Reconciling reconstructed and simulated features of the winter Pacific-North-American pattern in the early 19th century

Davide Zanchettin[1,2*], Oliver Bothe[3], Flavio Lehner[4†], Pablo Ortega[5], Christoph C. Raible[4], and Didier Swingedouw[6]

[1]Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

[2]University of Venice, Dept. of Environmental Sciences, Informatics and Statistics, Calle Larga Santa Marta, Dorsoduro 2137, 30123 Venice, Italy

[3]HZG, Helmholtz Center Geesthacht, Max Planck Strasse 1, 21502 Geesthacht, Germany

[4]Climate and Environmental Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland

[5]LOCEAN-IPSL/UPMC, 4, place Jussieu, Paris, France

[6]UMR CNRS 5805 EPOC - OASU - Université de Bordeaux, Allee Geoffroy St Hilaire, 33615 Pessac cedex, France

* corresponding author: davide.zanchettin@mpimet.mpg.de

[†]currently at National Center for Atmospheric Research, Boulder, USA

Abstract

Reconstructions of past climate behavior often describe prominent anomalous periods that are not necessarily captured in climate simulations. Here, we illustrate the contrast between an interdecadal strong positive phase of the winter Pacific/North American pattern (PNA) in the early 19th century that is described by a PNA reconstruction based on tree-rings from northwestern North America, and a slight tendency towards negative winter PNA anomalies during the same period in an ensemble of state-of-the-art coupled climate simulations. Additionally, a pseudo-proxy investigation with the same simulation ensemble allows assessing the robustness of PNA reconstructions using solely geophysical predictors from northwestern North America for the last millennium. The reconstructed early-19th-century positive PNA anomaly emerges as a potentially reliable feature, although the pseudo-reconstructions are subject to a number of sources of uncertainty and deficiencies are highlighted especially at multidecadal and centennial timescales. The pseudo-reconstructions demonstrate that the early-19th-century discrepancy between reconstructed and simulated PNA does not stem from the reconstruction process. Instead, reconstructed and simulated features of the early-19th-century PNA can be reconciled by interpreting the reconstructed evolution during this time as an expression of internal climate variability, which is unlikely to be reproduced in its exact temporal occurrence by a small ensemble of climate simulations. However, firm attribution of the reconstructed PNA anomaly is hampered by known limitations and deficiencies of coupled climate models and uncertainties in the early-19th-century external forcing and background climate state.

1. Introduction

The Pacific/North American pattern (PNA) is one of the dominant modes of interannual winter atmospheric variability of the northern extratropics (*e.g.*, Barnston and Livezey 1987; Wallace and Gutzler, 1981). It strongly affects the weather and the hydroclimate of the North American continent, and contributes to the atmospheric bridge linking Pacific and Atlantic climate variability (*e.g.*, Raible et al. 2001; Pinto et al., 2010; Baxter and Nigam, 2013). The behavior of this large-scale atmospheric circulation pattern before the observational period and its sensitivity to natural external forcing are less understood compared to other dominant climate modes, partly due to the limited number and temporal coverage of available PNA reconstructions. In fact, only one major winter PNA reconstruction, based on tree-rings from northwestern North America, is available and only goes back to 1725 (Trouet and Taylor, 2010, hereafter: TT2010) (see section 2.1 for details). Here, we investigate the PNA features described by a multi-model ensemble of state-of-the-art climate simulations of the last millennium and use pseudo-proxy experiments (*e.g.*, Lehner et al., 2012; Smerdon, 2012) applied to the same ensemble to improve our understanding of the PNA's behavior during the pre-industrial period, especially during the early 19th century, and to investigate compatibility between climate simulations and reconstructions.

The PNA pattern consists of a wave train spanning from the subtropical northeastern Pacific to the Gulf of Alaska, northwestern North America and the southeastern United States through centers of action of alternating sign (Figure 1). Accordingly, a classical definition of the PNA index is the sum of the differences between its positive and negative centers of action (Wallace and Gutzler, 1981). The PNA can be interpreted as an amplification and dampening of the climatological stationary wave characterizing the pattern of the polar jet across North America (*e.g.*, Notaro et al., 2006), which explains its reduced importance during boreal summer. The positive phase of the PNA includes an anomalously deep Aleutian low and an enhanced ridge-trough pattern across North America. It produces above-average temperatures over northwestern North America due to the stronger ridge over the North American Rockies with associated northward diversion of the westerly flow, and below-average temperatures and drier conditions across the south-central and southeastern United States due to increased southward penetration of cold Arctic air masses. The signature is reversed for the negative phase of the PNA.

On sub-monthly timescales the PNA variability and predictability are largely determined by internal dynamics of the mid-latitude atmosphere, while on longer timescales they are most prominently controlled by forcing from sea-surface temperature (SST) signals from the tropical Pacific (Younas and Tang, 2013). Horel and Wallace (1981) were the first to identify a connection between the PNA and the equatorial El Niño-Southern Oscillation (ENSO). Since then, observational and modelling studies have revealed that boundary conditions relevant for the PNA also include low-frequency SST signals in the extra-tropical North Pacific (Yu and Zwiers, 2007; Yu et al., 2007), remote forcing from the North Atlantic (Baxter and Nigam, 2013), and upstream conditions determined by the East Asian jet (*e.g.*, Gong et al., 2007).

Climate simulations of the last millennium indicate increased likelihood of a significantly weaker Aleutian low after strong tropical volcanic eruptions, suggesting that the PNA can dynamically respond to volcanic forcing on interannual to decadal time scales (Zanchettin et al., 2012; Wang et al., 2012). A connection between PNA variability and natural forcing is also suggested by the TT2010 reconstruction, which shows a prolonged strong positive phase of the PNA during the early 19^{th} century. Indeed, this period was characterized by a close succession of strong volcanic eruptions concomitant with a phase of weak solar activity, both contributing to exceptionally cold climate conditions (Cole-Dai et al., 2009). However, TT2010 attributed the early-19^{th}-century anomalous PNA phase to the decreased solar irradiance during the Dalton Minimum of solar activity (ca.

1790-1830), without discussing possible implications from the concomitant volcanic cluster.

The selection of proxy locations is crucial for the robustness and reliability of reconstructions of large-scale circulation modes (Lehner et al., 2012). Decadal-scale shifts in the centers of action of atmospheric modes like the North Atlantic Oscillation (NAO) or the PNA are associated with non-stationarities in the imprint of such teleconnection patterns on local precipitation and temperature (Raible et al. 2006; Coats, et al., 2013; Moore et al. 2013; Raible et al., 2014). Accordingly, the ring-width response to atmospheric modes like the PNA and the NAO is spatially heterogeneous due to the complex causal chain linking climate modes, local environment and seasonal tree growth (St. George, 2014). Reflecting such heterogeneity, the prolonged positive PNA phase in the early 19th century becomes less prominent if proxies from other PNA-sensitive regions are considered, such as river catchments in western-central British Columbia (Starheim et al., 2013) or lakes from the north-eastern United States (Hubeny et al., 2010).

Twentieth-Century-Reanalysis data suggest that the PNA centers of action are less variable in space than, *e.g.*, the NAO (Raible et al. 2014). However, the risk of insufficient coverage and representation of its different centers of action – hence of poor reconstruction – is considerable for the PNA due to the complexity of its pattern and the strong interdependencies with surrounding or even superposing modes of variability. For instance, over the last six decades and especially in winter the PNA is practically indistinguishable from the inverted North Pacific Index (NPI) describing the sea-level pressure (SLP) variability in the Aleutian low region. However, there is no indication so far about whether the NPI and the PNA (and their signatures on regional temperature and precipitation) can become distinguishable over periods of decades or longer. Similarly, late-winter temperature reconstructions in western North America, *i.e.*, a region where the PNA climatic imprint is strong, have been recently used to test whether strong tropical volcanic eruptions induce a preferred phasing of ENSO (Wahl et al., 2014). However, the simulated teleconnection between ENSO and North America climate is nonstationary on multidecadal timescales (Coats et al., 2013). The possible superposition of regional climate signatures of different large-scale modes (like ENSO, PNA and NPI) and their possible nonstationarity poses a challenge for any reconstruction attempt using climate-proxies from affected regions, in particular to produce robust and unambiguous reconstructions. There is therefore a need to assess the robustness of proxy-based PNA reconstructions: If the TT2010 reconstruction was found to accurately capture the past PNA behavior, the reconstructed prolonged strong positive PNA phase during the early 19th century would represent an important feature to address in coupled climate simulations. Its robust reproduction by climate simulations would be strong evidence for its externally-driven nature, while the opposite would suggest two possibilities: either an episodic excitation consistent with internal variability, or limited realism of climate models due to common deficiencies.

Thus, this study aims at answering the following questions: Is the reconstructed prolonged, strong positive PNA phase during the early 19th century reliable? What can we learn from available climate simulations of the last millennium about its attribution? To answer these questions we compare simulated and observed/reconstructed PNA features, and perform a series of pseudo-proxy experiments on a multi-model ensemble. We also extend our pseudo-proxy investigation to determine whether margins exist to substantially improve the PNA reconstruction.

2. Data and methods

2.1 The TT2010 PNA reconstruction

The TT2010 reconstruction of the winter PNA covers the period 1725–1999. It is based on a multiple regression model using three winter climate sensitive tree-ring records from the western United States as predictors (TT2010). The predictors were sampled from three regions, whose location is marked by the green boxes in Figure 2 and that are hereafter referred to as “Alaska” (northern box, here defined as: [60-70]ºN, [120-160]ºW), “Montana” (middle box, [50-60]ºN, [115-135)ºW) and “Wyoming” (southern box, [35-50)ºN, 107.5-125ºW]), respectively. The Alaskan predictor is most sensitive to winter temperature; the predictor from Montana captures winter precipitation in a relatively high-elevation site whereas the predictor from Wyoming captures both autumn/winter precipitation and summer temperature in a relatively low-elevation, semi-arid site (TT2010). The latter two predictors show opposite sensitivity to precipitation despite their close location, highlighting the role of regional topographical features in determining the relationship between the biological sensor and the local environmental conditions. The combination of the selected three tree-ring series explains 49 % of the variance of the winter PNA index for the calibration period 1949-1999 (TT2010).

The TT2010 PNA reconstruction shows a prolonged period of positive PNA, with a peak in 1800–1820. The early-19^{th}-century positive PNA phase was interpreted as a response to the decreased solar irradiance since it coincides with the period of weak solar activity known as Dalton Minimum and since subsequent periods of weak solar activity similarly correspond to positive PNA anomalies (TT2010). Radiatively forced warming of eastern tropical Pacific SST associated with cold SST anomalies in the Aleutian Low region - corresponding to an in-phase behavior between positive anomalies of both ENSO and the Pacific Decadal Oscillation - was reported as a possible dynamical explanation for the connection between decreased solar irradiance and positive PNA phase. TT2010 did not discuss the possible implications of the strong volcanic eruptions during the early 19^{th} century. Possible mechanisms underlying volcanically-forced PNA variability include both tropical-extratropical coupling via volcanically-forced changes of ENSO (*e.g.*, Li et al., 2013) and extratropical processes via, *e.g.*, sea-ice responses in the Gulf of Alaska and the Bering Strait region (*e.g.*, Zanchettin et al., 2014).

2.2 Observational and simulated data

We use monthly-mean data obtained from the NCEP reanalysis (Kalnay et al., 1996; Kistler et al., 2001) for the period 1948-2013 as reference data for the observational period. The data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. The NCEP reanalysis dataset suits our needs since it encompasses the calibration period used in the TT2010 reconstruction. The Twentieth Century Reanalysis (Compo et al., 2011) extends further back in time, but the ensemble teleconnection patterns over the Pacific from this dataset are poorly constrained during the first half of the 20^{th} century (*e.g.*, Raible et al., 2014).

We include outputs from the *past1000* and the follow-up *historical* climate simulations from seven coupled general circulation and Earth system models contributing to the third phase of the Paleoclimate Modelling Intercomparison Project (PMIP3, Braconnot et al., 2012). All simulations are full-forcing simulations, *i.e.*, they describe the combined effects of all major natural and anthropogenic external forcing factors acting during the last millennium (Schmidt et al., 2013). Two simulations are considered for the GISS-E2-R model (hereafter referred to as GISS-E2-R24 and GISS-E2-R25, respectively), which differ in the considered external forcing inputs. Table 1 provides a summary of the main characteristics of the models and simulations considered.

Bothe et al. (2013) provide an assessment of the probabilistic and climatological consistency of the PMIP3-past1000 simulations relative to proxy-based reconstructions under the paradigm of a statistically indistinguishable ensemble. They diagnose distributional inconsistencies between ensemble-simulated surface-air temperatures and the global temperature field reconstruction of Mann et al. (2009) over large areas of the globe, including PNA-sensitive regions over North America (see their Figure 1). These full-period inconsistencies originate mainly from differences in multicentennial to millennial trends (Bothe et al., 2013). By contrast, the ensemble was found to be probabilistically consistent with the reconstructed annual temperatures for the North America Southwest since year 1500 (Wahl and Smerdon, 2012).

2.3 Indices and definitions

The following indices and definitions are considered based on monthly-mean data:

- the PNA is calculated using the modified pointwise method currently adopted by the NOAA-CPC and applied to 500 hPa geopotential height (Z500) data. The index is defined as $Z^*_{[15\text{-}25ºN;\ 180\text{-}220ºE]} - Z^*_{[40\text{-}50ºN;\ 180\text{-}220ºE]} + Z^*_{[45\text{-}60ºN;\ 235\text{-}255ºE]} - Z^*_{[25\text{-}35ºN;\ 270\text{-}290ºE]}$, where Z* denotes monthly Z500 anomalies from the respective climatological value, and the suffix [x] indicates spatial averaging over the domain x. We briefly discuss a different definition of the PNA index in section 4;
- the NAO index is calculated based on the latitude-longitude 2-box method by Stephenson et al. (2006) applied on Z500 data, *i.e.*, as the pressure difference between spatial averages over [20-55ºN; 90ºW-60ºE] and [55-90ºN; 90ºW-60ºE];
- the NPI is calculated using the definition from Trenberth and Hurrel (1994) applied to SLP data. The index is computed as the spatial SLP averaged over [30-

65ºN; 160-220ºE], so that positive phases of the index indicate a weaker-than-normal Aleutian low and the opposite holds for the negative phases;

- the SOI is calculated based on a modified version of the Tahiti-Darwin index. It is defined as the difference between the average SLP over the domains [20-15ºS; 147-152ºW] and [15-10ºS; 128.5-133.5ºE]. The SOI is here preferred to SST-based ENSO indices since we focus on the atmospheric component of ENSO.

Indices are not standardized by default in order to highlight inter-model differences in the climatology and in the amplitude of fluctuations associated with the indices.

2.4 Pseudo-proxy experiments

Pseudo-proxy experiments are conducted to test the robustness and potential skills of PNA reconstructions using solely geophysical predictors from northern North America. Millennium-scale transient simulations from climate models provide a long and physically consistent framework where paleoclimate reconstruction methods can be altered and evaluated systematically in absence of the spatial and temporal discontinuities of the real-world climate-proxy networks (Smerdon, 2012). In particular, they may allow determining an upper limit in the accuracy of the reconstruction of large-scale modes under limited spatial sampling.

Our pseudo-proxy experimental approach is meant as a generalization of the method used in TT2010. Specifically, the reconstruction is based on a multi-linear least-square regression model of general form: $y^t = \Sigma_{i=1:N} a_i x_i^t + y_0 + \varepsilon^t$, where y^t is the reconstructed value at time step t, N is the number of predictors, a_i is the regression coefficient of the i^{th} predictor x_i, y_0 is the intercept and ε^t is the residual at time step t. All data are normalized based on the full period before the pseudo-proxy experiments are conducted.

Pseudo-reconstructions are performed as follows: first, the pool of candidate predictors including temperature and precipitation data is determined by defining three regions over North America. Then, an ensemble of predictor-sets is built by iteratively (up to 1000 times) and randomly sampling one grid-point data from each region. Finally, an ensemble of PNA pseudo-reconstructions is obtained by using the built sets in a multi-linear regression. So, the robustness of PNA pseudo-reconstructions with a TT2010-like design is tested using predictor sets that mimic the quality of the TT2010 reconstruction.

Accordingly, we consider only pseudo-reconstructions with R^2 skill metric in the range between 0.45 and 0.55 for the calibration period (R^2_c), *i.e.*, the selection is based on calibration skills instead of on a preliminary screening of climate proxies.

We follow a perfect-model approach with noise-free predictors, and the considered range of R^2_c is meant to account for the possible effects of noise in the actual climate proxies. The inclusion of noise in the predictors and its influence on the results are briefly investigated with a series of pseudo-proxy experiments where predictors are artificially perturbed by different types and levels of noise (section 4). Skill metrics calculated for such noise-free predictor sets and regression models, but using other climate indices as validation target instead of the PNA, clarify whether these pseudo-reconstructions distinguish the PNA from other modes influencing North American regional climates. Additionally, the PNA pseudo-reconstructions pertaining to the upper quartile of R^2_c for each simulation provide a crude estimate of the quality of PNA reconstructions obtainable with a TT2010-like method for the given set of sampling regions. An exemplary different set of regions is also considered in the same reconstruction approach to assess whether regions not included in the TT2010 design may allow for a notable improvement in the accuracy and robustness of PNA reconstructions.

Skill metrics include R^2 and coefficient of error (CE) (Cook et al., 1994). CE is defined as $[1 - \Sigma_{t=1:M}(x_t-y_t)^2/\Sigma_{t=1:M}(x_t-x_{mv})^2]$, where x_t and y_t are the observed and the predicted index in year t, respectively. x_{mv} is the observed mean index over the validation period and M is the number of years in the validation period. R^2 values are also calculated for successive 30-year periods to highlight the robustness of the pseudo-reconstructions over different interdecadal periods.

Unlike in TT2010, the predictors sampled herein from the “Montana” (middle) and “Wyoming” (southern) boxes are winter temperature and precipitation, respectively. This choice guarantees that our pseudo-reconstructions encompass the desired R^2_c range in all models, which is hardly achieved for pseudo-proxies from these regions following the original definition. In particular, reconstruction skills considerably degrade if the predictor for “Wyoming” is defined as summer temperature instead of winter temperature (see section 2.1). This does not affect the generality of our conclusions, since we aim at testing PNA reconstructions based solely on local geophysical predictors from northwestern North America, not at replicating the linkage between biological sensors and the local environmental forcing at the basis of the TT2010 reconstruction.

All the following analyses are performed using winter-average (DJF) data and using 1950-1999 as the calibration period, unless specified otherwise. Furthermore, unless specified otherwise, the validation period is defined as the period spanning from the beginning of the simulation to the last year before the calibration period. The use of unsmoothed data limits the effect of high autocorrelation leading to spurious high skills metrics (Macias-Fauria et al., 2012).

3. Results

3.1 Simulated PNA during the observational period

First, we assess whether the employed models/simulations represent the observed/reanalyzed dominant large-scale circulation and associated surface-air temperature and precipitation patterns with sufficient accuracy for the observational period. This comparison guarantees that they are suitable for the subsequent analyses.

The four centers of the observed PNA wave-pattern are generally well captured by the simulations (Figure 1). A number of simulations, and most noticeably GISS-R24/-R25 (Figure 1e,f), display a weaker tropical center in the Pacific suggesting a weaker connection between tropics and extra-tropics. Higher model resolution does not systematically improve the overall quality of the hemispheric pattern. For instance, the PNA imprint over the Arctic as well as the Pacific-Atlantic teleconnection are too strong in the highly-resolved CCSM4 (Figure 1c). The PNA pattern of the lowest-resolution model (FGOALS-gl) has an overall weaker hemispheric imprint and the negative center over Florida is displaced westward over Mexico (Figure 1d), possibly reflecting an inadequate representation of the Rocky Mountains.

A similar behavior is found for the simulated spatial patterns of NPI (see supplementary Figure S1). Most noticeably, the NPI pattern in FGOALS-gl includes strong negative correlations over Central North America, again pointing to low-resolution topographic issues. All simulations show a good representation of the NAO pattern over the North Atlantic/Europe and China, but often overestimate its signature over the North Pacific (Figure S2). The simulation of the SOI pattern is a challenge for most of the models, especially concerning its signature over the extra-tropical North Pacific and North America (Figure S3).

Our pseudo-reconstruction approach also requires that simulations produce reliable imprints of the PNA - as well as of NPI, SOI and NAO - on North American winter surface-air temperature and precipitation (Figures 2 and 3). The observed correlation pattern between PNA and continental temperature is characterized by an approximately meridional stretch of positive correlations along the eastern coast of the Pacific Ocean, which extends eastward into continental regions at mid to polar latitudes, and by a center of negative correlation over the Sargasso Sea/Florida (Figure 2a). Simulations capture both features with varying quality (Figure 2b-i). For instance, the Sargasso Sea/Florida center is displaced in BCC-CSM1-1 and FGOALS-gl, while it is underrepresented in GISS-E2-R25, IPSL-CM5A-LR and slightly so in MIROC-ESM. Overall, FGOALS-gl features the worst representation of this correlation pattern possibly due to the deficiencies noticed above in the 500 hPa PNA pattern. Similar conclusions can be drawn about the NPI signature on winter North American temperatures (Figure S4). Simulations and reanalyses consistently point to a limited imprint of NAO and SOI on winter North American temperatures (Figures S5 and S6), which for both modes partly superposes with PNA signals.

Similar considerations could be derived for winter precipitation, but correlation patterns between large-scale circulation modes and precipitation over land are patchier than for temperature. Overall, the quality of simulated precipitation patterns compared to reanalyses is clearly poorer than for temperature. Both reanalyses and simulations indicate locally significant negative correlations between PNA and precipitation in the mid-latitude United States (*i.e.*, wetter conditions under negative PNA, and *vice versa*), but with substantial differences in the details of the pattern (Figure 3). An important robust feature is that all simulations except GISS-E2-R25 indicate weak negative correlations over the central Rocky Mountains (Figure 3b-i), a region where precipitation-sensitive proxies were screened for the TT2010 reconstruction.

In summary, the correlation patterns reveal a marked heterogeneity between simulations in the quality of their representation of dominant large-scale circulation modes and associated imprint on the North American climate. Of course, the spatial patterns are derived from the chosen 50-year period within single transient simulations, and are therefore not necessarily representative of the quality of the different models. Still, some general features are recognizable: A coarsely resolved North American topography and a poor representation of tropical and extra-tropical Pacific interactions are likely two major challenges limiting the quality of the simulated PNA imprints. The most apparent issues concern FGOALS-gl, potentially due to its coarser resolution compared to the other models.

3.2 Simulated PNA features during the last millennium

The evolution of the winter PNA index throughout the last millennium shares little resemblance between the different simulations (Figure 4a), which is indicative of a limited effect of the external forcing since the latter is very similar across the ensemble. Decadal and interdecadal phases of strong positive or, similarly, strong negative PNA appear at different periods in different simulations, suggesting that, in general, the PNA is mostly determined by internal variability at these time scales. No simulation displays a prolonged strong positive PNA phase during the early 19th century as featured by the TT2010 reconstruction, but decadal-scale positive PNA anomalies of similar relative amplitude emerge sporadically in the ensemble during different periods (see dots in Figure 4a). Such prolonged positive-PNA events are, however, rare. Unlike in the reconstruction, the early 19th century is characterized by predominant negative PNA trends in the simulations (the period of discrepancy is highlighted by a horizontal red bar in Figure 4b).

Running-window correlations between the PNA index and the other indices provide a simple assessment of the variable strength of PNA teleconnections in the different simulations. Reanalysis data indicate that winter PNA and NPI are practically indistinguishable: the two indices are robustly highly anti-correlated (thick black line in Figure 5a). Simulations consistently feature significant negative PNA-NPI correlations through the last millennium, although with considerable differences within the ensemble concerning strength and stationarity of the statistics (Figure 5a). CCSM4 produces the strongest and most robust correlations, which overlap with values from reanalyses, whereas FGOALS-gl produces the weakest and most time-varying correlations. The simulated winter PNA-NAO correlations are generally weak and negative during the last millennium, in agreement with the non-significant and strongly varying statistics from reanalysis data (Figure 5b). Some simulations feature multidecadal periods when the negative correlation becomes statistically significant, suggestive of a temporarily strong atmospheric connection between North Pacific and North Atlantic sectors. This is especially the case for CCSM4 (compare also superposing patterns in Figure 1c and Figure S2c). The negative PNA-NAO correlations represent periods when the atmospheric bridge linking Pacific and Atlantic climate variability is active (for a dynamical description see, *e.g.*, Raible et al. 2001; Pinto et al., 2010; Baxter and Nigam, 2013). Decadal active phases of such bridge in the form of persistent negative PNA/positive NAO pattern have been attributed to both, internal variability (Pinto et al., 2010) and strong volcanic forcing (Zanchettin et al., 2012). The winter PNA-SOI correlation is significantly negative in the reanalyses, though not very strong (Figure 5c). CCSM4 produces PNA-SOI correlations that remain robustly around this observed value throughout the last millennium, while the other simulations produce generally lower and more variable correlations (Figure 5c). In BCC-CSM1-1, MPI-ESM-P and MIROC correlations between SOI and both PNA and NPI (the latter not shown) are only sporadically significant, meaning that these models feature a weak connection between tropical and extra-tropical North Pacific. Note that these results do not qualitatively change if running-window correlations are calculated over longer periods.

Changes in the relative importance of large-scale modes for North American winter climate variability during the last millennium are assessed by comparing the fractional variances of North American surface-air temperature and precipitation that are explained by the different indices over sliding 30-year periods paced at one-decade intervals. The winter PNA is the dominant mode of simulated North American winter temperature variability among the considered indices (PNA, NPI, ENSO), generally explaining around 20% of the total variance (Figure 6). Only for FGOALS-gl there are several periods when the NPI becomes more dominant than the PNA. The strength of all index signatures changes through time. The fraction of North American winter precipitation variability explained by the indices is generally below 10%, and the dominance of PNA over the other indices is less clear than for temperature (Figure S7).

In summary, internal variability is an important factor for the simulated PNA during the pre-industrial millennium. The ensemble markedly disagrees with the TT2010 reconstruction, whose strong positive phase in the early 19th century was interpreted as resulting from a strong PNA response to solar forcing (TT2010). Correlations between indices reveal substantial differences in the simulated representation of teleconnections both within the ensemble and in comparison to reanalyses. Among the considered indices, PNA generally explains the largest fraction of North American winter temperature variability. Only FGOALS-gl features prolonged periods when PNA and NPI explain comparable fractions of North American winter temperature variability. We are therefore confident that through proper sampling of precipitation and especially temperature proxies over North America, pseudo-reconstructions are able to express robust PNA signals rather than signals from other indices.

3.3 PNA pseudo-reconstructions

First, we validate the reconstruction approach in TT2010. This is done in a perfect model framework by testing whether a reconstruction design based solely on geophysical predictors from northwestern North America can provide meaningful pseudo-reconstructions of the PNA. We use only predictor sets that provide a calibration skill comparable to that of the actual TT2010 reconstruction (Figure S8 summarizes the full-ensemble of pseudo-reconstruction calibration skills).

The so-obtained PNA pseudo-reconstructions are generally skillful according to both employed full-validation metrics (R^2_v and CE): only a few pseudo-reconstructions give CE values below 0, meaning they have no predictive skill and are hence unacceptable (Figure 7a). R^2_v values can exceed the imposed R^2_c range (see columns of numbers in Figure 7a). In some simulations, different performance between the validation and calibration periods can be related to the presence of significant local trends in North American winter temperatures during the latter period (Figure S9) that are mostly due to strong anthropogenic forcing imposed in the second half of the 20th century. For some simulations and predictor sets, the validation-period R^2 values (R^2_v) for the PNA overlap with those from the NPI (Figure S10), meaning that in these cases the pseudo-reconstructions are hardly effective in distinguishing PNA-related features from other signals.

The robustness of a reconstruction through time is a major concern for its reliability. Figure 8 shows that for the same set of predictors used for Figure 7a the R^2 skill of the pseudo-reconstructions changes remarkably through time: There are periods when the pseudo-reconstructions from one simulation are consistently poorer, other periods when they are consistently better, and there is generally a large spread in the quality of the pseudo-reconstructions. The multidecadal and centennial variations of R^2 suggest that there is structural uncertainty on these time scales. Note that, since metrics are calculated based on deviations from the 30-year average, the so-defined R^2 describes the reconstruction skills mostly on interannual to decadal variability. The CE metric, which accounts for the 30-year mean state, is characterized by non-stationarity and inter-model spread similar to R^2 (not shown). Running-window statistics further indicate that the models substantially differ concerning the comparative skills of reconstructed indices (top panel of Figure 8): in some models, and most noticeably in GISS-E2-R, the skills for the PNA are often not better than for NPI; in other models, like MPI-ESM-P and BCC-CSM1-1, the skills for the PNA are better than for all other indices. We conclude that the approach is effective in distinguishing the reconstructed PNA from other indices only for the latter models.

Another of our main objectives is to assess the robustness of the interdecadal strong positive PNA phase identified by the TT2010 reconstruction in the early 19th century. Figure 8 shows particularly large inter-model spread in the R^2 metric during the late 18th and early 19th centuries, with some simulations performing very well (IPSL-CM5A-LR, GISS-E2-R25 and MPI-ESM-P) while others show some of the poorest skills in the entire millennium (*e.g.*, BCC-CSM1-1 and GISS-E2-R25). It should be noted that the PNA pseudo-reconstructions are generally biased towards negative PNA values (black vertical lines in Figure 9). This is mainly due to the fact that most simulations have strong local trends in 20th century temperatures (Figure S9), which result in PNA pseudo-reconstructions roughly following a hockey-stick shape over the last millennium. As a consequence of this bias, our pseudo-reconstructions tend to underestimate interdecadal phases of very strong positive PNA that are detected throughout the simulations (see the negatively-centered histogram in Figure 9a). By contrast, interdecadal phases of strong positive PNA in the pseudo-reconstructions seem to describe the actual PNA conditions more accurately, as shown by the rather symmetric, almost zero-centered probability distribution of ensemble residuals for these events (Figure 9b). So, whereas actually simulated prolonged strong positive PNA phases may not be correctly captured by pseudo-reconstructions, prolonged strong positive phases emerging in the pseudo-reconstructions are generally 'true'.

There are further concerns about the capability of pseudo-reconstructions to accurately capture the PNA low-frequency variability. We find that pseudo-reconstructions tend to overestimate the amplitude of multidecadal-to-centennial fluctuations (Figure 10). In some simulations, like BCC-CSM1-1, CCSM4, FGOALS-gl and MPI-ESM-P, the spectra of the pseudo-reconstructions entail large-amplitude peaks in this frequency band that do not appear in the spectrum of the actual index. GISS-E2-R and MIROC, by contrast, produce pseudo-reconstructions whose spectra agree fairly well with that of the simulated PNA. The different agreement between pseudo-reconstruction-spectra and the actual PNA spectrum implies that the models disagree about whether i) the low-frequency temperature and precipitation excursions captured by the pseudo-proxies are related to the PNA and/or whether ii) the PNA reacts to the same low-frequency forcing as the temperature and precipitation pseudo-proxies do. Combining evidence from Figures 8, 9 and 10, we conclude that the errors in the pseudo-reconstructed prolonged positive PNA phases and, more generally, the variable skills of pseudo-reconstructions on multidecadal and centennial time scales reflect misrepresentation of low-frequency PNA variability by the pseudo-reconstructions. This casts doubt on the reliability of the early-19^{th}-century PNA event identified by the TT2010 reconstruction.

Considering PNA pseudo-reconstructions instead of the actually simulated PNA indices does not solve the discrepancy between simulations and the TT2010 reconstruction during the early 19^{th} century (Figure 4c). In fact, none of the simulations displays significant positive winter temperature anomalies over northwestern North America during the period 1800-1820 (Figure S11), which would be consistent with a positive phase in the PNA pseudo-reconstructions following the current definition. Instead, the anomalous patterns are characterized by a marked heterogeneity, suggesting lack of a robust response to external forcing across the models, with no pattern resembling the typical PNA structure. Accepting the reconstructed PNA behavior during the early 19th century as accurate and the simulated climates as realistic, the apparent discrepancy can only be solved by interpreting the first as a particular event of internal climate variability, hence unlikely captured (in its temporal occurrence) in a small-size ensemble as the one at hand. Indeed, a similar discrepancy is found in the late 1940s for a reconstructed decadal-scale negative PNA phase (Figure 4c), a period not characterized by prominent (inter)decadal forcing events.

3.4 Designing new PNA reconstructions

A natural question at this point is whether there is margin to improve reconstructions of the PNA. Potential for improvement may come, for instance, from the inclusion of new and/or better predictors over northwestern North America. Simulations disagree about whether the skills of the actual TT2010 PNA reconstruction and of its synthetic analogs represent the limit of this reconstruction approach: Some simulations (*e.g.*, BCC-CSM1-1, FGOALS-gl) indicate that the calibration skill of the TT2010 reconstructions is close to the expectation for the method (Figure S8). In other simulations, including CCSM4 and IPSL-CM5A-LR, geophysical predictor sets from northwestern North America tend to produce R^2_c values above the 0.45-0.55 range, while this range is in the upper tail of the R^2_c distribution in MPI-ESM-P (Figure S8). The subsets of predictors yielding the highest R^2_c values among the considered random sets delineate the upper bound of this reconstruction method by the inclusion of improved predictors. Accordingly, analogously to Figure 7a, Figure 7b summarizes the skill metrics for the subset of pseudo-reconstructions with R^2_c values in the upper quartile of the R^2_c distribution. BCC-CSM1-1, CCSM4 and less so GISS-R24 and IPSL-CM5A-LR indicate upper potential for the TT2010 approach, especially in terms of R^2_v (compare panels a and b in Figure 7). This is not the case for FGOALS-gl, GISS-E2-R25 and MPI-ESM-P, whose best skill scores indicate that pseudo-reconstructions from northwestern North American predictors are unlikely capable of explaining substantially larger variance than obtained in the actual TT2010 reconstruction. The simulations also disagree about whether an improved selection of predictors would lead to more distinguishable reconstruction skills between PNA and NPI (not shown).

The correlation patterns between the residuals of the PNA pseudo-reconstructions illustrated in Figures 7a and 8 and North American winter temperatures (Figure 11) indicate that, consistently among the simulations, an approach only using predictors from northwestern North America lacks important information from the southwestern and southeastern United States. Both regions correspond to characteristic regions for the PNA signature on North American winter temperatures (Figure 2). The residual correlation patterns and their robustness reflect structural deficiencies, and suggest possible changes in the reconstruction design to improve PNA reconstructions. Inclusion of temperature information from the southeastern United States would, for instance, reduce the risk of erroneously interpreting periods of spatially uniform continental warming/cooling or moistening/drying over North America as positive/negative PNA phases.

Accordingly, Figure 7c outlines the potential of the reconstruction method through an improved selection of proxy locations and extended calibration period. In this case, the predictor sets include temperature sampled from a box located over Florida instead of precipitation sampled from the southern box over the western United States, so that the model can capture information from the easternmost negative PNA center (Figures 1 and 2). The potential quality of the PNA pseudo-reconstructions obtained with this design is greatly improved according to both considered skill metrics (compare panels b and c in Figure 7). The quality of the pseudo-reconstructions still varies substantially through the last millennium (not shown), but the risk of periods of unskillful reconstructions (CE<0) is much lower than for a design limited to northwestern North America. The pseudo-reconstructions further display improvement in the negative bias and a substantially better representation of low-frequency PNA variability (Figure S12). However, the models disagree about which factor (*i.e.,* extended calibration period or inclusion of a temperature predictor for the southeastern United States) more strongly contribute to the improved results.

In summary, pseudo-proxy experiments appear to be instrumental in both the designing and the assessment of future PNA reconstructions. Of course, the exemplary design proposed here represents an ideal setting, and future applications of this tool for real-world reconstructions would require the pseudo-proxy experiments to be designed based on quality and type of actually available proxies.

4. Discussion

In order to understand the implications of our results for real-world proxy reconstructions and for the interpretation of last-millennium climate simulations, our discussion concentrates on three aspects: limitations of the reconstruction methods and of our pseudo-reconstruction design in particular; weaknesses in the simulated representation of the PNA, and of its teleconnections and variability; and issues related to (regional) climate attribution before the observational period and uncertainties affecting the simulation of the early-19th-century climate.

Our pseudo-proxy investigation reveals the inherent limitations of a PNA reconstruction method solely relying on local geophysical predictors from northwestern North America. Assumptions of linearity and stationarity between local hydro-climate variability and the large-scale atmospheric circulation described by the PNA are further weaknesses of the approach. In our linear definition, the PNA robustly dominates winter North American climate variability (Figures 6), which is an encouraging premise for reconstruction attempts. Nonetheless, the so-defined PNA index may not capture shifts in the location of the mode's centers of action and in the associated teleconnections (Raible et al., 2014). Furthermore, our pseudo-reconstructions can be affected by the non-stationarity of the teleconnection pattern to North America of other modes of climate variability, like ENSO (Coats et al., 2013). Pseudo-reconstructions suggest that margins exist to substantially improve the quality of the reconstructed PNA based on a TT2010-like multi-linear regression method, for instance if the multiple PNA-sensitive regions over North America are more exhaustively represented in the predictors' set and if the calibration period is extended. However, including temperature-sensitive predictors from the southeastern North America and extending the calibration period to the full 20th century, as in our pseudo-proxy experiment (Figure 7c), may be difficult due to the nature of real-world climate proxies and limitation of observational data suitable for model calibration. First, as noted above, the ensemble teleconnection patterns in Twentieth-Century-

Reanalysis data – the longest reanalysis product now available – are poorly constrained during the first half of the 20th century over the Pacific (Raible et al., 2014). Then, the Northern Hemisphere's ring-width network shows that the proxy-responses to atmospheric modes like the PNA are determined by a complex causal chain linking large-scale circulation, local climate and seasonal tree growth (St. George, 2014). Accordingly, relatively few chronologies, mostly from the Pacific Northwest and northern Rockies, significantly respond to the winter PNA (St. George, 2014). More generally, real climate-proxies can be critically affected by noise (von Storch et al., 2009), and may suffer from non-stationary climate-proxy relationships that are neglected in our perfect-model framework (*e.g.*, Evans et al., 2013; D'Arrigo et al 2008). There exist, however, long winter precipitation-sensitive, and possibly also temperature-sensitive, proxies across the southern United States (*e.g.*, Stahle et al., 1994; St. George, 2014; St. George and Ault, 2014), upon which future designs of PNA pseudo-reconstruction exercises could be based. As shown in supplementary Figure S13, the skills of an ensemble of TT2010-like PNA pseudo-reconstructions progressively deteriorate for increasing levels of noise artificially introduced in the predictors. Skills depend more on the level of noise rather than on the type of noise, at least for low amounts of noise, in accordance with von Storch et al. (2009). For a signal-to-noise ratio of 1 (Figure S13c,f) explained variances for the validation period never reach 0.5, and red noise generally produces unskillful reconstructions. So, our pseudo-proxy investigation is only meant as an idealized example demonstrating the potential margins of improvement offered by the reconstruction method. Its application to a real-world PNA reconstruction requires scrutiny of available data, which we defer to a follow-up dedicated study.

Poor modeling of the PNA-related dynamics is a straightforward explanation of the early-19th century discrepancy between the reconstruction and the simulations. Furthermore, the realism of our PNA pseudo-reconstructions relies on the realism of simulated patterns, variability and teleconnections of the PNA as well as of other hemispheric modes imprinting on the North American climate. Accurate representation of observed dominant modes of climate variability and of their teleconnections still represents a challenge for coupled climate simulations (*e.g.*, about ENSO see: Guilyardi et al., 2012;

Zou et al., 2014). Unrealistic simulated representation of large-scale atmospheric circulation modes can arise due to biased ocean-atmosphere coupling over remote regions: Coupled climate models are still affected by considerable biases in regional SSTs and sea ice - especially in the North Atlantic Ocean - that are associated, in the Northern Hemisphere, to cold biases resembling the Northern Hemisphere's annular mode (Wang et al., 2014). This suggests that good model performance in simulating regional processes may be overridden by the effect of remote biases.

Our definition of the PNA index does not account for possible displacements of its centers of actions in simulated patterns compared to reanalyses. An alternative definition based on empirical orthogonal functions (EOF) results in PNA indices that share between half (MIROC-ESM) and almost the whole (CCSM4) total variance with the pointwise-based PNA indices over the observational period (see supplementary Table S1). Spatial differences between simulated EOF-based and pointwise-based patterns also vary considerably across the ensemble (Table S1). It is not yet clear whether and how these uncertainties related to the index definition affect the details of the pseudo-reconstructions. The validity of our general conclusions clearly stands for the sub-ensemble including only models with the most consistent PNA indices across the two definitions (CCSM4, IPSL-CM5A-LR, MPI-ESM-P).

The marked inter-model differences in the PNA-precipitation correlation patterns over North America and their general disagreement with the observed pattern (Figure 3) highlight the large uncertainties in the connection between large-scale circulation and local hydro-climates that still affect state-of-the-art coupled climate simulations. In this regard, topography largely determines the wave-like structure of the PNA and its surface signature. Its dominant role was already highlighted by TT2010 in describing the characteristics of their two precipitation-sensitive tree-ring series from Montana and Wyoming. Poor model topography likely leads to biases in representing the PNA pattern and more visibly its climate fingerprints. This was exemplified here by the stark contrast between the low-resolution model FGOALS-gl and the high-resolution model CCSM4 (compare panels c and d in Figures 1-3). With few exceptions, topography in the employed models misses critical plateau elevations that are crucial for the onset and sustenance of snow/ice-related feedbacks (Berdahl and Robock, 2013). These could be relevant for the reinforcement/dampening of the Canadian High during the development phase of a positive/negative PNA (Ge and Gong, 2009). Accordingly, these latter model deficiencies can partly explain why inclusion of precipitation over the Rockies as a predictor degrades the skills of our pseudo-reconstructions and yields much weaker skills than the actual TT2010 reconstruction (as discussed in section 2.4).

A possible solution to the discrepancy between the reconstruction and the simulations is to attribute the reconstructed early-19th-century positive PNA phase to internal variability. Supporting this hypothesis, interdecadal persistent positive PNA phases emerge in all simulations throughout the last millennium without consistent timing (Figure 4a). However, simulated events are generally weaker than the reconstructed event and the number of those longer than 20 years is exiguous (see supplementary Figure S14). Therefore, notwithstanding the caveats described above about the realism of simulated PNA dynamics and variability, the reconstructed early-19th-century positive PNA phase is compatible with an exceptional event of internal climate variability. A similar interpretation has been recently proposed, with the support of climate simulations, for reconstructed multidecadal droughts in the southwestern North America during the last millennium (Coats et al., 2015). Further supporting this hypothesis, the simulations ensemble does not point to coherent positive PNA anomalies during other periods of the last millennium with concomitant strong volcanic forcing and weak solar forcing, e.g., the mid 15th and the late 17th centuries (Figure 4a).

Under the alternative hypothesis that the reconstructed early-19th-century positive PNA phase is externally driven, the discrepancy between the reconstruction and the simulations can be explained by common model deficiencies in the simulated dynamical response to natural forcing and/or by uncertainty in the (reconstructed) imposed external forcing. Supporting this hypothesis, state-of-the-art coupled climate models still suffer from a deficient representation of stratospheric and coupled stratosphere-troposphere dynamics (Kodera et al., 1996; Woollings et al., 2010), which affect the simulated response to volcanic (Driscoll et al., 2012; Charlton-Perez et al., 2013; Muthers et al. 2014) and solar (Gray et al., 2010; Anet et al., 2014) forcing. Furthermore, inter-model disagreement about post-eruption oceanic evolutions (*e.g.*, Ding et al., 2014) shows that large uncertainties still exist about decadal-scale climate variability during periods of strong volcanic forcing and the role of the ocean in determining the surface-air temperature response (Canty et al., 2013). Sensitivity simulations performed with a chemistry–climate model demonstrate the importance of the Dalton Minimum of solar activity for the persistence of the hemispheric cold temperature anomalies of the early 19th century (Anet et al., 2014). Yet, the cold winter temperature anomalies depicted by these simulations over Alaska during the period 1805-1825 do not match with the imprint of a positive PNA. Single-model ensemble climate simulations have shown that the 1815 Tambora eruption produces robust large-scale atmospheric circulation anomalies, roughly corresponding to a positive PNA phase, only in the absence of additional external disturbances, whereas under full-forcing conditions such positive PNA-like features become hardly distinguishable (Zanchettin et al., 2013a). The same simulation-ensembles have further demonstrated that internal climate variability can be a source of uncertainty for the simulated early-19th-century decadal climate evolution as important as the (reconstructed) imposed forcing (Zanchettin et al., 2013a). Moreover, although climate simulations depict an interannual to decadal PNA/NPI response to strong tropical volcanic eruptions (Zanchettin et al., 2012; Wang et al., 2012), responses on longer time scales may be damped by the resilience of the interdecadal component of the Pacific Decadal Oscillation to natural external forcing (Zanchettin et al., 2013b).

Uncertainty in the external forcing factors acting on the early-19th-century climate further complicates the attribution of reconstructed and simulated variability. For instance, reconstructed variations in total solar irradiance are affected by considerable uncertainties (*e.g.*, Schmidt et al., 2011; Shapiro et al., 2011) as well as deficiencies in accounting for the spectrum variations for solar forcing and ozone response (Gray et al. 2010). Debate is ongoing about how changes in total solar irradiance affect the tropical oceans, with different observations and different simulations disagreeing about whether warming rather than cooling of the upper tropical Pacific is expected under enhanced solar activity (Misios and Schmidt, 2012). Moreover, the radiative impact of tropical volcanic eruptions is sensitive to the season of the eruption (Toohey et al., 2011; Froelicher et al., 2013), and the season of the 1809 tropical eruption is still insufficiently constrained (Cole-Dai et al., 2009).

5. Conclusions

Our results depict a discrepancy between reconstructed and simulated PNA behavior during the early 19th century, an exceptionally cold period in the Northern Hemisphere characterized by concomitant weak solar and strong volcanic forcing. According to our pseudo-proxy investigation, reconstructions based on northwestern North American geophysical predictors are potentially skillful in terms of two different metrics (coefficient of determination and coefficient of error). Such an approach following Trouet and Taylor (2010) is also likely capable of capturing strong interdecadal positive PNA phases, like the one reconstructed for the early 19th century. However, a number of sources of uncertainty and potential deficiencies are still present especially at multidecadal and centennial timescales. Furthermore, pseudo-reconstructions based solely on predictors from northwestern North America often cannot distinguish between the PNA and the North Pacific Index describing the strength of the Aleutian Low.

The PMIP3-*past1000* and *historical* simulations provide an overall satisfactory representation of the observed PNA spatial pattern and of its imprint on the North American climate. Simulated pre-industrial PNA evolutions show a predominance of internal variability over forced signals, which could be used as an argument to explain why simulations do not robustly exhibit the reconstructed positive PNA phase in the early 19th century. Shifting focus to attribution of the reconstructed anomaly requires confidence that simulations do not suffer from common deficiencies in the response to natural forcing, in the applied reconstructed forcing and/or in the internally-generated climate variability. We need therefore to better understand the relative role of externally-forced and internal climate variability during the pre-industrial period.

A refined topography associated with high horizontal model resolution appears to be essential for models to realistically capture the connection between the large-scale circulation and the local climatic/environmental conditions upon which a reliable PNA reconstruction depends. However, our pseudo-reconstructions also indicate that there is margin to substantially improve the available PNA reconstruction, in particular through a more exhaustive representation of the multiple PNA-sensitive regions over North America in the predictors' set. These results call for strengthened cooperation between the climate-proxy and climate modeling communities in order to improve our knowledge about the early-19th-century PNA and to solve the related reconstruction-simulations discrepancy.

Acknowledgments

This study stems from discussion during the PAGES2k/PMIP "Workshop on integrated analyses of reconstructions and multi-model simulations for the past two millennia" (http://www.palma-ucm.es/members/workshop/Home.html) in November 2013, funded by PAGES, FECYT-MINECO and IGEO. Comments from two anonymous reviewers helped improve the clarity of the manuscript. Thomas Kleinen provided useful comments on an earlier version of the draft. D. Zanchettin was supported by the BMBF (research program "MiKlip", FKZ:01LP1158A). F. Lehner and C. C. Raible are supported by the Swiss National Science Foundation. This work contributes to the MPI-M Integrated Project Millennium and CliSAP. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate-modelling groups for producing and making available their model output.

References

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

Anet, J. G., S. Muthers, E. V. Rozanov, C. C. Raible, A. Stenke, A. I. Shapiro, S. Brönnimann, F. Arfeuille, Y. Brugnara, J. Beer, F. Steinhilber, W. Schmutz, and T. Peter (2014) 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.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Mon. Wea. Rev., **115**, 1083–1126.

Baxter, S., and S. Nigam (2013) A subseasonal teleconnection analysis: PNA development and its relationship to the NAO. *J. Climate*, **26**, 6733-674.

Berdahl, M., and A. Robock (2013), Northern Hemispheric cryosphere response to volcanic eruptions in the Paleoclimate Modeling Intercomparison Project 3 last millennium simulations, J. Geophys. Res. Atmos., **118**, 12,359–12,370, doi:10.1002/2013JD019914.

Bothe, O., Jungclaus, J. H., and Zanchettin, D. (2013) Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble. Clim. Past, **9**, 2471-2487, doi:10.5194/cp-9-2471-2013.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao (2012) Evaluation of climate models using palaeoclimatic data, Nature Clim. Ch., **2**, 417-424, doi:10.1038/nclimate1456.

Canty, T., N. R. Mascioli, M. D. Smarte, and R. J. Salawitch (2013) An empirical model of global climate – Part 1: A critical evaluation of volcanic cooling. Atm. Chem. Phys., **13**, 3997–4031, doi:10.5194/acp-13-3997-2013.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., Davis, N. A., Gerber, E. P., Gillett, N., Hardiman, S., Kim, J., Krüger, K., Lee, Y.-Y., Manzini, E., Mc-Daniel, B. A., Polvani, L., Reichler, T., Shaw, T. A., Sigmond, M., Son, S.-W., Toohey, M., Wilcox, L., Yoden, S., Christiansen, B., Lott, F., Shindell, D., Yukimoto, S., and Watanabe, S.: On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models, J. Geophys. Res., 118, 2494–2505, doi:10.1002/jgrd.50125, 2013.

Coats, S., J. E. Smerdon, B. I. Cook, and R. Seager (2013) Stationarity of the tropical pacific teleconnection to North America in CMIP5/PMIP3 model simulations. Geophys. Res. Lett., 40(18), 4927–4932, doi:10.1002/grl.50938

Coats, S., J. E. Smerdon, B. I. Cook, and R. Seager (2015) Are Simulated Megadroughts in the North American Southwest Forced?. J. Climate, 28, 124–142. doi:http://dx.doi.org/10.1175/JCLI-D-14-00071.1

Cole-Dai, J., D. Ferris, A. Lanciki, J. Savarino, M. Baroni, and M. H. Thiemens (2009), Cold decade (AD 1810–1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption, Geophys. Res. Lett., **36**, L22703, doi:10.1029/2009GL040882.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason Jr., B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Broennimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, O., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J. (2011) The Twentieth Century Reanalysis Project, Q. J. Roy. Meteorol. Soc., **137**, 1–28.

Cook, E. R., K. R. Briffa, and P. D. Jones (1994) Spatial regression methods in dendroclimatology – a review and comparison of 2 techniques. Int. J. Climatol., **14**, 379–402.

Crowley, T. J. (2000) Causes of climate change over the past 1000 years, Science, **289**, 270–277.

Crowley, T. J., Zielinski, G., Vinther, B., Udisti, R., Kreutz, K., Cole-Dai, J., and Castellano, E. (2008) Volcanism and the little ice age, Pages News, **16**, 22–23.

D'Arrigo, R., R. Wilson, B. Liepert, and P. Cherubini (2008) On the 'Divergence Problem' in Northern Forests: A review of the tree-ring evidence and possible causes. Glob. Plan. Ch., **60**, 289-305

Ding, Y., J. A. Carton, G. A. Chepurin, G. Stenchikov, A. Robock, L. T. Sentman, and J. P. Krasting (2014) Ocean response to volcanic eruptions in Coupled Model Intercomparison Project 5 (CMIP5) simulations. J. Geophys. Res., in press, doi:10.1002/2013JC009780.

Driscoll, S., A. Bozzo, L. J. Gray, A. Robock, and G. Stenchikov (2012) Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, J. Geophys. Res., **117**, D17105, doi:10.1029/2012JD017607.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N. , Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., and Viovy, N. (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Clim. Dyn., **40**, 2123-2165.

Evans, M.N., S.E. Tolwinski-Ward, D.M. Thompson, and K.J. Anchukaitis (2013) Applications of proxy system modeling in high resolution paleoclimatology. Quat. Sc. Rev., **76**, 16-28, doi:10.1016/j.quascirev.2013.05.024.

Froelicher, T. L., F. Joos, C. C. Raible, J. L. Sarmiento (2013) Atmospheric CO2 response to volcanic eruptions: the role of ENSO, season, and variability. Global Biogeochemical Cycles, **27**, 239-251.

Gao, C., A. Robock, and C. Ammann (2008) Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models, J. Geophys. Res.-Atmos., **113**, D23111, doi:10.1029/2008JD010239.

Ge, Y., and G. Gong (2009) North American snow depth and climate teleconnection patterns. J. Clim., **22**, 217-233, doi:10.1175/2008JCLI2124.1.

Gong, D.-Y., R. Mao, P.-J. Shi, Y.-D. Fan (2007) Correlation between east Asian dust storm frequency and PNA. Gephys. Res. Lett., **34**, L14710, doi:10.1029/2007GL029944.

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G. A., Shindell, D., van Geel, B., and White, W. (2010) Solar influence on climate, Rev. Geophys., **48**, RG4001, doi:10.1029/2009RG000282.

Guilyardi, E., H. Bellenger, M. Collins, S. Ferrett, W. Cai, and A. Wittenberg (2012) First look at ENSO in CMIP5. CLIVAR Exchanges No. 58, Vol. 17, No.1.

Horel, J.D., and Wallace J.M. (1981) Planetary-scale atmospheric phenomena associated with the southern oscillation. Mon. Weath. Rev., **109**, 813–829.

Hubeny, J. B., J. W. King, M. Reddin (2010) Northeast US precipitation variability and North American climate teleconnections interpreted from late Holocene varved sediments. Proc. Nat. Ac. Sc., **108**, 17895–17900, doi:10.1073/pnas.1019301108.

Jones, P., and Mann, M. (2004) Climate over past millennia, Rev. Geophys., **42**, RG2002, doi:10.1029/2003RG000143.

Jungclaus, J. H., K. Lohmann, and D. Zanchettin (2014) Enhanced 20th century heat transfer to the Arctic simulated in the context of climate variations over the last millennium. Clim. Past Discuss. 10, 2895-2924, doi:10.5194/cpd-10-2895-2014

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, Roy Jenne, and D. Joseph (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull. Amer. Meteor. Soc., **77**, 437–471.

Kistler, R., W. Collins, S. Saha, G. White, J. Woollen, E. Kalnay, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino, (2001) The NCEP–NCAR 50–Year Reanalysis: Monthly means CD–ROM and Documentation. Bull. Amer. Meteor. Soc., **82**, 247–267.

Kodera, K., Chiba, M., Koide, H., Kitoh, A., and Nikaidou, Y. (1996) Interannual variability of the winter stratosphere and troposphere in the Northern Hemisphere, J. Meteorol. Soc. Jpn., **74**, 365–382.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., and Teng, H. (2011) Last millennium climate and its variability in CCSM4, J. Climate, **26**, 1085–1111, doi:10.1175/jcli-d-11-00326.1.

Lehner, F., C. C. Raible, and T. F. Stocker (2012) Testing the robustness of a precipitation proxy-based North Atlantic Oscillation reconstruction. Quat. Sc. Rev. **45**, 85-94, doi:10.1016/j.quascirev.2012.04.025

Li, J., Xie, S.-P., Cook, E. R., Morales, M. S., Christie, D. A., Johnson, N. C., Chen, F., D'Arrigo, R., Fowler, A. M., Gou, X., and Fang, K. (2013) El Niño modulations over the past seven centuries, Nature Cl. Ch., 3, 822–826, doi:10.1038/NCLIMATE1936

Macias-Fauria,M., A. Grinsted, S. Helama, and J. Holopainen (2012) Persistence matters: Estimation of the statistical significance of paleoclimatic reconstruction statistics from autocorrelated time series. Dendrochronol., **30**, 179-187.

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F. (2009) Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, Science, **326**, 1256–1260, doi:10.1126/science.1177303.

Misios, S., H. Schmidt (2012) Mechanisms Involved in the Amplification of the 11-yr solar cycle signal in the tropical Pacific ocean. *J. Climate*, **25**, 5102–5118.

Moore, G. W. K., I. A. Renfrew, and R. S. Pickart (2013) Multidecadal mobility of the North Atlantic Oscillation, J. Climate, **26**, 2453–2466.

Muthers, S., J. G. Anet, E. Rozanov, C. C. Raible, T. Peter, A. Stenke, A. Shapiro, J. Beer, F. Steinhilber, S. Broennimann, F. Arfeuille, Y. Brugnara, and W. Schmutz (2014) Sensitivity of the winter warming pattern following tropical volcanic eruptions to the background ozone climatology, J. Geophys. Res., **119**, 1340-1355. DOI:10.1002/2013JD020138.

Notaro, M., W.-C. Wang, W. Gong (2006) Model and Observational Analysis of the Northeast U.S. Regional Climate and Its Relationship to the PNA and NAO Patterns during Early Winter. *Monthly Weather Review* **134**, 3479-3505.

Pinto, J. G., Reyers, M., Ulbrich, U. (2011) The variable link between PNA and NAO in observations and in multi-century CGCM simulations. Clim. Dyn. **36**, 337-354, doi:10.1007/s00382-010-0770-x.

Raible, C. C., U. Luksch, K. Fraedrich, and R. Voss (2001) North Atlantic decadal regimes in a coupled GCM simulation, Clim. Dyn., **18**, 321-330.

Raible, C. C., Casty, C., Luterbacher, J., Pauling, A., Esper, J., Frank, D. C., Büntgens, U., Roesch, A. C., Wild, M., Tschuck, P., Vidale, P.-L., Schär, C., and Wanner, H. (2006) Climate variability – observations, reconstructions and model simulations, Climatic Change, **79**, 9–29.

Raible, C. C., F. Lehner, J. F. González-Rouco, and L. Fernández-Donado (2014) Changing correlation structures of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD. Clim. Past, **10**, 537-550, doi:10.5194/cp-10-537-2014.

Shapiro, A. I., W. Schmutz, E. Rozanov, M. Schoell, M. Haberreiter, A. V. Shapiro and S. Nyeki (2011) A new approach to the long-term reconstruction of the solar irradiance leads to large historical solar forcing. Astr. & Astrophys., **529**, A67, doi: http://dx.doi.org/10.1051/0004-6361/201016173.

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 paleo-climate comparisons to constrain future projections in CMIP5, Clim. Past Discuss., 9, 775–835, doi:10.5194/cpd-9-775-2013, 2013.

Smerdon, J. E. (2012) Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. WIREs Clim Change, **3**, 63–77, doi: 10.1002/wcc.149.

Stahle, D. W., and M. K. Cleaveland (1994) Tree-ring reconstructed rainfall over the southeastern USA during the medieval warm period and the little ice age. Clim. Change, **26**, 199-212.

St. George, S. (2014) An overview of tree-ring width records across the Northern Hemisphere. Quat. Sc. Rev., **95**, 132-150, doi:10.1016/j.quascirev.2014.04.029.

St. George, S., and T. R. Ault (2014) The imprint of climate within Northern Hemisphere trees. Quat. Sc. Rev., **89**, 1-4, doi:10.1016/j.quascirev.2014.01.007.

Starheim, C.C.A., Smith, D.J., Prowse, T.D., 2013, Dendrohydroclimate reconstructions of July-August runoff for two nival-regime rivers in west central British Columbia. Hydrol. Proc., **27**, 405-420.

Stephenson, D. B., V. Pavan, M. Collins, M. M. Junge, and R. Quadrelli (2006) North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment. Clim. Dyn., **27**, 401–420, doi:10.1007/s00382-006-0140-x.

Toohey, M., Krüger, K., Niemeier, U. and Timmreck, C. (2011) The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions Atmosph. Chem. Phys., **11**, 12351-12367, doi:10.5194/acp-11-12351-2011.

Trenberth, K. E., and J. W. Hurrell (1994) Decadal atmosphere-ocean variations in the Pacific. Clim. Dyn., **9**, 303-319.

Trouet, V., and A. H. Taylor (2010) Multi-century variability in the Pacific North American circulation pattern reconstructed from tree rings. Clim. Dyn., **35**, 953–963, doi:10.1007/s00382-009-0605-9.

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

von Storch, H., E. Zorita, F. Gonza´lez-Rouco (2009) Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation. Int. J. Earth. Sci. (Geol Rundsch) 98:67–82, doi:10.1007/s00531-008-0349-5

Wahl, E. R., and J. E. Smerdon (2012) Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors, Geophys. Res. Lett., **39**, L06703, doi:10.1029/2012GL051086.

Wahl, E. R.,Diaz, H. F., Smerdon, J. E., and C. M. Ammann (2014) Late winter temperature response to large tropical volcanic eruptions in temperate Western North America: Relationship to ENSO phases, Glob. Plan. Ch., online first, doi:10.1016/j.gloplacha.2014.08.005.

Wallace, J. M., and D. S. Gutzler (1981) Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. Mon. Wea. Rev., **109**, 784–812, doi:http://dx.doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso (2014) A global perspective on CMIP5 climate model biases. Nature Clim. Ch., **4**, 201–205, doi:10.1038/nclimate2118.

Wang, T., O. H. Otterå, Y. Gao, and H. Wang (2012), The response of the North Pacific Decadal Variability to strong tropical volcanic eruptions. Clim. Dyn., **39**, 2917-2936, doi:10.1007/s00382-012-1373-5.

Woollings, T., Charlton-Perez, A., Ineson, S., Woollings, T., Charlton-Perez, A., Ineson, S., Marshall, A. G., and Masato, G. (2010) Associations between stratospheric variability and tropospheric blocking, J. Geophys. Res., **115**, D06108, doi:10.1029/2009JD012742.

Younas, W., and Y. Tang. (2013) PNA predictability at various time scales. *J. Climate* **26**, 9090-9114.

Yu, B., F. W. Zwiers. (2007) The impact of combined ENSO and PDO on the PNA climate: A 1,000-year climate modeling study. *Clim. Dyn.,* **29**, 837-851.

Yu, B., A. Shabbar, F. W. Zwiers. (2007) The Enhanced PNA-Like climate response to Pacific interannual and decadal variability. *J. Climate* **20**, 5285-5300.

Zanchettin, D., C. Timmreck, H.-F. Graf, A. Rubino, S. Lorenz, K. Lohmann, K. Krueger, and J. H. Jungclaus (2012) Bi-decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic eruptions. Clim. Dyn., **39**, 419-444, doi:10.1007/s00382-011-1167-1.

Zanchettin, D., O. Bothe, H. F. Graf, S. J. Lorenz, J. Luterbacher, C. Timmreck and J. H. Jungclaus (2013a) Background conditions influence the decadal climate response to strong volcanic eruptions, J. Geophys. Res. Atmos., **118**, doi:10.1002/jgrd.50229.

Zanchettin, D., Rubino, A., Matei, D., and J.H. Jungclaus. (2013b) Multidecadal-to-centennial SST variability in the MPI-ESM simulation ensemble for the last millennium. Clim. Dyn., **40**, 1301-1318, doi:10.1007/s00382-012-1361-9

Zanchettin, D., O. Bothe, C. Timmreck, J. Bader, A. Beitsch, H.-F. Graf, D. Notz and J. H. Jungclaus (2014) Inter-hemispheric asymmetry in the sea-ice response to volcanic forcing simulated by MPI-ESM (COSMOS-Mill). Earth Syst. Dynam., 5, 223–242, doi:10.5194/esd-5-223-2014

Zhou, T., Li, B., Man, W., Zhang, L., and Zhang, J. (2011) A comparison of the Medieval Warm Period, Little Ice Age and 20th century warming simulated by the FGOALS climate system model, Chinese Sci. Bull., **56**, 3028–3041, doi:10.1007/s11434-011-4641-6.

Zou,Y., J.-Y. Yu, T. Lee, M.-M. Lu, and S. T. Kim (2014), CMIP5 model simulations of the impacts of the two types of El Niño on the U.S. winter temperature, J. Geophys. Res. Atmos., **119**, doi:10.1002/2013JD021064.

TABLES

Model/simulation	**Components and resolutions**	**Natural forcing**	**Time intervals**	**References**
BCC-CSM1-1 (r1i1p1)		S: Vieira, V: Gao	P (850-1850) H (1851-2012)	-
CCSM4 (r1i1p1)	CAM4 (1.25°x0.9°L26)/Parallel Ocean Model 2 (1°L60)	S: Vieira, V: Gao	P (850-1850) H (1851-2005)	Landrum et al. (2011)
FGOALS-gl (r1i1p1)		S/V: Crowley (2000), Jones and Mann	P (1000-1999)	Zhou et al. (2011)
GISS-E2-R (r1i1p124)	ModelE(2°x 2.5°L40)/Russell (1°x 1.25°L32)	S: Vieira, V: Crowley (2008)	P (850-1850) H (1851-2005)	-
GISS-E2-R (r1i1p125)	ModelE(2°x 2.5°L40)/Russell (1°x 1.25°L32)	S: Vieira, V: Gao	P (850-1850) H (1851-2005)	http://data.giss.nasa.gov/modelE/ar5/
IPSL-CM5A-LR (r1i1p1)	LMDZ5A (1.875°x3.75° L39)/NEMO (2° , with refinement at the equator of 0.5°, L31)	S: Vieira, V: Ammann et al. (2007)	P (850-1850) H (1851-2005)	Dufresne et al. (2013)
MIROC-ESM (r1i1p1)	-	-	P (850-1850) H (1851-2005)	-
MPI-ESM-P (r1i1p1)	ECHAM6 (T63L47)/MPIOM(GR15 L40)	S: Vieira, V: Crowley (2008)	P (850-1850) H (1851-2005)	Jungclaus et al. (2014)

Table 1 – Simulations considered in this study. Columns, from left: model and, in brackets, simulation; atmospheric and oceanic components (with resolution in brackets); applied external forcings for solar (S) and volcanic (V); considered periods of the past1000 (P) and historical (H) integrations; references/sources of information. Names of models and simulations follow the acronyms adopted in the CMIP5 repository. Full

references for the applied solar and volcanic forcing are: Vieira et al. (2011), Gao et al. (2008), Crowley (2000), Jones and Mann (2004) and Crowley et al. (2008).

Figure captions

Figure 1 – Observed and simulated correlation maps between the winter PNA index and winter Z500 time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the calculation of the PNA index. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the domain north of 20º N (to this purpose NCAR data were regridded to the model grid).

Figure 2 – Observed and simulated correlation maps between the winter PNA index and winter surface-air-temperature time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the shown land-only domain north of 12º N (to this purpose NCAR data were regridded to the model grid).

Figure 3 – Observed and simulated correlation maps between the winter PNA index and winter precipitation time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the shown land-only domain north of 12º N (to this purpose NCAR data were regridded to the model grid).

Figure 4 – Simulated, reconstructed and pseudo-reconstructed evolutions of the winter (djf) PNA index. a) smoothed PNA time series from the simulations for the whole last millennium; b) comparison between smoothed and, then, normalized (over the period 1725-1999) PNA time series from the simulations and the TT2010 reconstruction; c) comparison between normalized (over the period 1950-1999) pseudo-reconstructions (shown as agreement between pseudo-reconstructions from all the simulations) and the TT2010 reconstruction. Smoothing in panels a and b was performed through an 11-year running moving average. Dots in panel a individuate, for each simulation, occurrences of prolonged periods of positive PNA (defined as periods where the normalized smoothed index is above 1 for at least 15 consecutive years). The red bar in panels b and c highlights the approximate period of the simulations-reconstruction discrepancy.

Figure 5 – Running-window (31-year) correlations of the winter (djf) PNA with winter NPI (a), NAO (b) and SOI (c) indices for simulations (colored lines) and reanalysis (black thick lines) data. Dots mark when correlation is statistically significant at 95% confidence accounting for autocorrelation.

Figure 6 – Fractions of total variance of winter North American surface air temperatures (land only grid points within the domain 20-70ºN, 190-300ºE) explained by winter PNA, NPI, NAO and SOI indices for individual models. Values are calculated over decadally-paced 30-year periods.

Figure 7 – Skill metrics (coefficient of determination (R^2v) and coefficient of error (CE)) of the ensemble PNA pseudo-reconstructions for the full-validation period. Different panels illustrate results from different reconstruction designs, summarized on the title of each panel: a) reconstructions based on geophysical predictors from northwestern North America, with R^2c comparable to that of the actual TT2010 reconstruction (see methods); b) same as panel a, but for best R^2c values; c) best R^2c values from an idealized design including a temperature predictor over Florida. The numbers inside each panel indicate the minimum and maximum R^2v values obtained for each model. Insets in each panel map the three boxes from where gridded data are sampled to be included as predictors, with the name reported in each box (tas: surface air temperature, pr: precipitation).

Figure 8 – Skill metric (R^2) for an ensemble of PNA reconstructions based on geophysical predictors from northwestern North America for subsequent 30-year periods (paced at 3-decade intervals). To be comparable with TT2010, only the subset of reconstructions with R^2 for the 1950-1999 calibration period in the range [0.45-0.55] are shown, as for Figure 7a. For each 30-year period, dots are minimum, mean and maximum of R^2 values, vertical lines indicate the inter-quartile interval of R^2 values. The top symbols indicate 30-year periods when the R^2 value for NPI (square), SOI (triangle) or NAO (circle) is, for at least one predictor set, better than the worse PNA value.

Figure 9 – Pseudo-reconstructions' accuracy in describing interdecadal positive PNA phases. Histograms are ensemble (all simulations) empirical probability distributions of residuals (predicted value minus true value) from the winter PNA pseudo-reconstructions obtained following an approach similar to TT2010 and illustrated in Figures 7a and 8 for (a) target 21-year smoothed PNA values above the 90^{th} percentile and (b) pseudo-reconstructed 21-year smoothed PNA values above the 90^{th} percentile. The black vertical lines indicate the full-period average residuals from individual simulations. 90^{th} percentiles are calculated over the full simulation and therefore reflect also full-period biases in the pseudo-reconstructions. The smoothing is meant to mimic the approximately 20-year duration of the early-19^{th}-century positive PNA phase in the TT2010 reconstruction. The considered positive PNA phases are sampled throughout the simulations, regardless of their timing.

Figure 10 – Power spectral density of the winter PNA index (blue line) for individual simulations with associated 95% confidence level (blue dashed line) and agreement between the spectra of the pseudo-reconstructions (shading) obtained following an

approach similar to TT2010 and illustrated in Figures 7a and 8. Agreement is defined, for a given frequency, as the fraction of total pseudo-reconstructions having power within 0.1 units^2/year intervals. All indices are standardized according to the 1950-1999 climatology.

Figure 11 – Correlation maps between the ensemble-average residuals (predicted value minus true value) from the winter (djf) PNA pseudo-reconstructions obtained following the TT2010 approach and illustrated in Figures 7a and 8 and winter surface-air-temperature time series for the pre-industrial period up to 1849. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction.

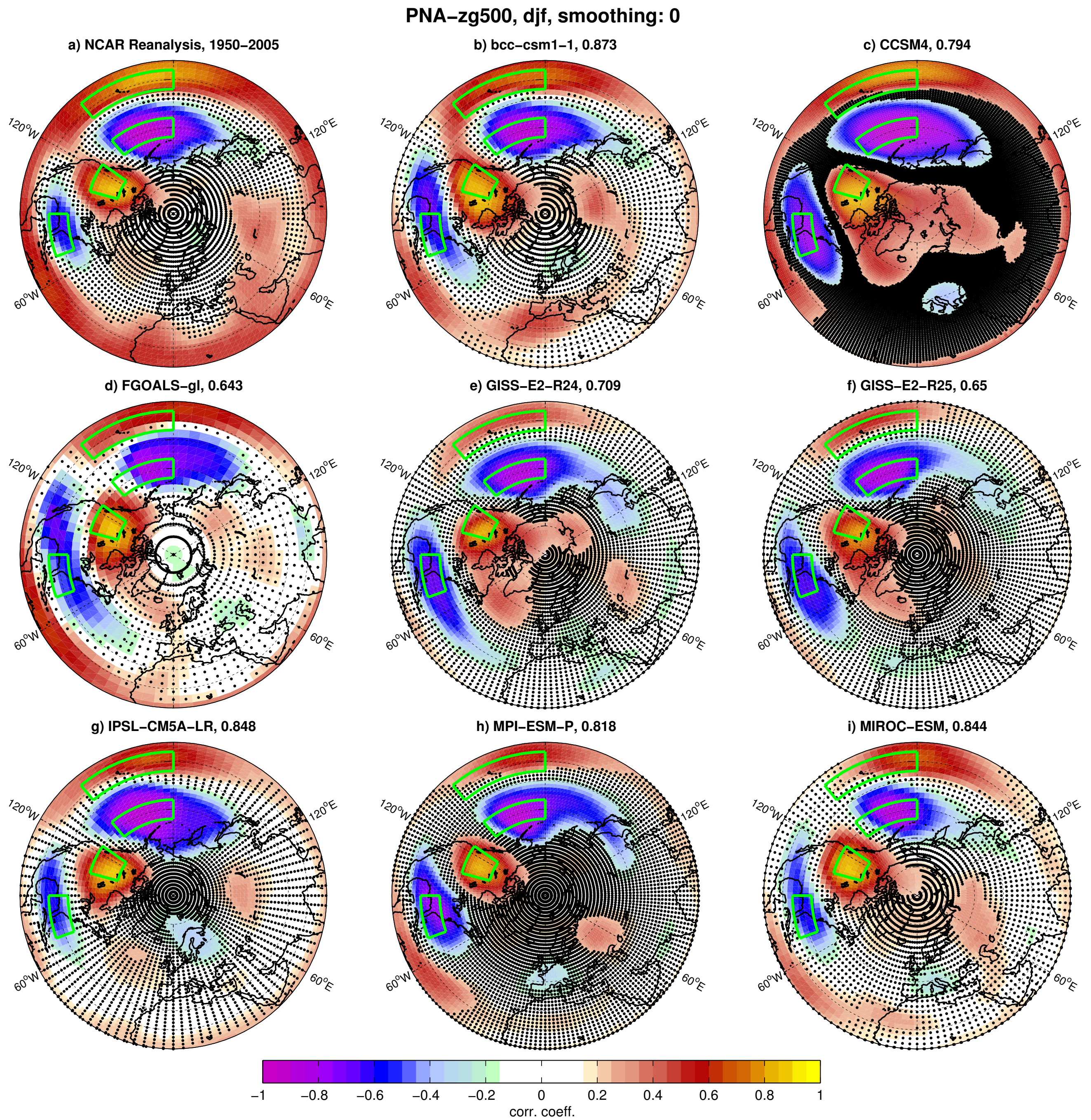

Figure 1 – Observed and simulated correlation maps between the winter PNA index and winter Z500 time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the calculation of the PNA index. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the domain north of 20º N (to this purpose NCAR data were regridded to the model grid).

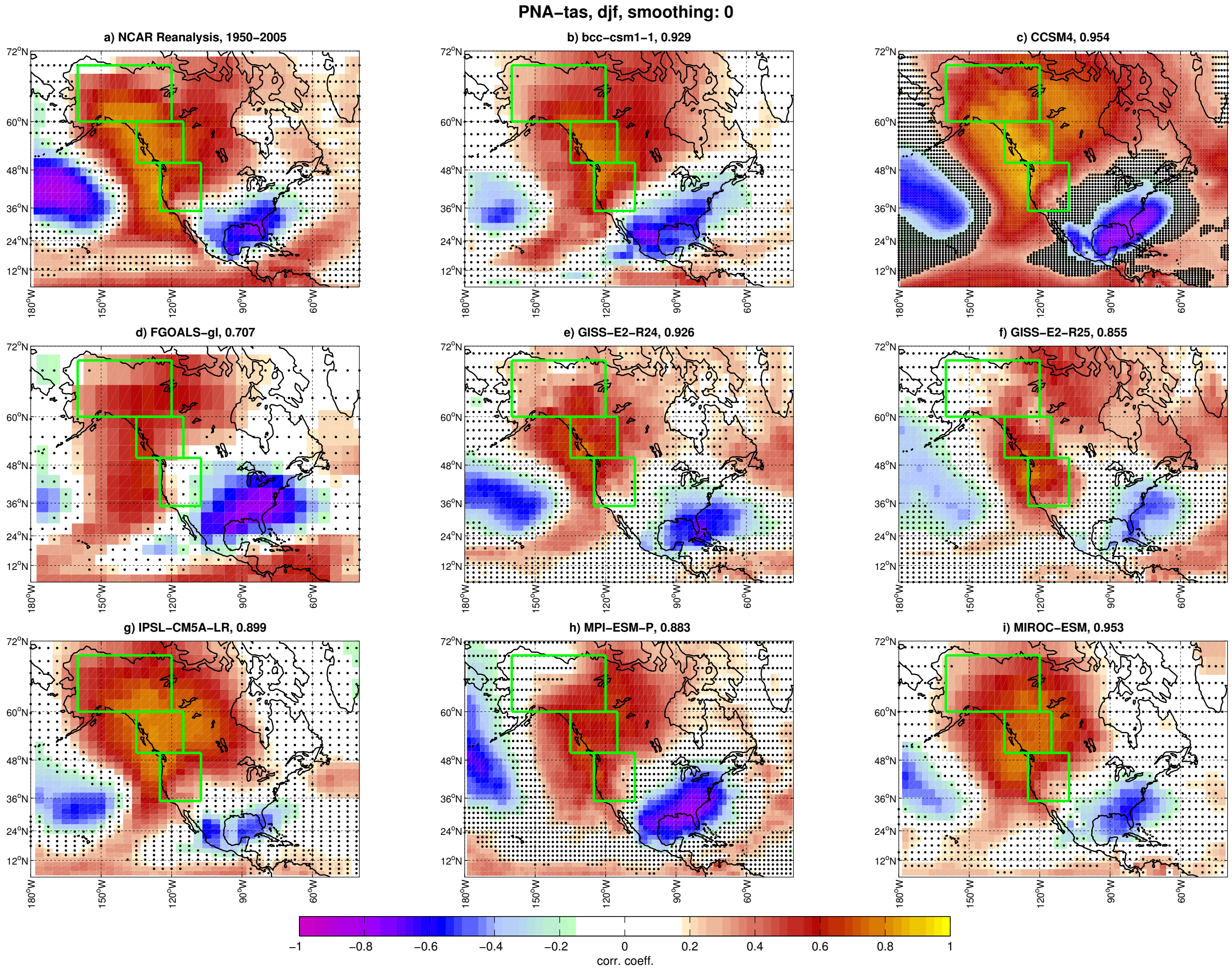

Figure 2 – Observed and simulated correlation maps between the winter PNA index and winter surface-air-temperature time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the shown land-only domain north of 12° N (to this purpose NCAR data were regridded to the model grid).

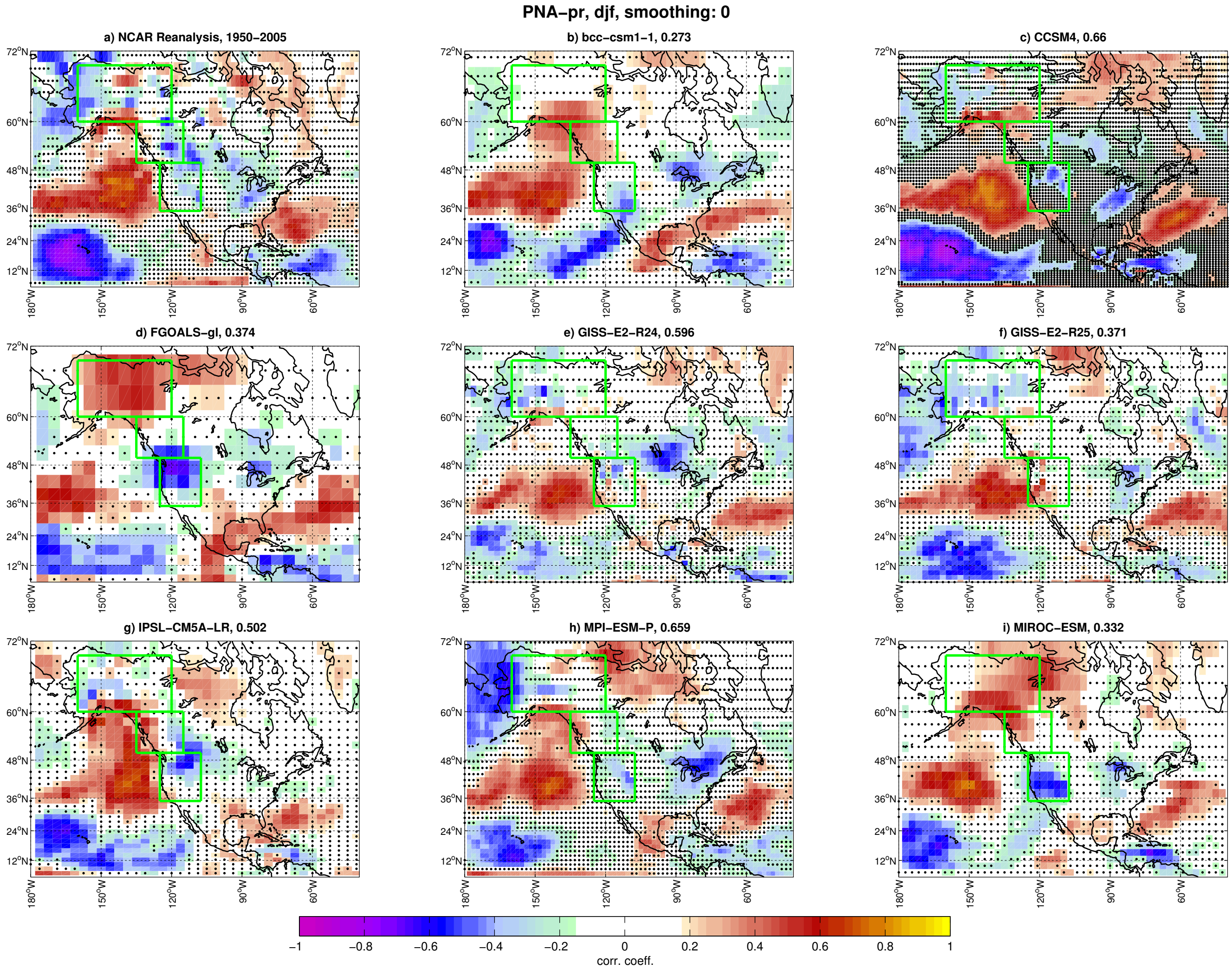

Figure 3 – Observed and simulated correlation maps between the winter PNA index and winter precipitation time series for the period 1950-2005. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction. In panels b-i, the numbers reported in the title are the spatial correlations between observed and simulated patterns calculated for the shown land-only domain north of 12º N (to this purpose NCAR data were regridded to the model grid).

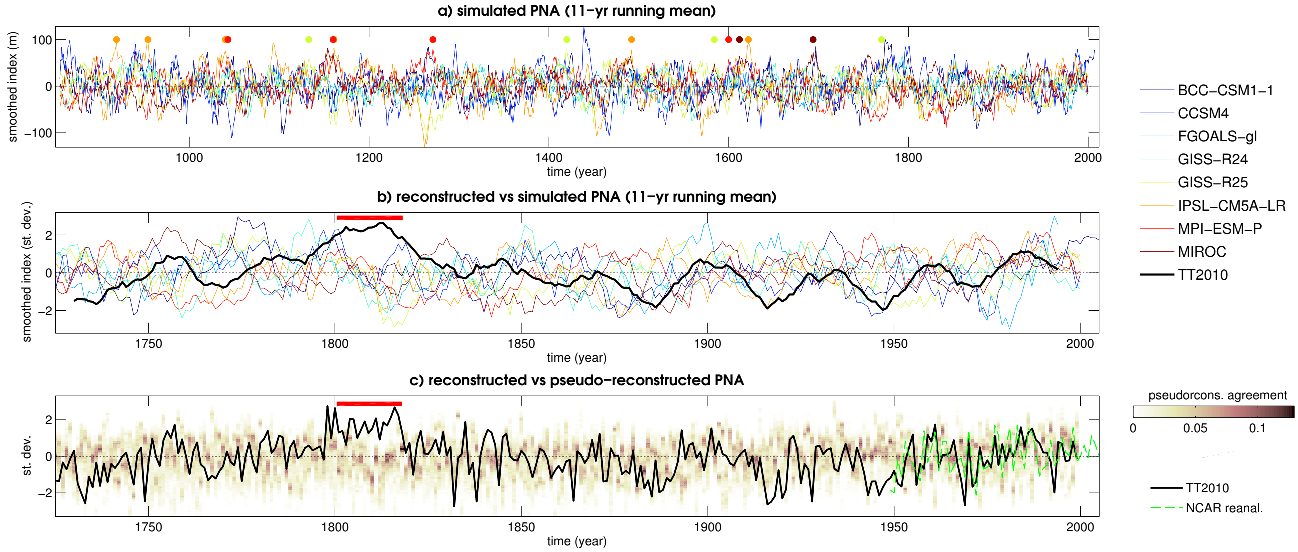

Figure 4 – Simulated, reconstructed and pseudo-reconstructed evolutions of the winter (djf) PNA index. a) smoothed PNA time series from the simulations for the whole last millennium; b) comparison between smoothed and, then, normalized (over the period 1725-1999) PNA time series from the simulations and the TT2010 reconstruction; c) comparison between normalized (over the period 1950-1999) pseudo-reconstructions (shown as agreement between pseudo-reconstructions from all the simulations) and the TT2010 reconstruction. Smoothing in panels a and b was performed through an 11-year running moving average. Dots in panel a individuate, for each simulation, occurrences of prolonged periods of positive PNA (defined as periods where the normalized smoothed index is above 1 for at least 15 consecutive years). The red bar in panels b and c highlights the approximate period of the simulations-reconstruction discrepancy.

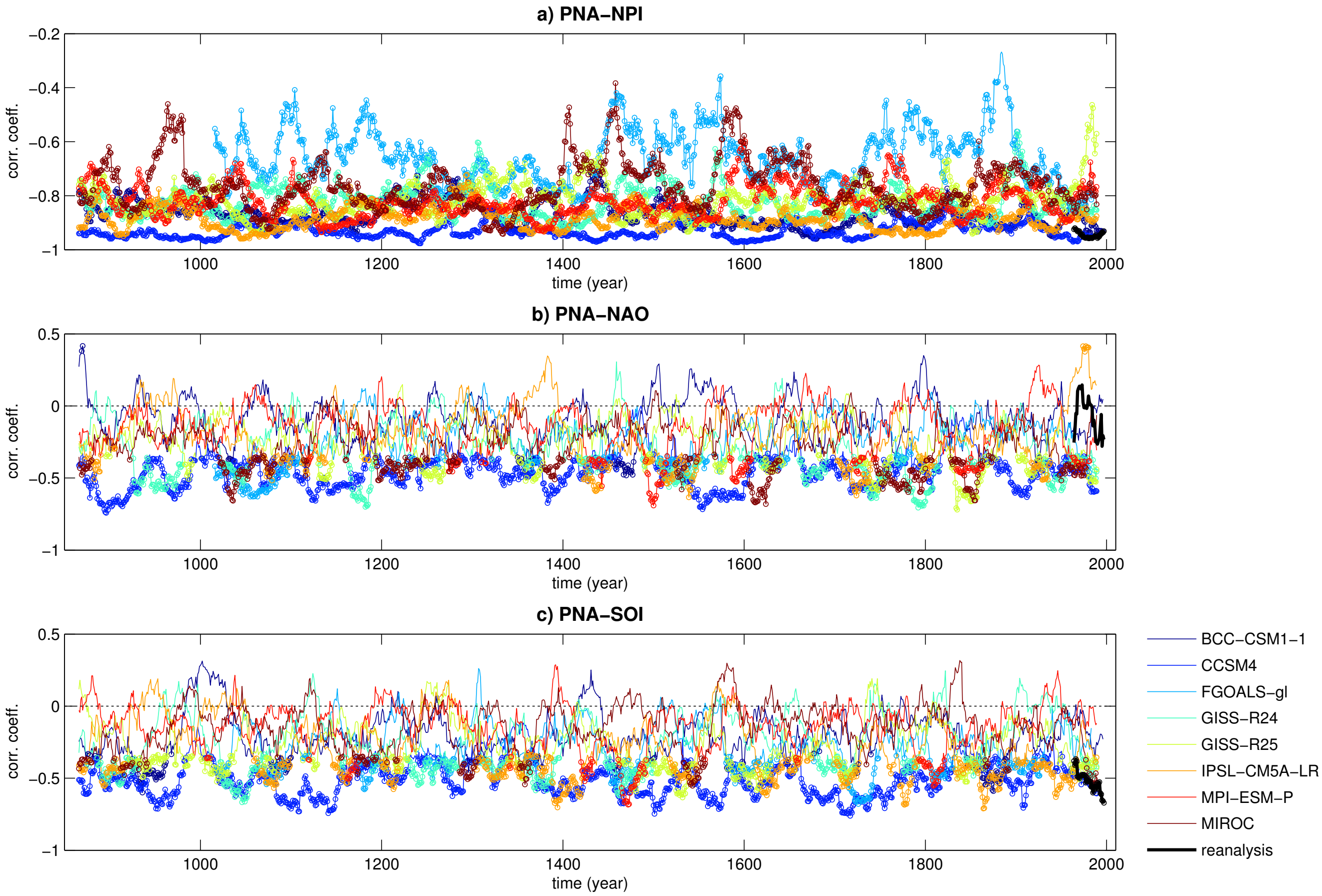

Figure 5 – Running-window (31-year) correlations of the winter (djf) PNA with winter NPI (a), NAO (b) and SOI (c) indices for simulations (colored lines) and reanalysis (black thick lines) data. Dots mark when correlation is statistically significant at 95% confidence accounting for autocorrelation.

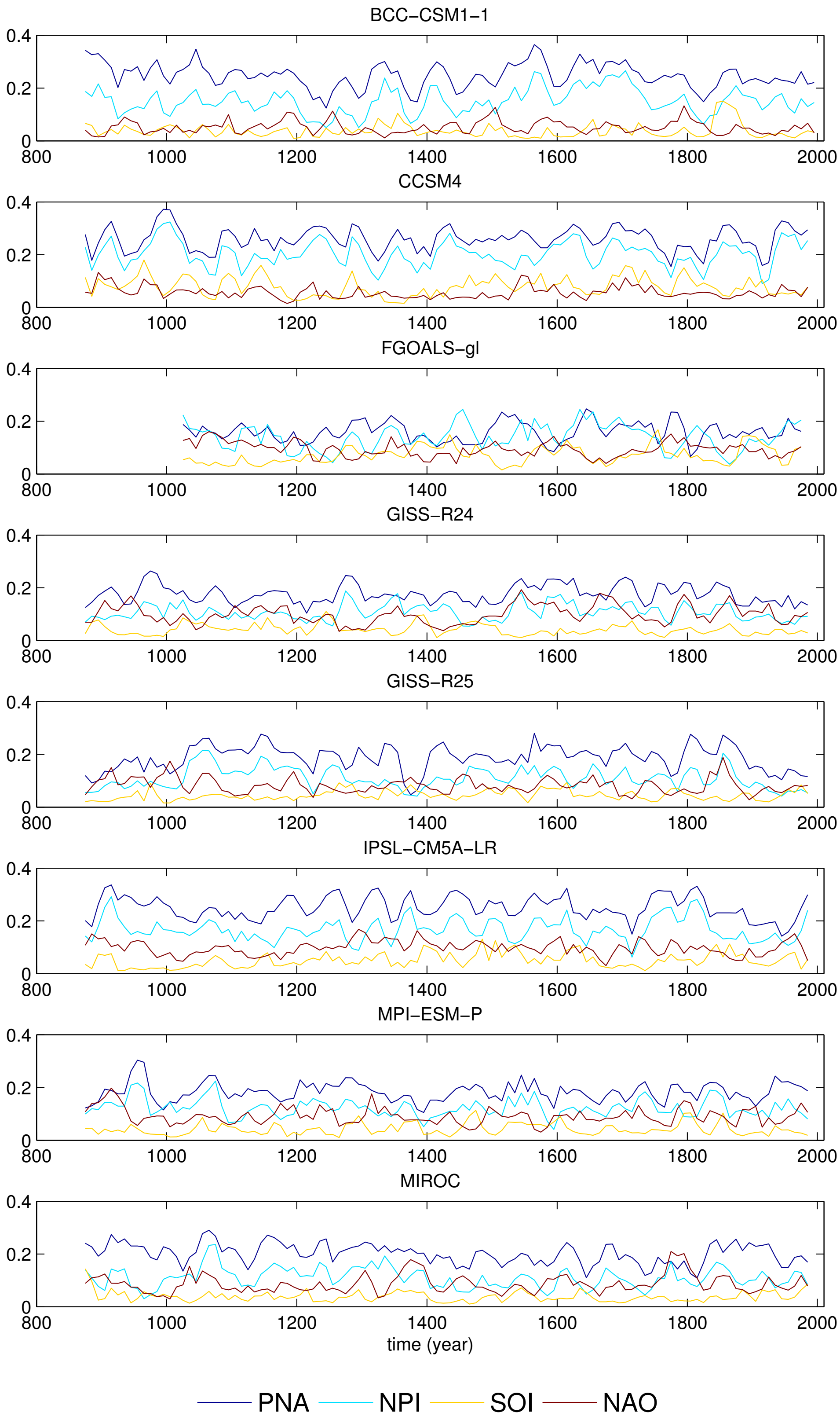

Figure 6 – Fractions of total variance of winter North American surface air temperatures (land only grid points within the domain 20-70ºN, 190-300ºE) explained by winter PNA, NPI, NAO and SOI indices for individual models. Values are calculated over decadally-paced 30-year periods.

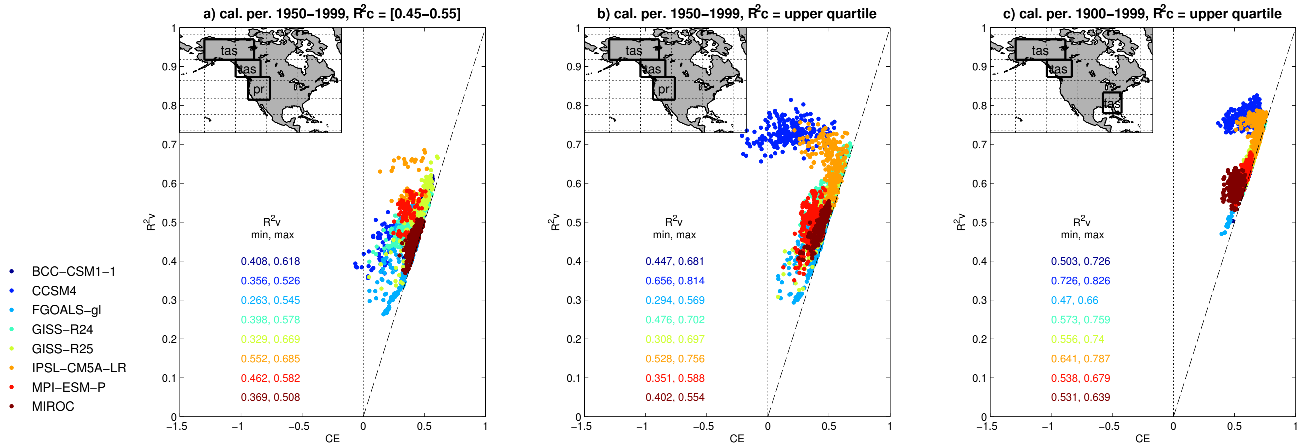

Figure 7 – Skill metrics (coefficient of determination (R^2v) and coefficient of error (CE)) of the ensemble PNA pseudo-reconstructions for the full-validation period. Different panels illustrate results from different reconstruction designs, summarized on the title of each panel: a) reconstructions based on geophysical predictors from northwestern North America, with R^2c comparable to that of the actual TT2010 reconstruction (see methods); b) same as panel a, but for best R^2c values; c) best R^2c values from an idealized design including a temperature predictor over Florida. The numbers inside each panel indicate the minimum and maximum R^2v values obtained for each model. Insets in each panel map the three boxes from where gridded data are sampled to be included as predictors, with the name reported in each box (tas: surface air temperature, pr: precipitation).

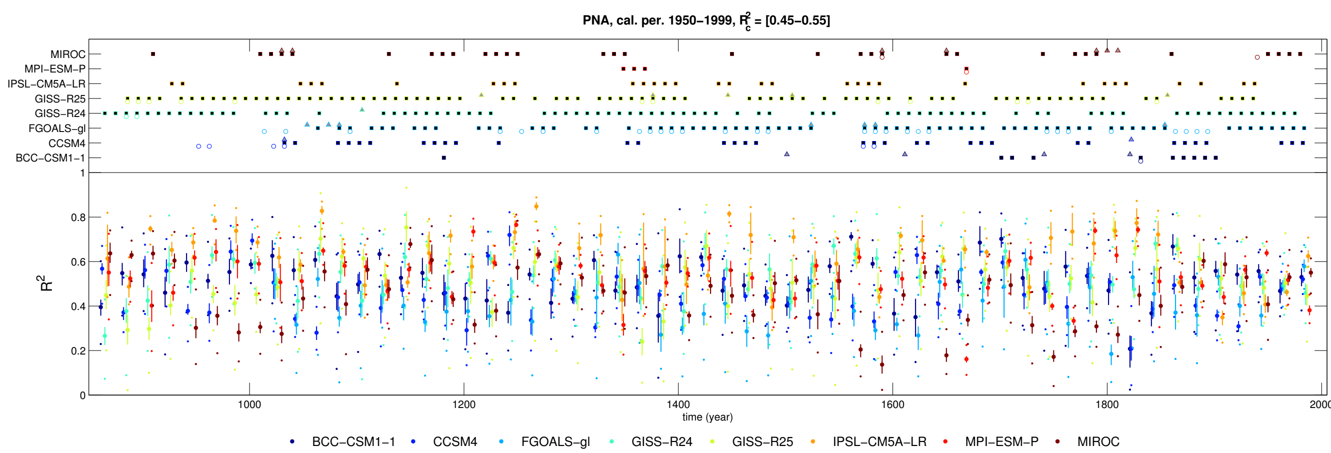

Figure 8 – Skill metric (R^2) for an ensemble of PNA reconstructions based on geophysical predictors from northwestern North America for subsequent 30-year periods (paced at 3-decade intervals). To be comparable with TT2010, only the subset of reconstructions with R^2 for the 1950-1999 calibration period in the range [0.45-0.55] are shown, as for Figure 7a. For each 30-year period, dots are minimum, mean and maximum of R^2 values, vertical lines indicate the inter-quartile interval of R^2 values. The top symbols indicate 30-year periods when the R^2 value for NPI (square), SOI (triangle) or NAO (circle) is, for at least one predictor set, better than the worse PNA value.

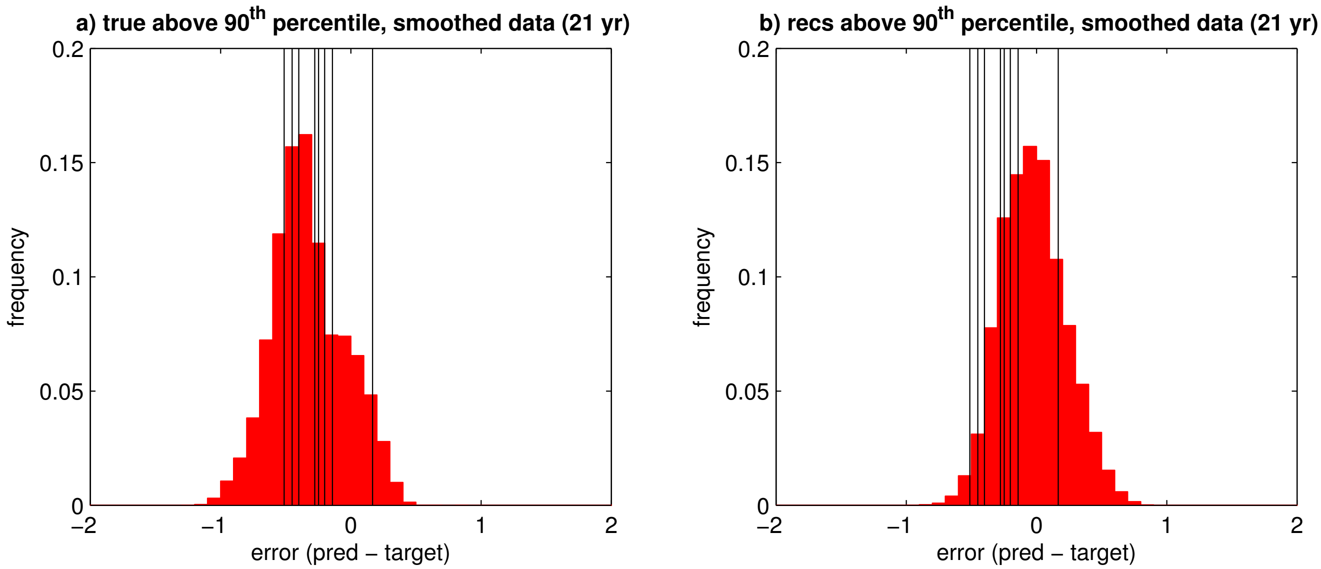

Figure 9 – Pseudo-reconstructions' accuracy in describing interdecadal positive PNA phases. Histograms are ensemble (all simulations) empirical probability distributions of residuals (predicted value minus true value) from the winter PNA pseudo-reconstructions obtained following an approach similar to TT2010 and illustrated in Figures 7a and 8 for (a) target 21-year smoothed PNA values above the 90th percentile and (b) pseudo-reconstructed 21-year smoothed PNA values above the 90th percentile. The black vertical lines indicate the full-period average residuals from individual simulations. 90th percentiles are calculated over the full simulation and therefore reflect also full-period biases in the pseudo-reconstructions. The smoothing is meant to mimic the approximately 20-year duration of the early-19th-century positive PNA phase in the TT2010 reconstruction. The considered positive PNA phases are sampled throughout the simulations, regardless of their timing.

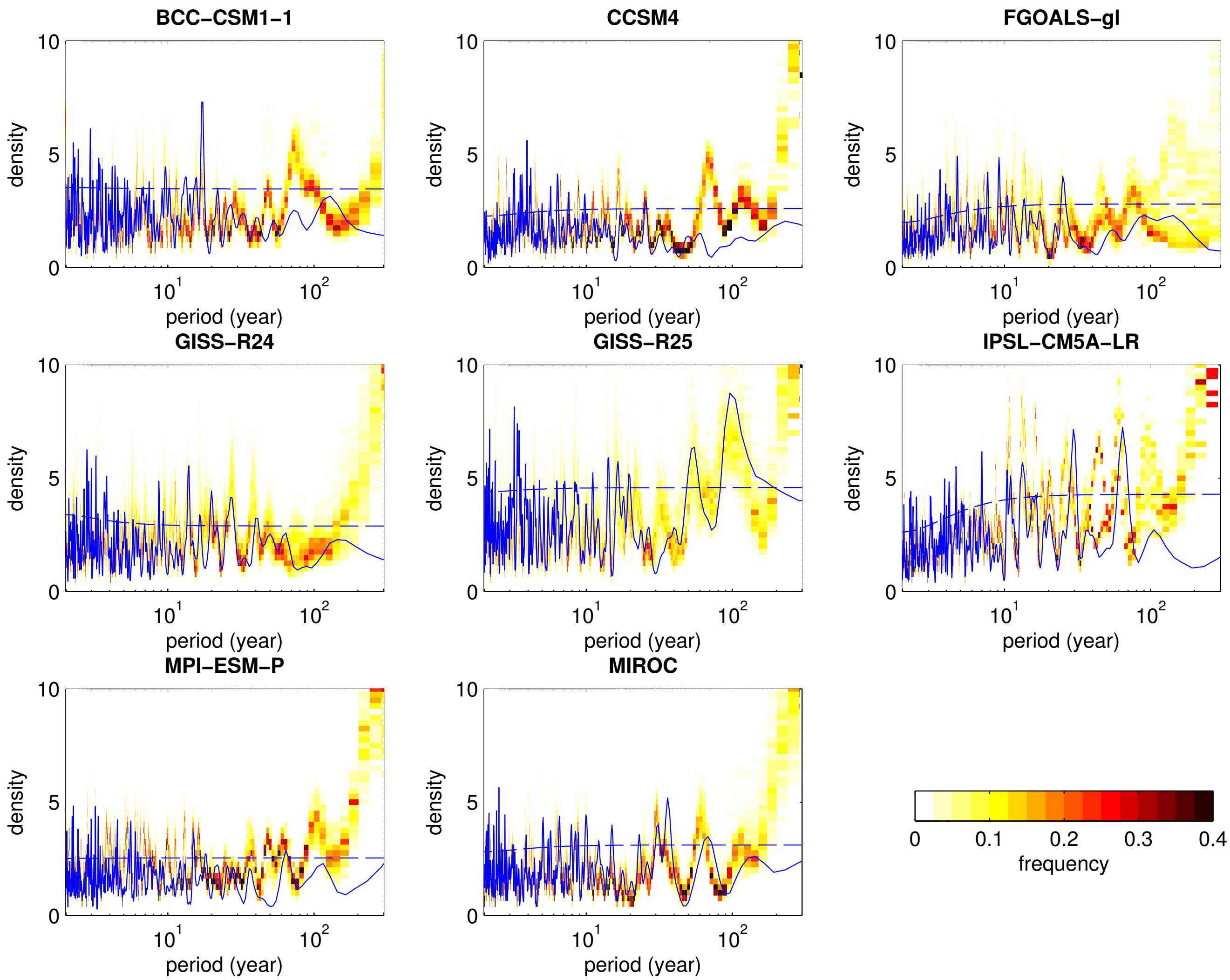

Figure 10 – Power spectral density of the winter PNA index (blue line) for individual simulations with associated 95% confidence level (blue dashed line) and agreement between the spectra of the pseudo-reconstructions (shading) obtained following an approach similar to TT2010 and illustrated in Figures 7a and 8. Agreement is defined, for a given frequency, as the fraction of total pseudo-reconstructions having power within 0.1 units^2/year intervals. All indices are standardized according to the 1950-1999 climatology.

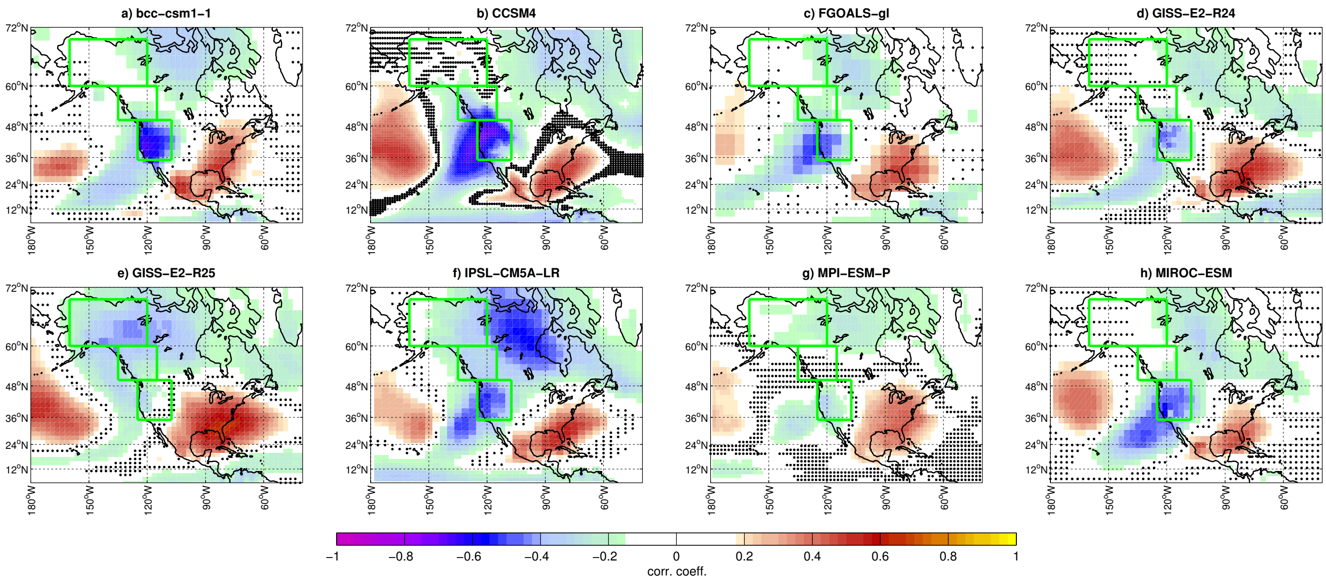

Figure 11 – Correlation maps between the ensemble-average residuals (predicted value minus true value) from the winter (djf) PNA pseudo-reconstructions obtained following the TT2010 approach and illustrated in Figures 7a and 8 and winter surface-air-temperature time series for the pre-industrial period up to 1849. Dots mark grid points where the correlation is not significant at 95% confidence accounting for autocorrelation. The green contours mark the boxes used for the TT2010 reconstruction.