDFs of winter and summer precipitation. The differences between the Late Maunder Minimum, the Dalton Minimum and the late 20th century climatology are spatially less homogeneous across regions. Generally, the mean is underestimated and the extremes are overestimated for southern European winter precipitation, while summers are generally less dry in those regions in the simulation. On the other hand northern Europe shows the opposite for both seasons.

3.3 SAT anomalies during key periods

Given its relevance assessing climate sensitivity and being an important benchmark for climate reconstructions, we analyse SAT anomalies around a prominent cold period in the preindustrial period, the Dalton Minimum (DM). This event is characterised by the simultaneous occurrence of lower TSI and two strong tropical explosive volcanic eruptions. Fig. 6 shows the anomalies of winter and summer SAT in the simulation and the reconstruction. Note that the other seasons show an intermediate behaviour and are omitted here. The simulation (top row) exhibits a clear cold period, in particular in northeastern Europe in winter and central and south Europe in summer. These results, and the particularly cold summers in Iberia, are consistent with results obtained by Gómez-Navarro et al. (2011). The reconstruction (bottom row) shows slightly negative SAT anomalies in northern Europe,

particularly around the Baltic Sea. Compared to the simulation, summer exhibits no cold anomaly at all. A similar spatial distribution can be found for the period Late Maunder Minimum (1675–1715), and the same conclusions can be drawn from the comparison (not shown). Again, this lack of agreement can have multiple explanations. Given the relatively small variability of the reconstructions (see Gómez-Navarro et al., 2011, and the results in the next section), especially in summer, this mismatch might be partly attributable to an underestimation of variance of the SAT reconstruction.

A remarkable feature in reconstructed winter SAT is the strong warming trend during the first decades of the 18th century in several parts of northern Europe. Indeed, this warming is embedded in a very anomalous period characterised by a large climatic variability, and culminated with an exceptionally cold winter in 1740 (Luterbacher et al., 2002; Jones and Briffa, 2006; Zorita et al., 2010). The anomalous warming trend is mostly detected in areas such as SCA or EEU, and less notably in ALP or CAR. Fig. 7 depicts the winter SAT anomalies in the reconstruction and the simulation in the 1700–1750 period with respect to the preceding century. There is an apparent warm anomaly in winter temperatures extending from north to southeast Europe. Such an anomaly is not reproduced in the simulation, neither in this nor any other period prior to the late 20th century. The inability of the model to reproduce such a noticeable anomaly has several implications: On the one hand, internal variability could be responsible for such an anomalous event, rendering an agreement very unlikely. On the other hand, the fact that such an anomalous period is not reproduced in any other period of the simulation, points towards fundamental limitations in the simulation unrealistically restricting the spectrum of possible simulated extreme events (see also the discussions in Wetter et al. (2014).

4 Dynamical consistency of simulation and reconstructions

The available gridded reconstructions of winter and summer temperatures, precipitation and sea level pressure allow not only evaluating the temporal evolution at certain locations, but also to analyse the spatial structures of dominant modes of variability. Moreover, their evolution in different periods and the relation between modes of different variables can be investigated with canonical correlation analysis (CCA). With the latter approach we gain insight in the dynamical consistency among reconstructions and between reconstructions and the simulation (Luterbacher et al., 2010a, b).

4.1 Modes of variability for SAT and PRE

Figure 8 shows the first EOF for winter (left) and summer (right) SAT for the CRU data set (top row), MM5-ECHO-G (middle) and the Luterbacher et al. (2004, 2007) reconstructions (bottom row). The patterns are based on observations for the 1901–1990 period, whereas for the model and the reconstructions they are calculated for the period 1500–1990. The time period used to calculate the EOFs appears to be of minor relevance. Indeed, the patterns are robust, exhibiting only minor changes when the 1901–1990 period is used in the simulation and reconstructions (see discussion below).

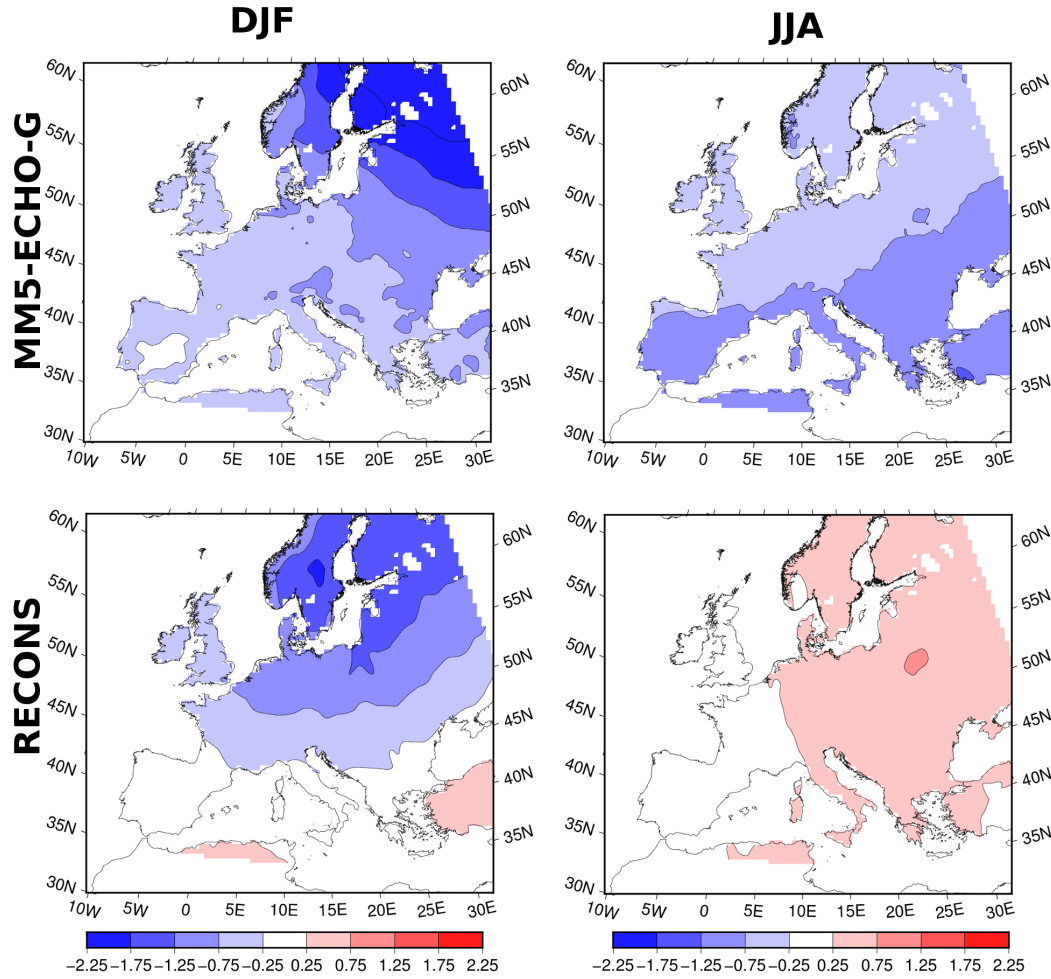


Figure 6. SAT anomalies in winter (left) and summer (right) around the Dalton minimum (1780-1820) with respect to the control period (1900-1990). Top and bottom rows show the results corresponding to the simulation and the reconstruction, respectively.

420 The second and third EOFs, also representing a remarkable amount of variance, are discussed here just briefly and shown in the supplementary material. Note that the EOF patterns, i.e. the eigenvectors, are not normalised, but contain the corresponding units for each variable, so the spatial integral of the square of the pattern is proportional to the variance explained by the respective pattern. In order to facilitate the comparison, the same color scale is used in all maps. Therefore the patterns are multiplied by different scaling factors, indicated in the top right corner of each panel (Fig. 8).

425 Reconstruction and simulation agree well on the shape of the main EOF pattern for winter and summer SAT variability. They represent similar amounts of variability (indicated in each map), and also the total variance is similar. Note, for example, that the scaling factors consistently vary among data sets, and that summer maps had to be multiplied by a larger factor, indicating that summer series

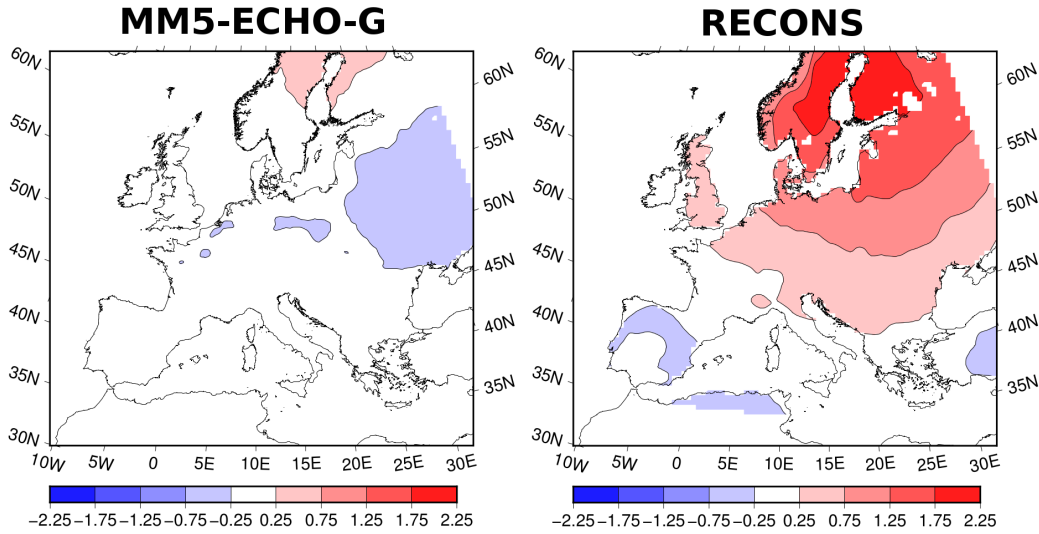


Figure 7. SAT anomalies in winter during the first decades of the 18th century (1700-1750) with respect to the previous century (1600-1700). Left and right maps show the results corresponding to the simulation and reconstruction, respectively.

show less variability, as already pointed out by (Gómez-Navarro et al., 2013) and discussed in the section 3. Although only the leading variability mode is shown, this general conclusion applies also to the higher indexed EOFs (see supplementary material). The simulation, coherently with the observations, exhibits a monopole pattern centred over Eastern Europe, whereas this centre is slightly shifted towards the Baltic Sea in the reconstruction in both seasons. The resemblance between observations and reconstructions increases when the 1901–1990 period alone is considered (not shown), resulting in the slight sensitivity of the pattern to the choice of period. Note that a resemblance between the CRU data and the reconstructions can be expected, especially when the same period is used for the calculation. This is due to the fact that the reconstruction is calibrated against observations and the reconstructions are bound by PCA-regression to show very similar EOF patterns through the whole period (Raible et al., 2006). In the simulation, there is a larger agreement between 20th century and the full-period EOFs (not shown), suggesting that the main patterns of variability are not very sensitive to their respective base period, and more importantly, that the arguably short length of observation records appears to be adequate to calibrate the proxy data.

The simulated and reconstructed SAT tend to attribute more variance to the first EOF in winter (71% and 72% of total variance in the model and reconstructions, respectively) compared to observations (61%). This difference is stronger in summer, when the leading mode in the observations represents 36% of the total variance compared to 57% and 48% in the model and reconstructions, respectively. This indicates that the simulated temperature covariance matrix is too homogeneous, particularly in summer, which is a reminder of the limitations of climate simulations: the zonal cir-

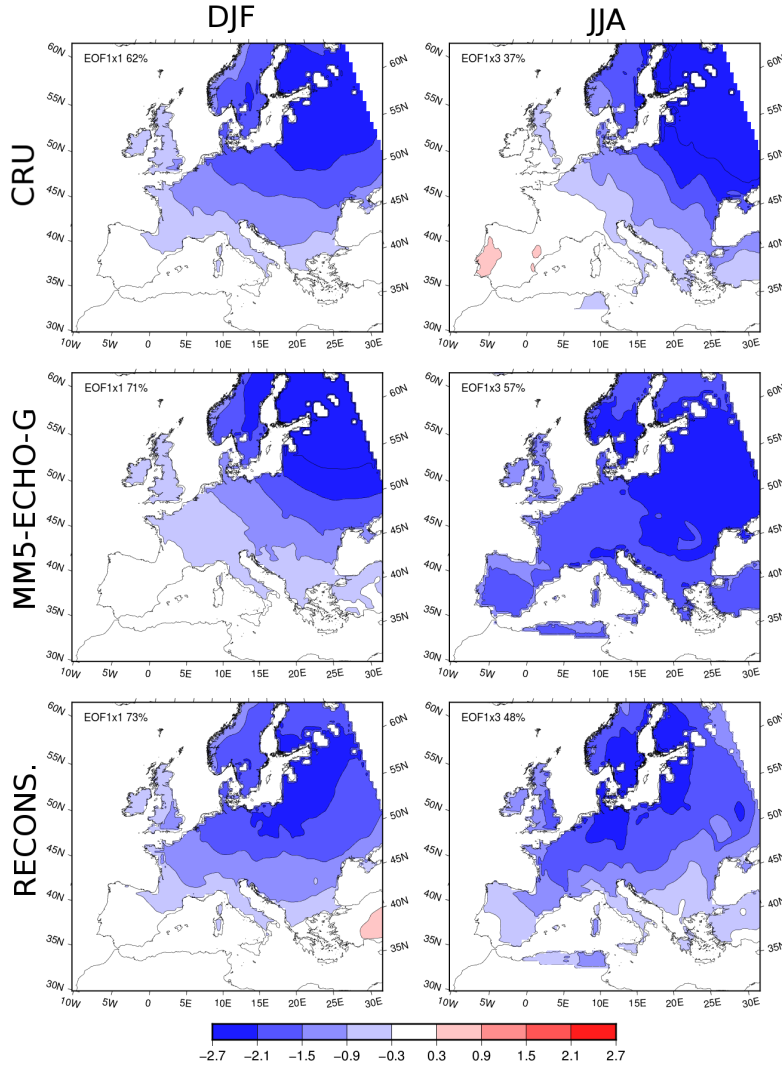


Figure 8. First EOF of winter (left) and summer (right) SAT. Rows depict the results for the CRU data set (top), the MM5-ECHO-G simulation (middle) and the reconstructions (bottom). For the first case the 1901–1990 period is employed, whereas for the other the period 1500–1990 is considered. Note that the patterns carry the units of the variable, and thus they are proportional to the squared root of the variance that each pattern represents. Hence, and to facilitate the comparison, each pattern has been multiplied by a scaling factor, indicated in the top right corner of the figure. The percentage of total variance represented by each pattern is also indicated. The units are $^{\circ}\text{C}$.

450 culation in the driving GCM is too strong. This leads to a circulation regime in the RCM that is reminiscent of that observed in winter. Regarding the reconstruction, the larger proportion of variance represented by the reconstruction's leading EOF highlights again that using a truncated EOF-basis in the PCA-regression results only in partial representation of the true variability. The second and third EOFs are broadly similar in the reconstructions and observations for summer temperature, although their order is inverted. They still show similar gradient-like patterns, with the direction of the greatest gradient slightly tilted in the simulation compared to that in the reconstructions and the observations, respectively.

Figure 9 is similar to Figure 8 but for PRE (higher order modes of variability are shown in the supplementary material). In winter all data sets agree well and show a strong North-South dipole with the node at about 55°N . This pattern highlights the well-known difference between the Mediterranean area and Northern Europe. However, although the spatial structure agrees, the first mode represents more variance in the reconstruction than in the observations. The simulated leading variability mode represents 34% of the winter variance compared to 30% in the CRU data set. However the difference is larger in the reconstruction, where this mode explains up to 46% of the total variance. In summer the leading mode of variability represents just 15% in the observations. This can be explained by the fact that the precipitation regime is less influenced by the large-scale circulation. Despite the too strong zonal circulation in the driving global simulation, this is consistent with the regional simulation, where the leading EOF also represents a low percentage of variance (12%). However this is in strong contrast to the reconstruction, where the first EOF alone is able to account for 40% of total variance. For the summer season, the spatial pattern of the observed and the simulated precipitation agree relatively well, while the North-South gradient observed in these data sets is changed mostly to a strong pole over the Alpine region, with a slight gradient to the North-East. The dominating first mode in the reconstructions shows that the reconstructed precipitation regime is too homogeneous. This conclusion matches similar findings obtained through Pseudoproxy Experiments (Gómez-Navarro et al., 2014), where it has been shown how the linear regression used in Pauling et al. (2006) tends to underestimate the high spatial variability of precipitation.

In the following we briefly describe how the main variability modes compare between the GCM and the RCM. This comparison allows characterising when the downscaling adds value, and represents an aspect of the analysis not shown by (Gómez-Navarro et al., 2013). The main variability modes of SAT exhibit very similar patterns in both models and seasons, although the GCM reproduces less spatial variability, as expected from its coarser spatial resolution (not shown). The percentage of variability represented by the main mode is 72% in winter, indistinguishable from the RCM (Fig. 8). In summer this percentage drops to 38%, in better agreement with observations, although the spatial structure shows generally less resemblance with observations, with a lower Southwest-Northeast gradient. For PRE, the GCM compares worse than the RCM with CRU. In winter, the GCM is able to reproduce the characteristic main variability mode dominated by a North-South gra-

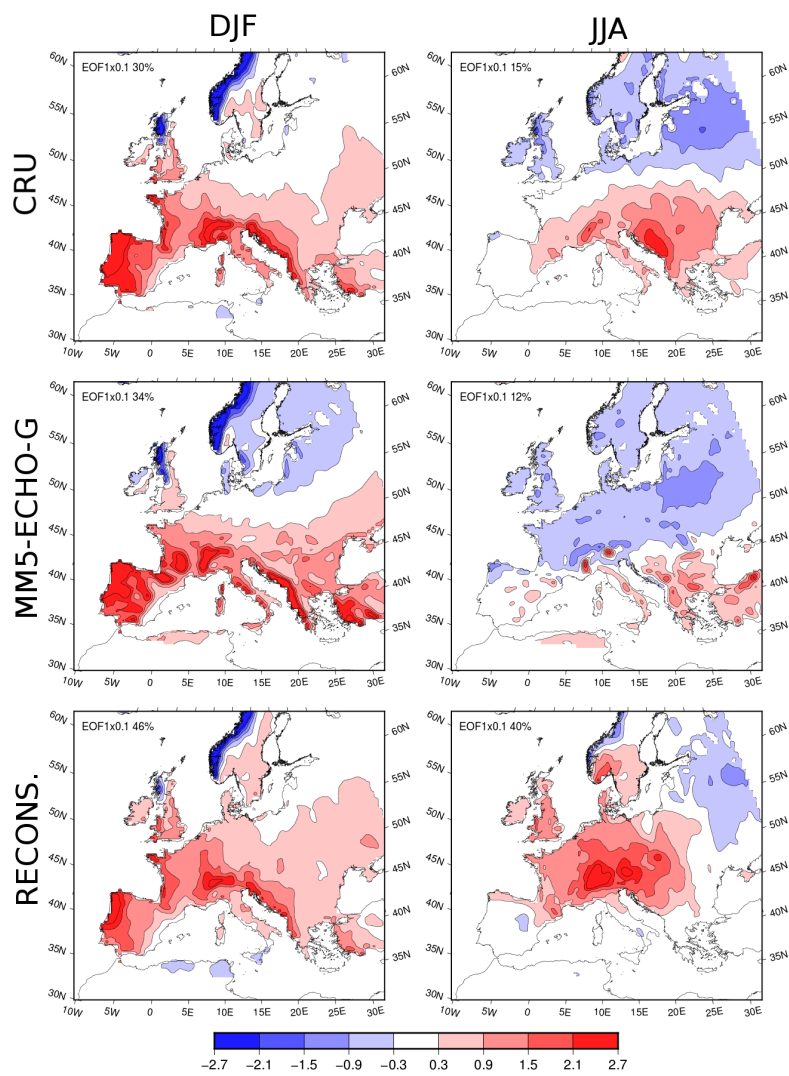


Figure 9. As Fig. 8 but for PRE. The units are mm/month.

dient shown in other data sets (see left column in Fig. 9). However, the imprint of orography that can be appreciated in the RCM is not seen in the GCM, resulting in an too spatially homogeneous pattern. In summer not only the spatial structure is not realistic, but also the main variability mode represents 26% of variance. Thus, results indicate that main variability modes are similar in both simulations, resulting from the strong forcing provided by the GCM through the boundaries of the domain. Still, the RCM is able to add regional details to the simulated fields. However, this depends on the variable and season. SAT is more strongly influenced by the driving conditions than precipitation, where the presence of complex orography is more important. This is especially evident for summer, where precipitation in the GCM is barely able to reproduce the observed patterns. These results agree with similar findings described in other RCM studies (Gómez-Navarro et al., 2011, 2014).

4.2 Dynamical consistency between variables

Canonical Correlation analysis (CCA) provides insight in the interrelation between different variables in the spatial domain. Comparing observed relationships with the corresponding simulated ones provides an assessment of the model skill. Evaluating these relationships in reconstructions of different variables gives an indication of the consistency among independent reconstructions (e.g. Luterbacher et al., 2010a). Figure 10 shows the canonical pair of patterns of SLP and SAT, and of SLP and PRE, with the largest canonical correlation, as simulated by the MM5-ECHO-G and their counterpart in the observational record in winter. Note that in summer the evolution of temperature and especially precipitation is driven to a lesser by the large-scale circulation. This is reflected by small canonical correlations. Hence, CCA is more useful for the winter season, and therefore only results for this season are discussed in detail.

Figure 10 shows the results for the observations and the simulation in the control period. Considering the first canonical pair of SLP and SAT (top row), the canonical correlation is 0.93 for the observations. The patterns represent 42% of total variance for SLP and 53% for SAT, respectively. The SLP resembles the NAO pattern and is related to a North-South gradient pattern in SAT. The physical explanation for this correlation is the well-known relationship between NAO and European temperature: a more zonal circulation in the North of Europe advects oceanic warm and moist air eastwards, leading to a positive temperature and precipitation anomaly in Northern Europe (Luterbacher et al., 2010a).

A similar SLP pattern and physical mechanism is found for the SLP-PRE pair (third column), with a correlation of 0.95 (Fig. 10). The SLP-PRE pair roughly resembles the SLP-SAT pair, although the zonal circulation is shifted southwards. Despite the fact that the zonal circulation supports the same physical relation between variables, in this case the canonical correlation is lower ($\rho=0.75$). The SAT pattern represents a large amount of variance and indeed resembles the leading EOF (see Fig. 8). The leading canonical pair of SLP-PRE exhibits a centre of high pressures in the North Atlantic which

reinforces the northwestern component of wind and is responsible for increasing precipitation in Western Europe, whereas it produces precipitation deficits in Norway and Turkey. This mechanism results in a strong link, producing a correlation of 0.91, although it explains a relatively small amount of winter precipitation variability (only 19% in the simulation).

Using the period 1750-1990, where also SLP reconstructions are available, show that none of the patterns for the longer period resembles perfectly the pair in the observations (compare Figs. 10 and 11), indicating that relationships between variables are sensitive to the period used. There are two potential reasons for this lack of robustness: First, the strong forcing in the 20th century may influence the canonical pairs either due to the strong anthropogenic trend in the zonal circulation in the driving simulation or due to a strong trend-component in the temperature field. Second, we have to keep in mind the simplified covariance and the potentially reduced signal in the reconstruction. The simulated canonical pair of SLP-SAT has a canonical correlation of 0.79 whereas the correlation for the reconstruction is 0.28. Again, the canonical pairs appear to be dominated by the temperature variability. The leading pairs for reconstruction and simulation both show a temperature gradient from the South-West to the North-East, which is dynamically related to a slight wave-like disturbance of the zonal flow and related changes in the advection of air masses. The reconstruction and the simulation disagree on the location and character of flow centres.

The first SLP-PRE pair in the simulation (fourth row in Fig. 11), corresponds to the second canonical pair over the 1948–1990 period in the observations (not shown). Note that the first two pairs derived from observations are very similar, especially with respect to canonical correlations but also considering the representation of variances. However, the second pair represents more SLP-variance than the first one. The separation between both pairs is more distinct in the longer period of analysis and in that case the ranking of the two leading pairs is exchanged. Hence we show the third canonical pair for the reconstruction, which is the apparent dynamic equivalent to the simulated one but shows much smaller canonical correlations (0.12 in the reconstruction and 0.89 in the simulation) while representing broadly consistent amount of variance. The small correlation signals that dynamical relations between both patterns may be weak. Indeed we would expect the NAO-like SLP-pattern to link the intensified zonal flow to a decrease of precipitation in Southern Europe, which is opposite to the pattern implied by the reconstructed pair.

5 Discussion and conclusions

This study investigates agreements and disagreements between a regional climate (high-resolution) simulation for Europe and empirical proxy based reconstructions for SAT, PRE and SLP from the 16th century to the 20th century. Our analyses complement the work by Gómez-Navarro et al. (2013), who compared the same simulation to observations for the 20th century.

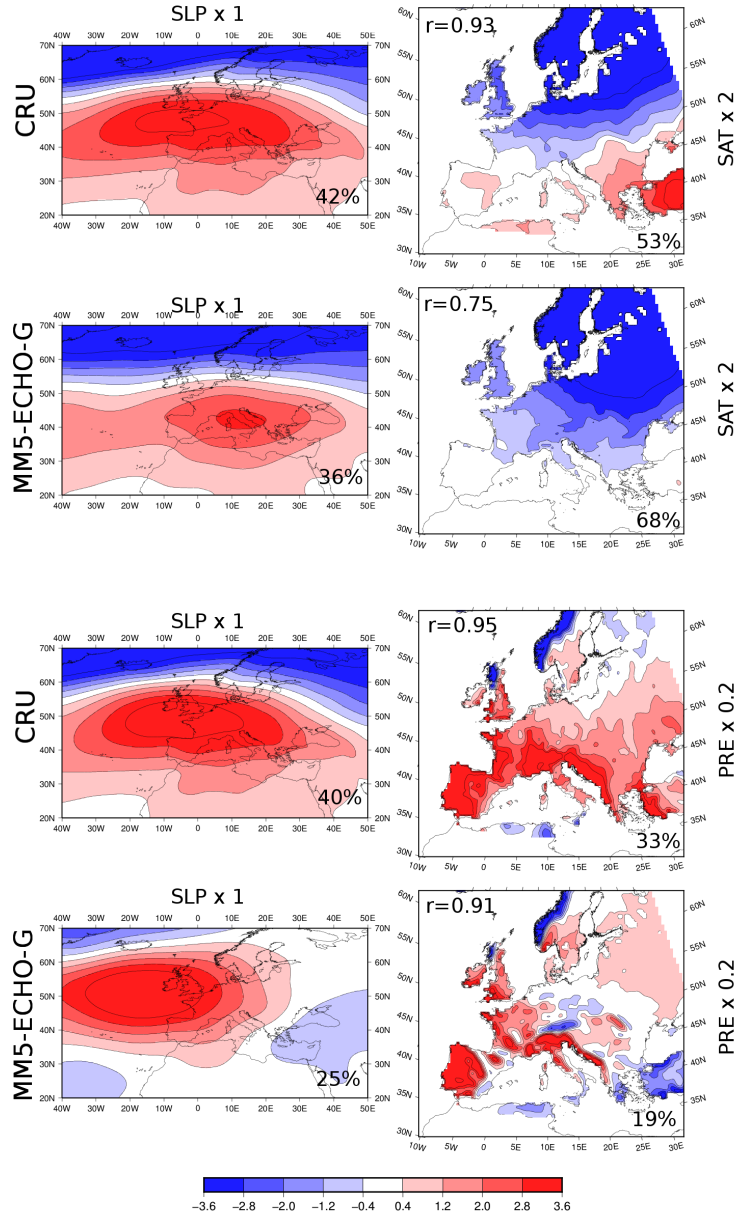


Figure 10. Canonical correlation pattern pairs of SLP and SAT (rows 1 and 2), and SLP and precipitation (rows 3 and 4) in winter. Each panel depicts the percentage of variance explained by each pattern and the canonical correlation associated with the pair. The results are calculated in the observational record (rows 1 and 3) and in the MM5-ECHO-G data set (rows 2 and 4) during the period 1901–1990. Note that the SLP has been obtained directly from the driving GCM, since the window of interest lies outside the RCM domain. As in Figs. 8 and 9, the patterns have been multiplied by a scaling factor that allows using the same color scale in every map. The SAT units are °C, SLP is shown in Pa, whereas precipitation units are mm/month.

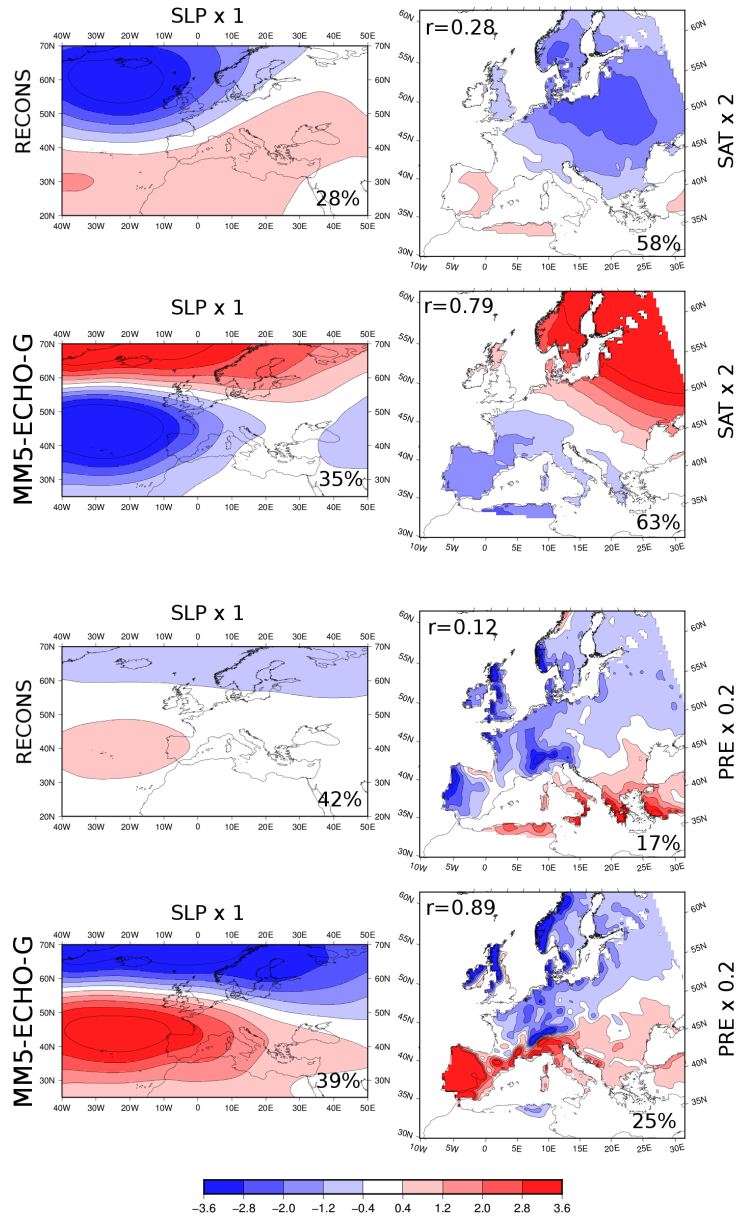


Figure 11. As in Fig. 10 but for the simulation and reconstructions. The calculations are based on the overlap period of the simulation and the SLP reconstructions, 1750–1990

Results indicate biases in regional means, especially noteworthy for summer temperature and winter precipitation. The biases between the simulation and reconstructions are similar to those described when comparing the model with an observational data set. In part they are explained by an enhanced zonal circulation in the GCM simulation that cannot substantially be ameliorated by the RCM, rather than by deficiencies within the reconstructions. Although reconstructions and the simulation seem to correctly reproduce most of the spatio-temporal variability, there is little agreement in their temporal evolution. The mismatch in the temperature, especially in the last decades, can originate from the missing anthropogenic aerosol forcing in the simulation. Additionally, early instrumental time series can show warm biases caused by the lack of modern thermometer screens (Frank et al., 2007a, b). Although we do not necessarily expect the reconstructed and simulated temperature evolution to agree in the earlier periods due to the potentially dominant internal variability, we also acknowledge that the lack of stratospheric dynamics in both the regional and the global simulation may account for some disagreement. Specifically, a too low top atmospheric layer in the model and no ozone chemistry reduce the ability of the model to correctly represent the potential top-down influences of solar activity changes on the atmospheric circulation in the North Atlantic sector, e.g., the North Atlantic Oscillation and in turn European climate variability (Shindell et al., 2001; Anet et al., 2013). Finally, the simplification of using reduced TSI for volcanic forcing might be an additional source of errors reducing the agreement between the simulation and reconstructions.

Obviously, the reconstructions also suffer from uncertainties, which have to be considered in addressing the reliability of the simulation by comparing to the proxy-based data sources. A prominent disagreement is the winter warming trend within the first half of the 18th century (Jones and Briffa, 2006), which stands out in the reconstructions but lacks in the simulation. This disagreement could be an indication of a too simplistic simulated climate, which is not able to produce extreme situations comparable to this event recorded in the reconstructions. Also internal variability could dominate the temporal evolution, effectively hiding the imprint of external forcing on the regional scale. A further source of error complicating the comparison between models and reconstructions relates to method-specific non-climatic errors. These can be related to simplified physics and a too coarse resolution in the models and proxy-type specific uncertainties in empirical reconstructions.

Internal variability, reconstruction uncertainty and potential shortcomings of the simulation in representing forced climate may also explain the disagreement in the magnitude of change between recent decades and the periods of the Maunder and Dalton minima. Again, the lack of 20th century anthropogenic aerosol forcing is likely the most important factor.

EOF and CCA analysis unveiled the lack of dynamic consistency between reconstructions and the weak explanatory power of dominant canonical pairs. Although this is not surprising, it highlights the large uncertainties in our estimates about past climates. This further implies that we should not expect to understand past climate changes based on one data source alone. On the other hand, the plausibility of simulated dynamics has to be assessed through tests with proxy-based hypotheses.

Other assessments of consistency among independent reconstructions have been carried out in the literature. Casty et al. (2007) employed gridded reconstructions of SAT, precipitation and geopotential height at 500 hPa to investigate combined patterns of climate variability over Europe for the 1766–2000 period. A prominent difference with the data sets employed in the present analysis is that the three reconstructions employed by Casty et al. (2007) use completely independent indicators, entirely based on instrumental data for each variable. This reduces the length of the reconstructions, but in turn ensures independence, which enabled the authors to evaluate the consistency between reconstructions through EOF analysis applied to the combined fields of the three variables. The authors reported similar NAO-like behaviour as that described in this study for the observations and simulations, with the large-scale flow driving seasonal temperature and precipitation over Europe, especially in winter. They also analysed the co-variability between SAT and precipitation. This study carefully avoids establishing such link, since the data sets used here are not fully independent (both SAT and precipitation reconstructions share some indications). However, the CCA approach adopted here allows studying the co-variability between SLP and the other two variables. The weaker and physically inconsistent link we identify, especially with respect to the Pauling et al. (2006) reconstruction, raises concerns about the reliability of these reconstructions.

Coordinated reconstruction efforts as, for instance, related to PAGES2k (PAGES 2k Consortium, 2013) will increase the number of available proxy records. This, in conjunction with newly developed reconstruction methods, is expected to provide more realistic uncertainty estimates of the spatial fields and spatially averaged reconstructions. In addition, proxy system models (e.g. Evans et al., 2013) will provide a better basis for proxy-model comparison as they enable a direct modelling of the proxy under consideration within the virtual world of a climate model. This may help to evaluate e.g. the stationarity of proxy-climate relationships and the different sources and degrees of uncertainty implicit in empirical reconstruction methods.

In conclusion, although regional climates are generally better represented by the RCM compared to the driving GCM (Gómez-Navarro et al., 2013), the downscaling is not able to compensate for biases in the driving circulation. This leads to biases in the comparison with the reconstructions that are clearly attributable to model deficiencies. However, we cannot describe simulated and reconstructed anomalies with respect to today’s climate as generally inconsistent, although the temporal evolution is different enough to raise concerns over the ability of the simulation to produce exceptionally anomalous situations as those recorded by the SAT reconstructions during the first decades of the 18th century. Further, the dynamical inconsistencies that we identify between the reconstructions of SLP and SAT and PRE hamper addressing the reliability of forced changes in the dynamics. It remains an open question whether a lack of common forced signals is due to weak forcing effects relative to the internal variability of the climate system, due to erroneous representation of climate dynamics in the model or due to uncertainty in the reconstructions.

630 *Acknowledgements.* This work was funded by the PRIME2 project (priority program INTERDYNAMIK, German Research Foundation) and the SPEQTRES project (Spanish Ministry of Economy and Competitiveness, ref. CGL2011-29672-C02-02). Jürg Luterbacher was supported by the LOEWE Large Scale Integrated Program (Excellency in research for the future of Hesse), FACE2FACE – Folgen des Klimawandels, Anpassung an den Klimawandel und Verminderung der Treibhausgas-Emissionen bis 2050. Juan José Gómez-Navarro thanks the
635 funding provided by the Oeschger Centre for Climate Change Research and the Mobiliar lab for climate risks and natural hazards (Mobilab).

References

- Andreae, M. O., Jones, C. D., and Cox, P. M.: Strong present-day aerosol cooling implies a hot future, *Nature*, 435, 1187–1190, doi:10.1038/nature03671, 2005.
- 640 Anet, J. G., Muthers, S., Rozanov, E., Raible, C. C., Peter, T., Stenke, A., Shapiro, A. I., Beer, J., Steinhilber, F., Brönnimann, S., Arfeuille, F., Brugnara, Y., and Schmutz, W.: Forcing of stratospheric chemistry and dynamics during the Dalton Minimum, *Atmospheric Chemistry and Physics*, 13, 10 951–10 967, doi:10.5194/acp-13-10951-2013, 2013.
- 645 Anet, J. G., Muthers, S., Rozanov, E. V., Raible, C. C., Stenke, A., Shapiro, A. I., Brönnimann, S., Arfeuille, F., Brugnara, Y., Beer, J., Steinhilber, F., Schmutz, W., and Peter, T.: Impact of solar versus volcanic activity variations on tropospheric temperatures and precipitation during the Dalton Minimum, *Clim. Past*, 10, 921–938, doi:10.5194/cp-10-921-2014, 2014.
- Casty, C., Raible, C. C., Stocker, T. F., Wanner, H., and Luterbacher, J.: A European pattern climatology 1766–2000, *Climate Dynamics*, 29, 791–805, doi:10.1007/s00382-007-0257-6, 2007.
- 650 Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, *Climatic Change*, 81, 7–30, doi:10.1007/s10584-006-9210-7, 2007.
- Eden, J. M., Widmann, M., Maraun, D., and Vrac, M.: Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing, *Journal of Geophysical Research: Atmospheres*, pp. 11 040–11 053, doi:10.1002/2014JD021732, 2014.
- 655 Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Science Reviews*, 76, 16–28, doi:10.1016/j.quascirev.2013.05.024, 2013.
- Frank, D., Büntgen, U., Böhm, R., Maugeri, M., and Esper, J.: Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at the moving target, *Quaternary Sci. Rev.*, 26, 3298–3310, doi:10.1016/j.quascirev.2007.08.002, 2007a.
- 660 Frank, D., Esper, J., and Cook, E. R.: Adjustment for proxy number and coherence in large-scale temperature reconstruction, *Geophysical Research Letters*, 34, L16 709, doi:10.1029/2007GL030571, 2007b.
- Franke, J., González-Rouco, J. F., Frank, D., and Graham, N. E.: 200 years of European temperature variability: insights from and tests of the proxy surrogate reconstruction, *Climate Dynamics*, 37, 133–150, doi:10.1007/s00382-010-0802-6, 2010.
- 665 Franke, J., Frank, D., Raible, C. C., Esper, J., and Brönnimann, S.: Spectral biases in tree-ring climate proxies, *Nature Climate Change*, 3, 360–364, doi:10.1038/nclimate1816, 2013.
- Gómez-Navarro, J. J. and Zorita, E.: Atmospheric annular modes in simulations over the past millennium: No long-term response to external forcing, *Geophysical Research Letters*, 40, 3232–3236, doi:10.1002/grl.50628, 2013.
- 670 Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P., Lorente-Plazas, R., González-Rouco, J. F., and Zorita, E.: A regional climate simulation over the Iberian Peninsula for the last millennium, *Clim. Past*, 7, 451–472, doi:10.5194/cp-7-451-2011, 2011.

- 675 Gómez-Navarro, J. J., Montávez, J., Jiménez-Guerrero, P., Jerez, S., Lorente-Plazas, R., González-Rouco, J.,
and Zorita, E.: Internal and external variability in regional simulations of the Iberian Peninsula climate over
the last millennium, *Clim. Past*, 8, 25–36, doi:10.5194/cp-8-25-2012, 2012.
- Gómez-Navarro, J. J., Montávez, J. P., Wagner, S., and Zorita, E.: A regional climate palaeosimulation for
Europe in the period 1500–1990 – Part 1: Model validation, *Clim. Past*, 9, 1667–1682, doi:10.5194/cp-9-
680 1667-2013, 2013.
- Gómez-Navarro, J. J., Werner, J., Wagner, S., Luterbacher, J., and Zorita, E.: Establishing the skill of climate
field reconstruction techniques for precipitation with pseudoproxy experiments, *Climate Dynamics*, pp. 1–
19, doi:10.1007/s00382-014-2388-x, 2014.
- Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., and Mann, M. E.: Using paleoclimate proxy-data
685 to select optimal realisations in an ensemble of simulations of the climate of the past millennium, *Climate
Dynamics*, 27, 165–184, doi:10.1007/s00382-006-0128-6, 2006.
- Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y.: The medieval climate anomaly in
Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with
data assimilation, *Global and Planetary Change*, 84–85, 35–47, doi:10.1016/j.gloplacha.2011.07.002, 2012.
- 690 Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations
– the CRU TS3.10 Dataset, *International Journal of Climatology*, 34, 623–642, doi:10.1002/joc.3711, 2014.
- Jones, P. D. and Briffa, K. R.: Unusual Climate in Northwest Europe During the Period 1730 to 1745 Based
on Instrumental and Documentary Data, *Climatic Change*, 79, 361–379, doi:10.1007/s10584-006-9078-6,
2006.
- 695 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White,
G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak,
J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Re-
analysis Project, *Bulletin of the American Meteorological Society*, 77, 437–471, doi:10.1175/1520-
0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- 700 Kim, J.-W., Chang, J.-T., Baker, N. L., Wilks, D. S., and Gates, W. L.: The Statistical Problem of Climate
Inversion: Determination of the Relationship between Local and Large-Scale Climate, *Monthly Weather
Review*, 112, 2069–2077, doi:10.1175/1520-0493(1984)112<2069:TSPOCI>2.0.CO;2, 1984.
- Küttel, M., Xoplaki, E., Gallego, D., Luterbacher, J., García-Herrera, R., Allan, R., Barriendos, M., Jones,
P. D., Wheeler, D., and Wanner, H.: The importance of ship log data: reconstructing North Atlantic, Eu-
705 ropean and Mediterranean sea level pressure fields back to 1750, *Climate Dynamics*, 34, 1115–1128,
doi:10.1007/s00382-009-0577-9, 2010.
- Luterbacher, J., Xoplaki, E., Dietrich, D., Rickli, R., Jacobeit, J., Beck, C., Gyalistras, D., Schmutz, C., and
Wanner, H.: Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to
1500, *Climate Dynamics*, 18, 545–561, doi:10.1007/s00382-001-0196-6, 2002.
- 710 Luterbacher, J., Liniger, M. A., Menzel, A., Estrella, N., Della-Marta, P. M., Pfister, C., Rutishauser,
T., and Xoplaki, E.: Exceptional European warmth of autumn 2006 and winter 2007: Historical con-
text, the underlying dynamics, and its phenological impacts, *Geophysical Research Letters*, 34, L12 704,
doi:10.1029/2007GL029951, 2007.

- Luterbacher, J., Koenig, S. J., Franke, J., Schrier, G. v. d., Zorita, E., Moberg, A., Jacobeit, J., Della-Marta, P. M.,
715 Küttel, M., Xoplaki, E., Wheeler, D., Rutishauser, T., Stössel, M., Wanner, H., Brázdil, R., Dobrovolný, P.,
Camuffo, D., Bertolin, C., Engelen, A. v., Gonzalez-Rouco, F. J., Wilson, R., Pfister, C., Limanówka, D.,
Nordli, O., Leijonhufvud, L., Söderberg, J., Allan, R., Barriendos, M., Glaser, R., Riemann, D., Hao, Z.,
and Zerefos, C. S.: Circulation dynamics and its influence on European and Mediterranean January–April
720 climate over the past half millennium: results and insights from instrumental data, documentary evidence
and coupled climate models, *Climatic Change*, 101, 201–234, doi:10.1007/s10584-009-9782-0, 2010a.
- Luterbacher, J., Xoplaki, E., Küttel, M., Zorita, E., González-Rouco, J. F., Jones, P. D., Stössel, M., Rutishauser,
T., Wanner, H., Wibig, J., and Przybylak, R.: Climate Change in Poland in the Past Centuries and its Rela-
tionship to European Climate: Evidence from Reconstructions and Coupled Climate Models, in: *The Polish
Climate in the European Context: An Historical Overview*, edited by Przybylak, R., pp. 3–39, Springer
725 Netherlands, 2010b.
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual tem-
perature variability, trends, and extremes since 1500, *Science*, 303, 1499–1503, 2004.
- PAGES 2k Consortium: Continental-scale temperature variability during the past two millennia, *Nature Geo-
science*, 6, 339–346, doi:10.1038/ngeo1797, 2013.
- 730 PAGES2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k re-
gional temperature reconstructions over the past millennium, *Climate of the Past Discussions*, 11, 2483–
2555, doi:10.5194/cpd-11-2483-2015, 2015.
- Pauling, A., Luterbacher, J., Casty, C., and Wanner, H.: Five hundred years of gridded high-resolution precip-
itation reconstructions over Europe and the connection to large-scale circulation , *Climate Dynamics*, 26,
735 387–405, doi:10.1007/s00382-005-0090-8, 2006.
- Phipps, S. J., McGregor, H. V., Gergis, J., Gallant, A. J. E., Neukom, R., Stevenson, S., Ackerley, D., Brown,
J. R., Fischer, M. J., and van Ommen, T. D.: Paleoclimate Data–Model Comparison and the Role of Climate
Forcings over the Past 1500 Years*, *Journal of Climate*, 26, 6915–6936, doi:10.1175/JCLI-D-12-00108.1,
2013.
- 740 Raible, C. C., Casty, C., Luterbacher, J., Pauling, A., Esper, J., Frank, D. C., Büntgen, U., Roesch, A. C.,
Tschuck, P., Wild, M., Vidale, P.-L., Schär, C., and Wanner, H.: Climate variability-observations, reconstruc-
tions, and model simulations for the Atlantic-European and Alpine region from 1500-2100 AD, *Climatic
Change*, 79, 9–29, doi:10.1007/978-1-4020-5714-4_2, 2006.
- Raible, C. C., Lehner, F., González-Rouco, J. F., and Fernández-Donado, L.: Changing correlation structures
745 of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD, *Clim. Past*, 10, 537–550,
doi:10.5194/cp-10-537-2014, 2014.
- Schimanke, S., Meier, H. E. M., Kjellström, E., Strandberg, G., and Hordoir, R.: The climate in the Baltic
Sea region during the last millennium simulated with a regional climate model, *Clim. Past*, 8, 1419–1433,
doi:10.5194/cp-8-1419-2012, 2012.
- 750 Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P.,
Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C.,
Thompson, D., Timmermann, A., Tremblay, L.-B., and Yiou, P.: Using palaeo-climate comparisons to con-
strain future projections in CMIP5, *Clim. Past*, 10, 221–250, doi:10.5194/cp-10-221-2014, 2014.

- Shindell, D. T., Schmidt, G. A., Mann, M. E., Rind, D., and Waple, A.: Solar Forcing of Regional Climate
755 Change During the Maunder Minimum, *Science*, 294, 2149–2152, doi:10.1126/science.1064363, 2001.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together
the past: statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, 35, 1–22,
doi:10.1016/j.quascirev.2012.01.012, 2012.
- Vieira, L. E. A., Solanki, S. K., Krivova, N. A., and Usoskin, I.: Evolution of the solar irradiance during the
760 Holocene, *Astronomy & Astrophysics*, 531, A6, doi:10.1051/0004-6361/201015843, 2011.
- von Storch, H. and Zwiers, F.: *Statistical Analysis in Climate Research*, Cambridge University Press, 1999.
- Wagner, S., Fast, I., and Kaspar, F.: Comparison of 20th century and pre-industrial climate over South America
in regional model simulations, *Clim. Past*, 8, 1599–1620, doi:10.5194/cp-8-1599-2012, 2012.
- Wetter, O., Pfister, C., Luterbacher, J., Werner, J., Siegfried, W., Glaser, R., Riemann, D., Himmelsbach, I.,
765 Contino, A., Burmeister, K., Litzenburger, L., Barriendos, M., Brazdil, R., Kiss, A., Camenisch, C., Li-
manowka, D., Pribyl, K., Labbé, T., Retsö, D., Bieber, U., Rohr, C., Spring, J., Nordli, O., Soderberg, J., and
Alcoforado, M.: The year-long unprecedented European heat and drought of 1540—a worst case scenario,
Climatic Change, 125, 349–363, 2014.
- Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J.: Using data assimilation to study
770 extratropical Northern Hemisphere climate over the last millennium, *Clim. Past*, 6, 627–644, doi:10.5194/cp-
6-627-2010, 2010.
- Yoshimori, M., Stocker, T. F., Raible, C. C., and Renold, M.: Externally Forced and Internal Variabil-
ity in Ensemble Climate Simulations of the Maunder Minimum, *Journal of Climate*, 18, 4253–4270,
doi:10.1175/JCLI3537.1, 2005.
- 775 Zanchettin, D., Timmreck, C., Graf, H.-F., Rubino, A., Lorenz, S., Lohmann, K., Krüger, K., and Jungclaus,
J. H.: Bi-decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic
eruptions, *Climate Dynamics*, 39, 419–444, doi:10.1007/s00382-011-1167-1, 2012.
- Zorita, E., Moberg, A., Leijonhufvud, L., Wilson, R., R., B., Dobrovolný, P., Luterbacher, J., Böhm, R., Pfister,
C., Glaser, R., Söderberg, J., and González-Rouco, F.: European temperature records of the past five cen-
780 turies based on documentary information compared to climate simulations, *Climatic Change*, 101, 143–168,
doi:10.1007/s10584-010-9824-7, 2010.