

**Supplement to:**

**Consistency of the Multi-Model CMIP5/PMIP3-past1000 Ensemble**

*Oliver Bothe*<sup>1,2,3\*</sup> (*ol.bothe(at)gmail.com*), *Johann H. Jungclaus*<sup>2</sup>, *Davide Zanchettin*<sup>2</sup>

<sup>1</sup> Klimacampus, University of Hamburg, Grindelberg 5, 20144 Hamburg, Germany

<sup>2</sup> Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany

<sup>3</sup> now at Leibniz-Institut für Atmosphärenphysik e.V. an der Universität Rostock,  
Schlossstraße 6, 18225 Kühlungsborn

\* corresponding author

*Includes supplements:*

**A. Additional Figures for the analysis of the global field data** (Figures S1 to S5)

**B. Figures illustrating the Gao- and Crowley-subensembles**

**C. Additional analysis on northern hemispheric mean temperature**

## **A. Additional Figures for the analysis of the global field data**

The Figures S1 to S5 supplement Figure 2 of the main manuscript by extending the selection of grid-points.

Caption for all five Figures:

Top rows: Residual quantile-quantile plots for a random selection of 15 grid-points. For the first (light gray), second (dark gray) and last (colored) sub-periods of 28 records. Bottom rows: Rank histogram counts for the random selection of grid-points for the three sub-periods (first to last, light to dark gray) and the full period (black). Full-period counts are scaled to match the frequencies of the sub-periods. Large (small) red squares mark grid-points where spread or bias deviations are significant over the full (the individual sub-)period. Small squares from left to right for the first, second and third sub-period. Blue squares indicate not significant statistics. Locations are given in titles of individual panels. For color coding see Figure 2 of the main manuscript.

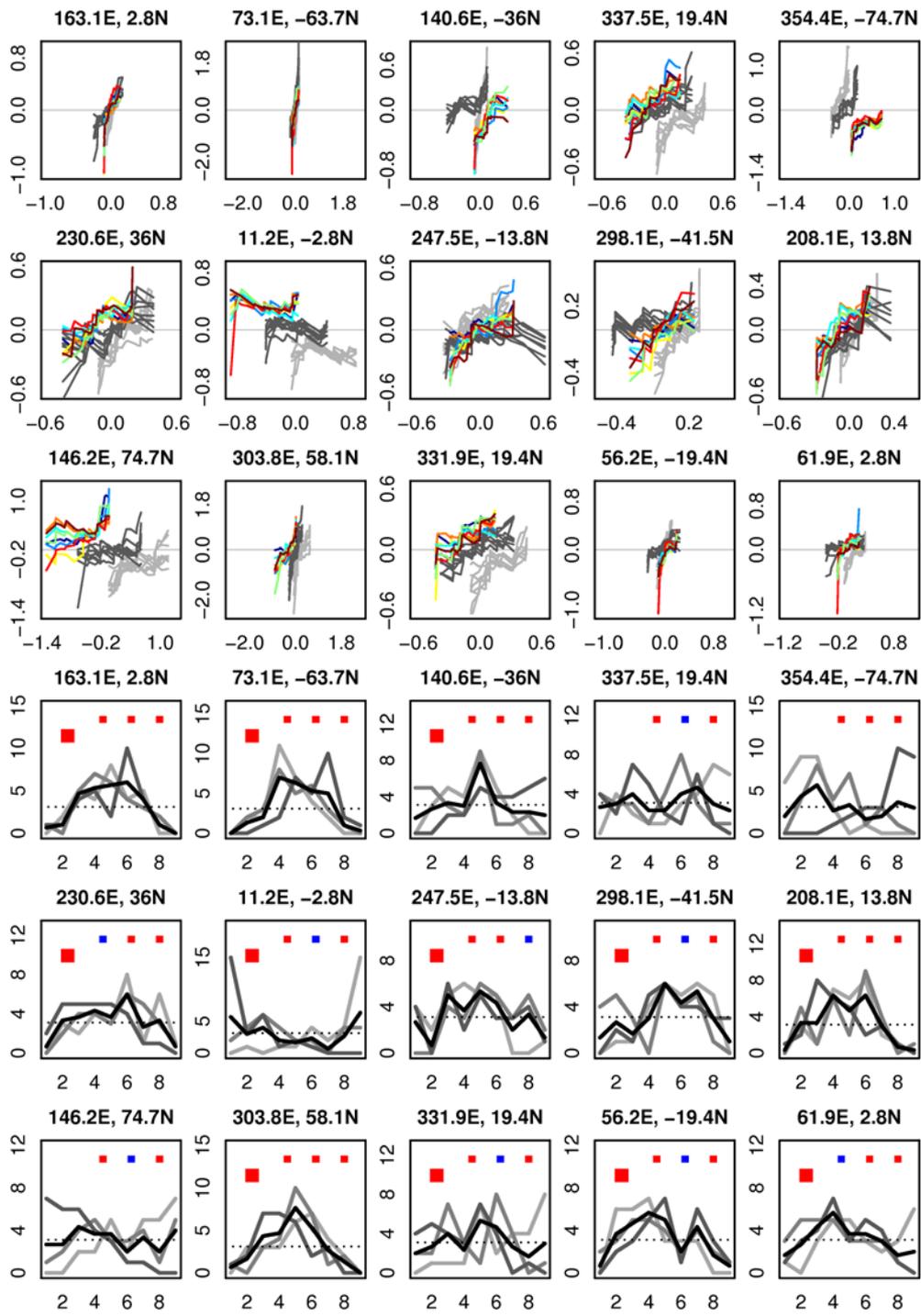


Figure S1

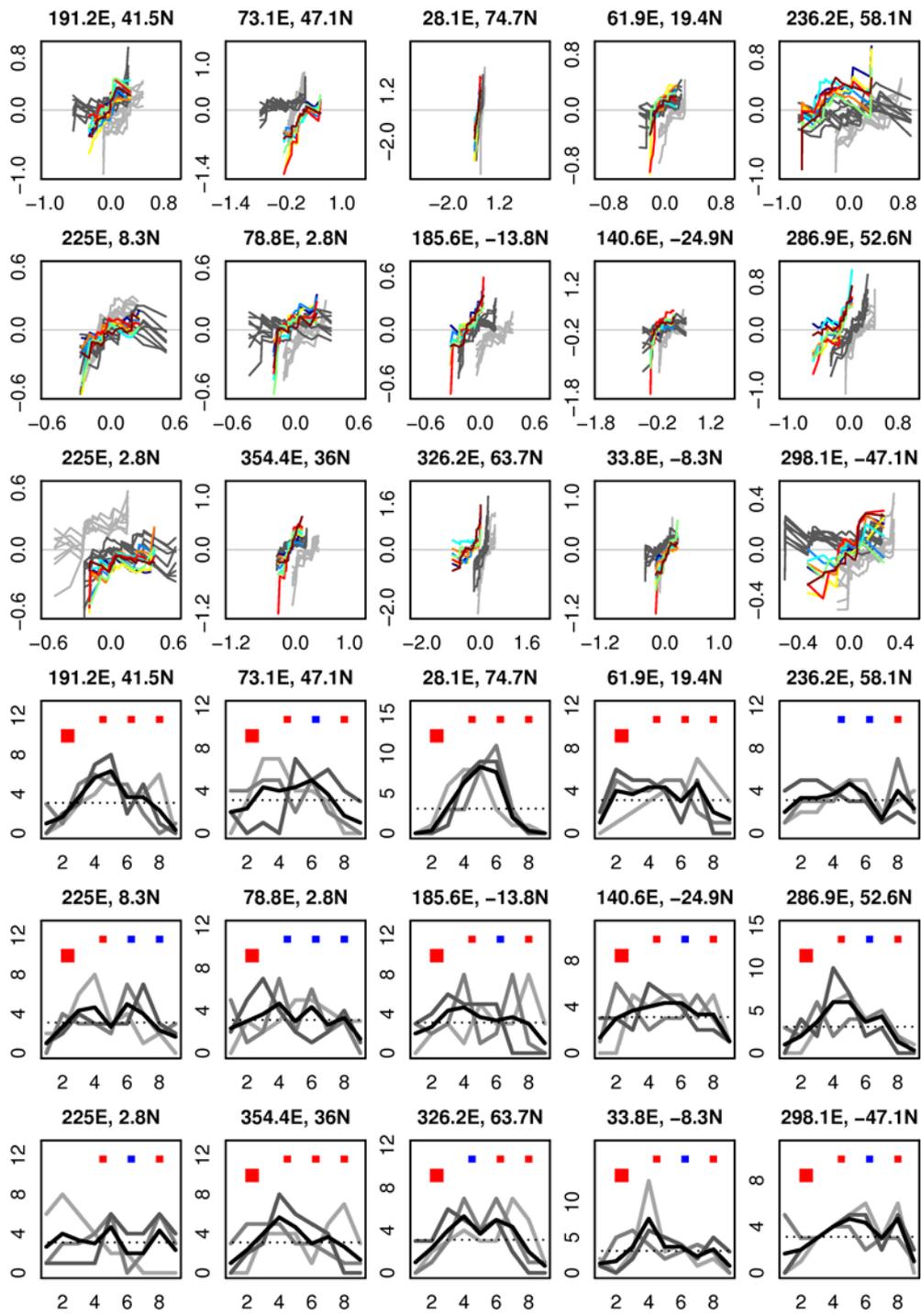


Figure S2

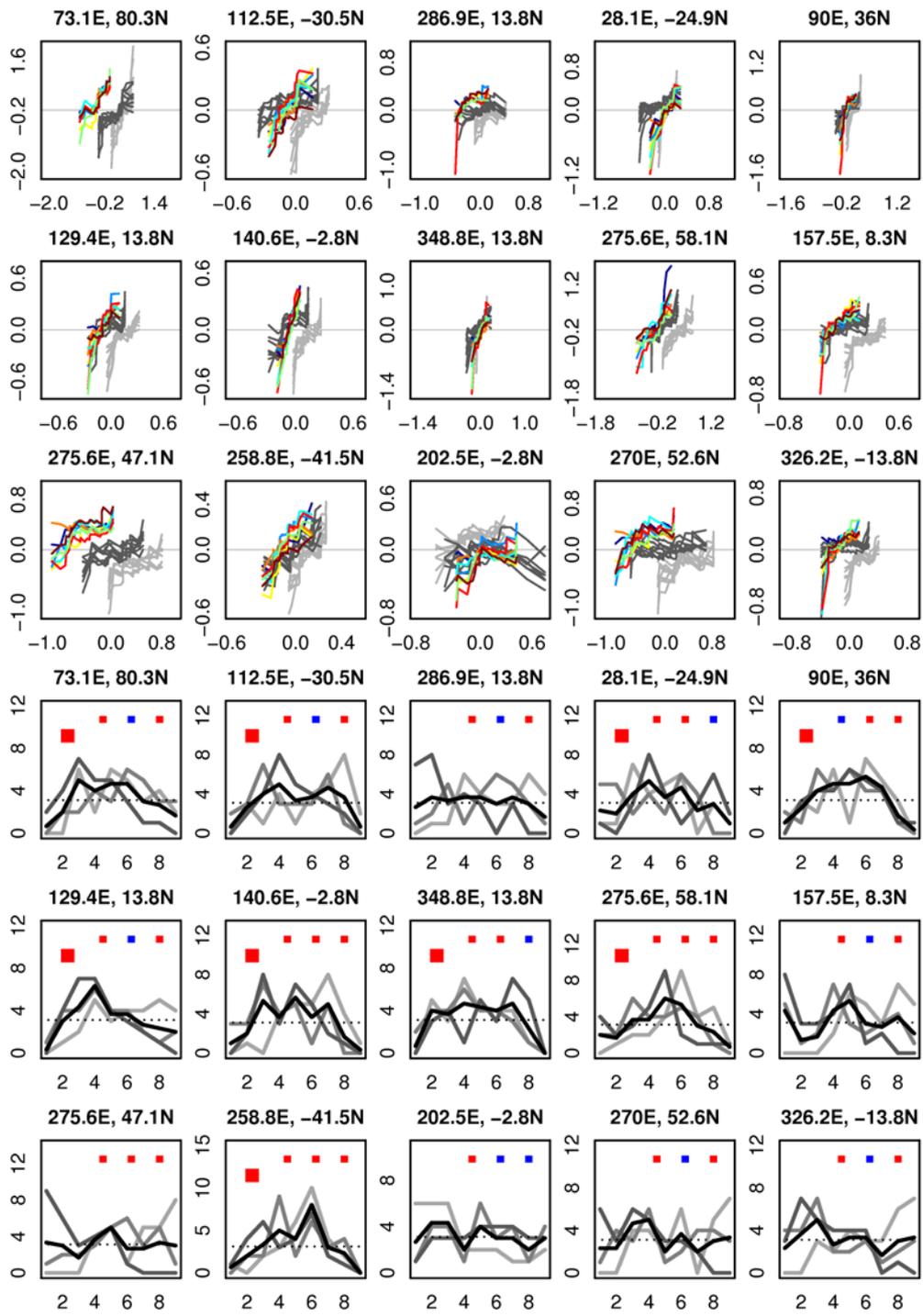


Figure S3

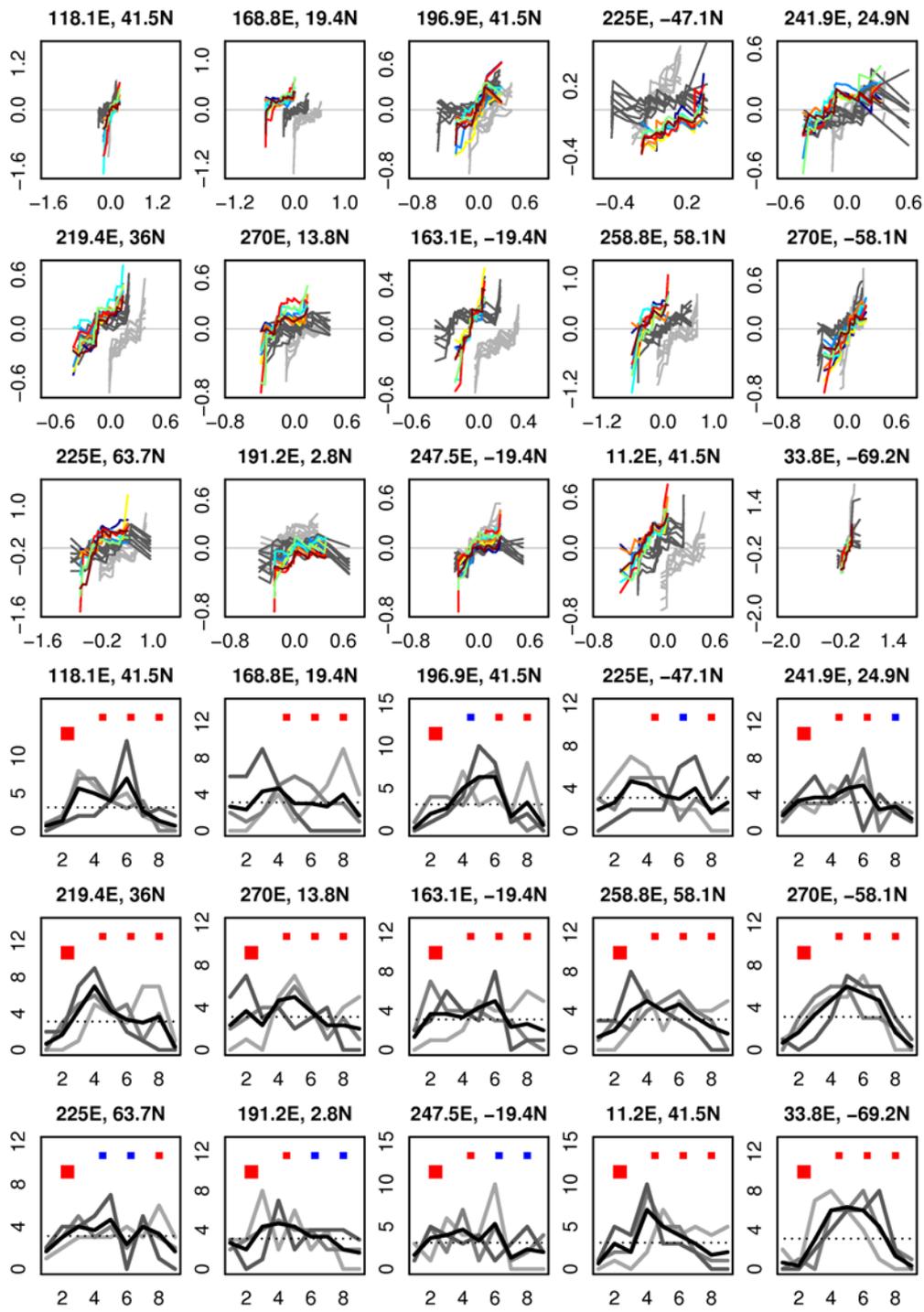


Figure S4

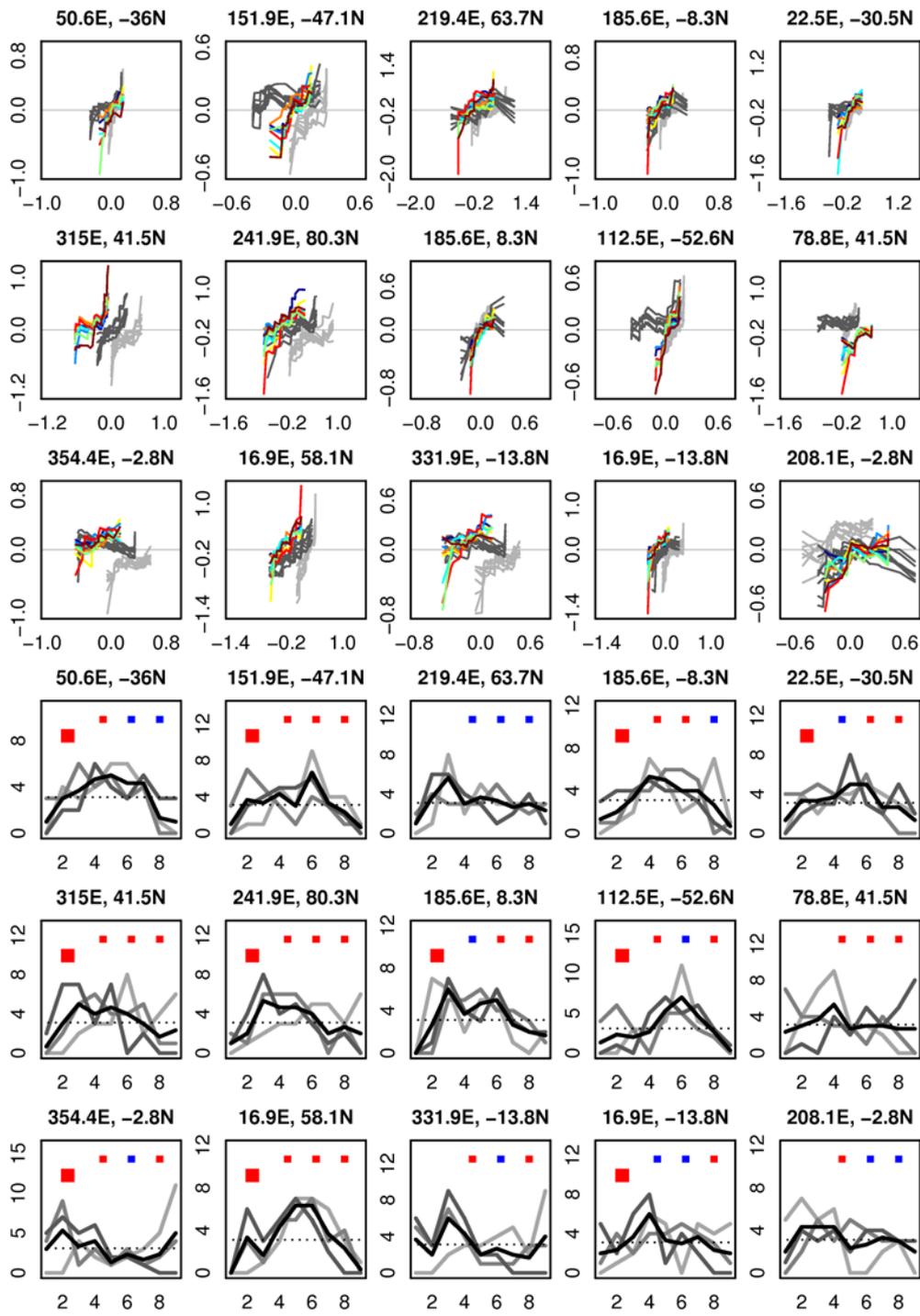


Figure S5

## B. Figures illustrating the Gao- and Crowley-subensembles

Figures S6 to S8

The following Figures illustrate the differences between the simulations using the volcanic forcing data by Gao et al. (2008) and those using the data by Crowley (e.g. Crowley et al., 2008).

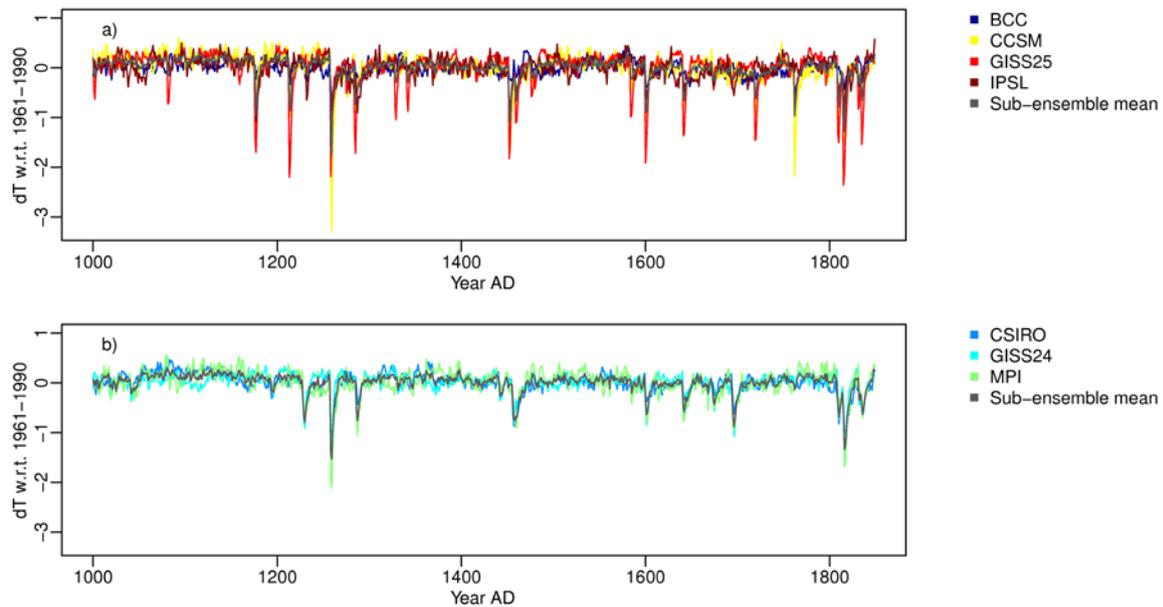


Figure S6: Annual northern hemisphere mean temperature time-series of the simulations using (a) the Gao et al. (2008) and (b) the Crowley data for the volcanic forcing.

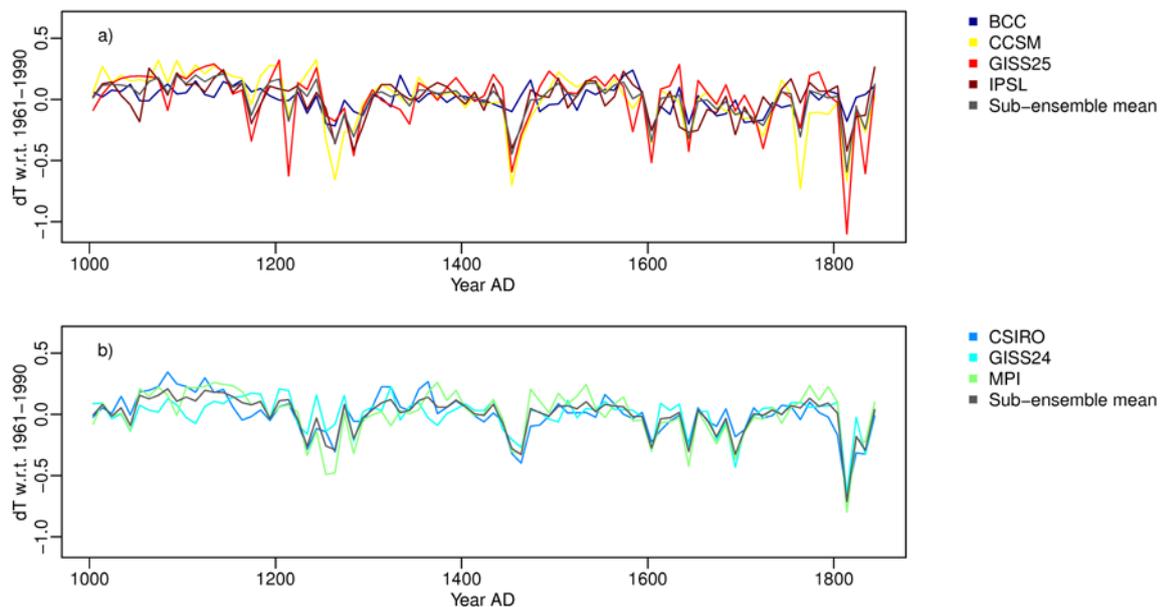


Figure S7: As Figure S6 but for the decadal mean data.

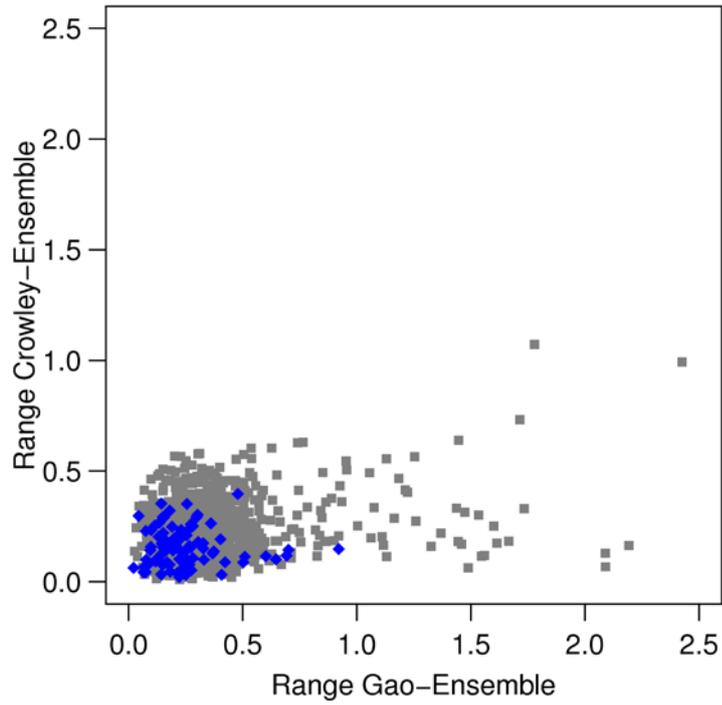


Figure S8: Range of the volcanic sub-ensembles, i.e. largest minus smallest Northern Hemisphere temperature anomaly for each date. Grey squares for the annual resolution, blue diamonds for the decadal resolution. The x-axis plots the range of the Gao-ensemble and the y-axis the range of the Crowley-ensemble.

## C. Additional analysis on northern hemispheric mean temperature

In the main manuscript we concentrate on the global temperature field reconstruction by Mann et al. (2009). We do not explicitly consider large-scale or hemispheric-mean reconstructions like, e.g. the one provided by Mann et al. (2008). In the following we add a short discussion on the consistency of the PMIP3 past1000 ensemble with a number of reconstructions of northern hemispheric-mean temperature. We concentrate on the Frank et al. (2010) reconstruction-recalibration ensemble for the northern hemisphere (from 0 to 90N) but also include the reconstruction by Ljungqvist (2010) for the northern hemisphere extra-tropics (30N to 90N). For a discussion of the ensembles see Frank et al. (2010), Bothe et al. (2013) and the main manuscript.

### *C.1 Ljungqvist 2010*

We start by shortly discussing the consistency of the PMIP3 past1000 ensemble with the reconstruction by Ljungqvist (2010). Ljungqvist provides a dedicated uncertainty estimate for his extra-tropical reconstruction.

Full-period rank histograms are overall flat (Figure S9c,d); this uniformity of the rank counts indicates consistency of the ensemble. The sub-period assessment suggests that the simulation ensemble displays significant biases during the early and late sub-periods (Figure S9d, the critical value for our test is 2.706, see Methods in main manuscript). The negative bias in the early sub-period and the positive bias in the late sub-period balance each other over the full period.

Climatologic residual quantiles suggest strong over-estimation of the cold quantiles over the full period and all sub-periods (Figure S9a,b). The quantiles are quite small for FGOALS over the full period and generally rather small (except for the cold tail) for the early sub-period. The other sub-periods indicate over-dispersive climatologies in the simulations as seen by the positive slope in the residuals.

Thus the long-term trends differ prominently between the simulations and the reconstruction. We also note the obvious difference in inter-decadal variability between the reconstruction and the ensemble with the Ljungqvist data showing only very small variations (Figure S9e).

## *C.2 Frank et al., 2010*

We do not consider the full Frank et al. ensemble but discuss only three exemplary reconstructions. These are the data by Jones et al. (1998), by Frank et al. (2007) and by Mann et al. (2008). For each reconstruction we choose the ensemble members recalibrated to the periods 1850 to 1890, 1920 to 1960, 1850 to the last available date and the last available 40 year period. As uncertainty estimate we use the ensemble standard deviation of the series calibrated to these periods for all nine original reconstructions included by Frank et al. (2010).

We now shortly present the consistency of the PMIP3 past1000 ensemble with the chosen members of the three reconstruction sub-ensembles provided by Frank et al. (2010). We show for all three reconstructions the results for the series recalibrated to the period 1920 to 1960 and further plots are added where necessary.

### *C.2.1 Jones et al., 1998*

First, we shortly discuss the reconstruction by Jones et al. (1998). Rank counts over the full period are uniform indicating consistency if we consider the series recalibrated to the period 1920 to 1960 (Figure S10c). Consistent results are generally found for the different sub-periods. However, the single deviation  $\chi^2$  goodness-of-fit test suggests a significant (warm) bias of the simulation ensemble in the late sub-period (Figure S10d).

The residual quantiles depict a more intricate situation compared to the Lundqvist et al. reconstruction. Some simulations, e.g., GISS25 and CSIRO, are under-dispersive over the full period, others, e.g., BCC and FGOALS, over-estimate the cold tail quantiles while under-estimating the warm tail (Figure S10a). A slight warm bias is consistently seen across the simulations in the late sub-period, confirming the probabilistic assessment (Figure S10b). In the central sub-period, the ensemble looks climatologically consistent except for a large over-estimation of the cold and the warm tail quantiles. In the early sub-period, the simulations generally over-estimate the cold tail while under-estimating the warm tail as already found for some individual simulations over the full period.

The time-series suggest that the reconstruction shows stronger centennial to multicentennial

variability than the simulations. Variability is overall of comparable amplitude in both data sources on decadal time-scales (Figure S10e).

The Jones et al. (1998) series from the Frank et al. reconstruction-ensemble agrees with the conclusion made in the main manuscript based on the Mann et al. (2009) global field data that the disagreement between PMIP3-past1000 simulations and large-scale reconstructions is mostly due to differences in the longer-term trends (multidecadal-to-centennial and longer timescales) of the corresponding indices.

We find the following for the other three recalibrated series. The earliest recalibration period (1850 to 1890, not shown) leads to nearly unequivocal rejection of consistency. All measures suggest strong over-dispersion of the simulation ensemble except probabilistically for the early sub-period. Over-dispersion is increased when the recalibration period 1850 to 1990 is considered (Figure S11). For the recalibration period 1950 to 1990, the bootstrap significance intervals allow for consistency over the full-period although displaying significant probabilistic over-dispersion over the full-period and over the central sub-period according to the  $\chi^2$  goodness-of-fit test (not shown).

The different results for the four recalibrated reconstructions may be interpreted as an indication of lacking robustness of reconstruction results. Depending on the recalibration period we may see either very large over-dispersion, near consistency or potentially under-dispersion.

### *C.2.2 Frank et al., 2007*

Though the data by Frank et al. (2007) displays considerably more multidecadal variability than the simulations for all four recalibration periods, the simulation ensemble is found to be consistent with this reconstruction (Figure S12). Nevertheless we also see in the residual quantiles that some simulations prominently over-estimate the cold tail quantiles (Figure S12a,b).

### *C.2.3 Mann et al., 2008*

Over-dispersive relations are found between the ensemble and the reconstruction by Mann et

al. (2008) (Figure 13a,b), which are mainly due to weaker interdecadal variability in the reconstruction for the recalibration period 1920 to 1960 (Figure S13e). Lacks of consistency are not significant only for the early sub-period and the probabilistic assessment (Figure S13d). These results hold for the recalibration periods 1850 to 1890 and 1850 to 1990.

The simulation ensemble is consistent with the Mann et al. (2008) reconstruction recalibrated to the period 1950 to 1990 over the full period and the central sub-period (Figure S14c,d). However, significant opposite probabilistic biases are found for the early (cold bias) and late sub-periods (warm bias) (Figure S14d). This recalibrated reconstruction shows more interdecadal variability than the simulations (Figure S14e). (Slight) climatological biases are found over the sub-periods, and full period residuals are either under-dispersive (BCC), consistent or underestimate warm and cold tails equally (especially GISS25, CCSM and CSIRO).

### *C.3 Summary*

Our assessment of the consistency of the PMIP3 ensemble relative to a number of commonly analysed reconstructions of Northern Hemisphere annual-mean temperatures allows for the following general conclusions.

Reconstructions potentially show weaker inter-decadal variability compared to the simulations, but vary more on centennial and longer time-scales. Related to the differing variability the simulations generally differ from the reconstructions in their long-term trend. Simplified, simulations overall exhibit a less pronounced Medieval Warm Period and a less pronounced or at least shorter Little Ice Age compared to the reconstructions.

We find generally a prominent lack of consistency of the simulation ensemble relative to the reconstructions. However, the simulation ensemble is consistent with some reconstructions over some periods. Indeed, we cannot reject consistency relative to the reconstruction by Frank et al. (2007) over all the considered recalibration periods.

The recalibrated series obtained from a certain reconstruction may differ prominently in their amount of variability on decadal to multicentennial time-scales, i.e. the latter depends on the considered recalibration period. This complicates any assessment of the consistency between

a simulation ensemble and a reconstruction. It highlights the inherent uncertainty in our estimates for the climate of the last millennium and stresses the necessity of increasing our confidence in reconstructions, in simulations and in the forcing estimates used for the simulations.

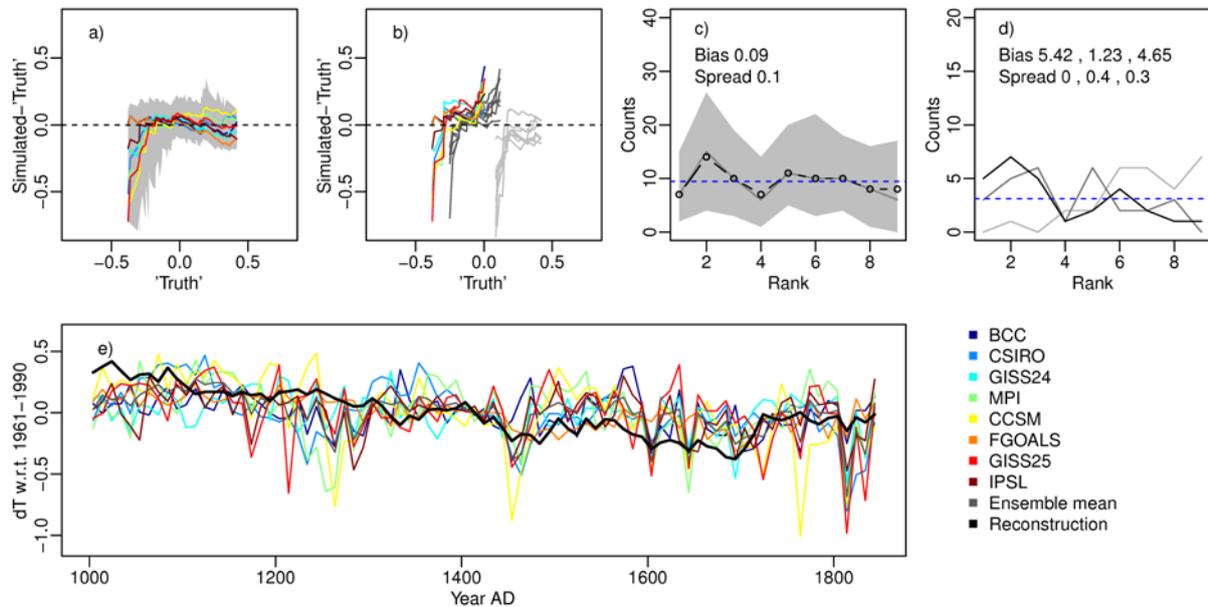


Figure S9: Analysis of the consistency of the PMIP3 simulation ensemble relative to the reconstruction of extratropical annual northern hemispheric mean temperature by Ljungqvist (2010). (a,b) Residual quantile-quantile plots for (a) the full period and (b) three sub-periods (defined as for Fig. 2 in the main manuscript) of 28 records (early, light gray, middle, dark gray, late, colored). (c,d) Rank histogram counts for (c) the full period and (d) the three sub-periods (light gray to black). Numbers are the 2 statistics for the periods. In (d) numbers refer, from left to right, to the early to late sub-periods. Blue horizontal lines give the expected average count for a uniform histogram. (e) Time series. Color-code as in legend except for shading. Shading for residual-quantiles and rank-counts (a, c) gives the 90% envelope of block-bootstrapping 2000 replicates of block-length 5.

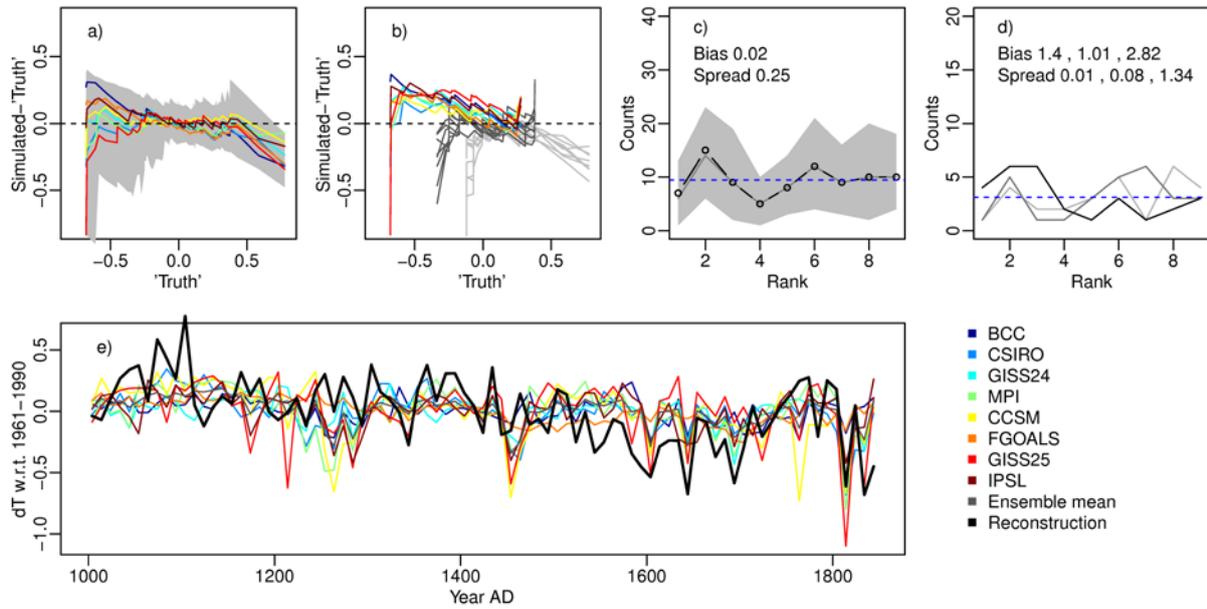


Figure S10: As Figure S9 but for the reconstruction by Jones et al. (1998) recalibrated by Frank et al. (2010) to the period 1920 to 1960.

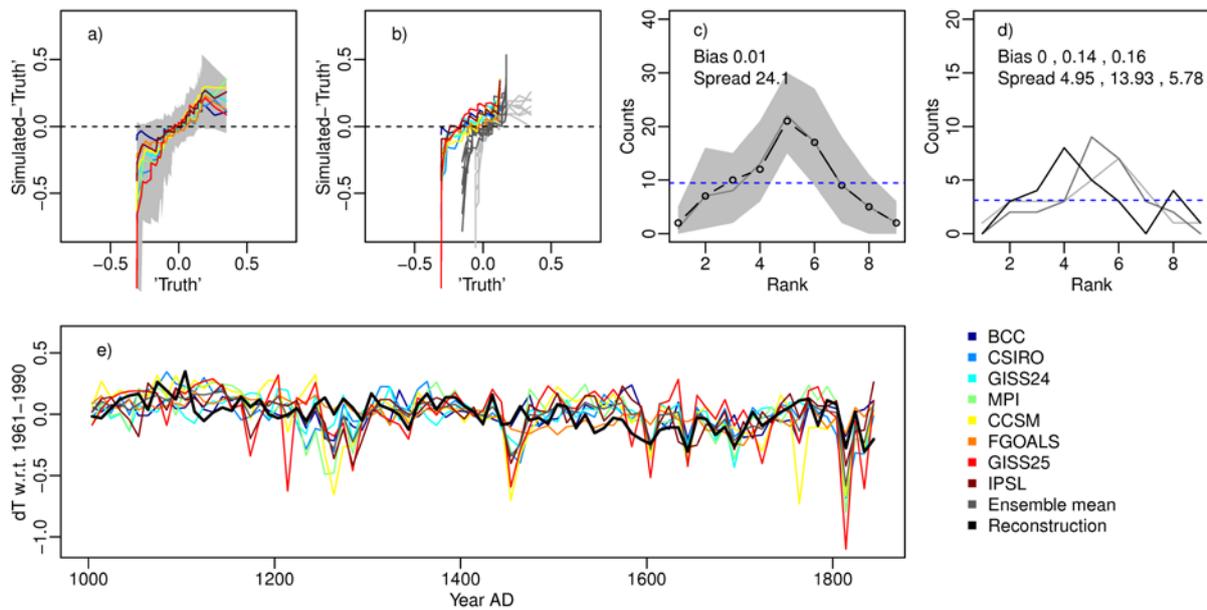
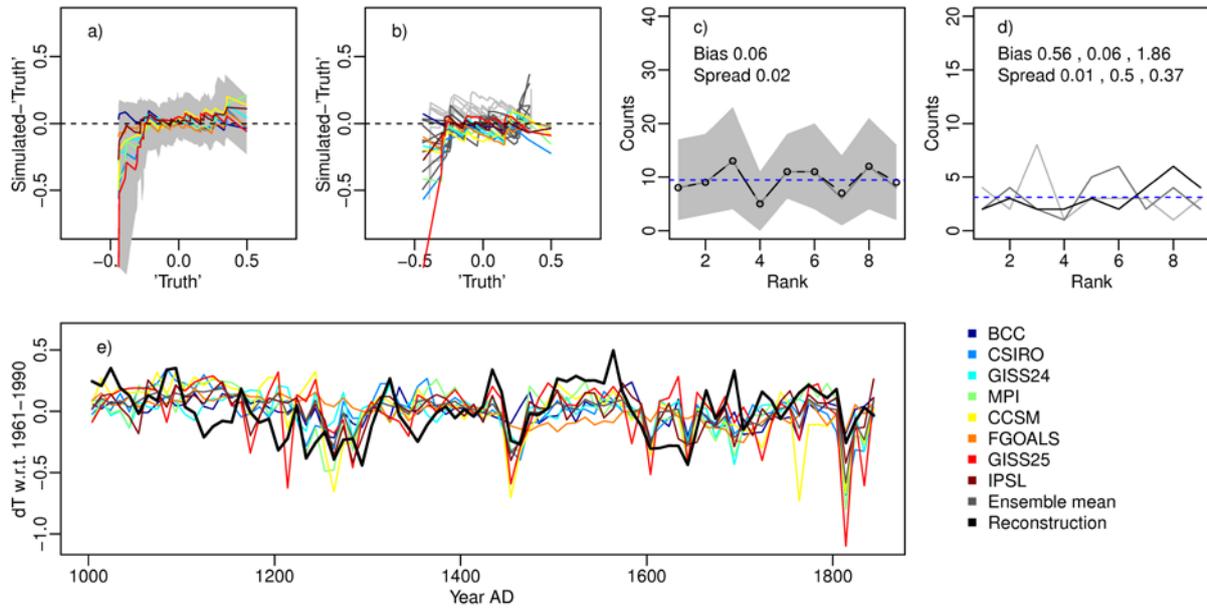


Figure S11: As Figure S9 but for the reconstruction by Jones et al. (1998) recalibrated by Frank et al. (2010) to the period 1850 to 1990.



Figures S12: As Figure S9 but for the reconstruction by Frank et al. (2007) recalibrated by Frank et al. (2010) to the period 1920 to 1960.

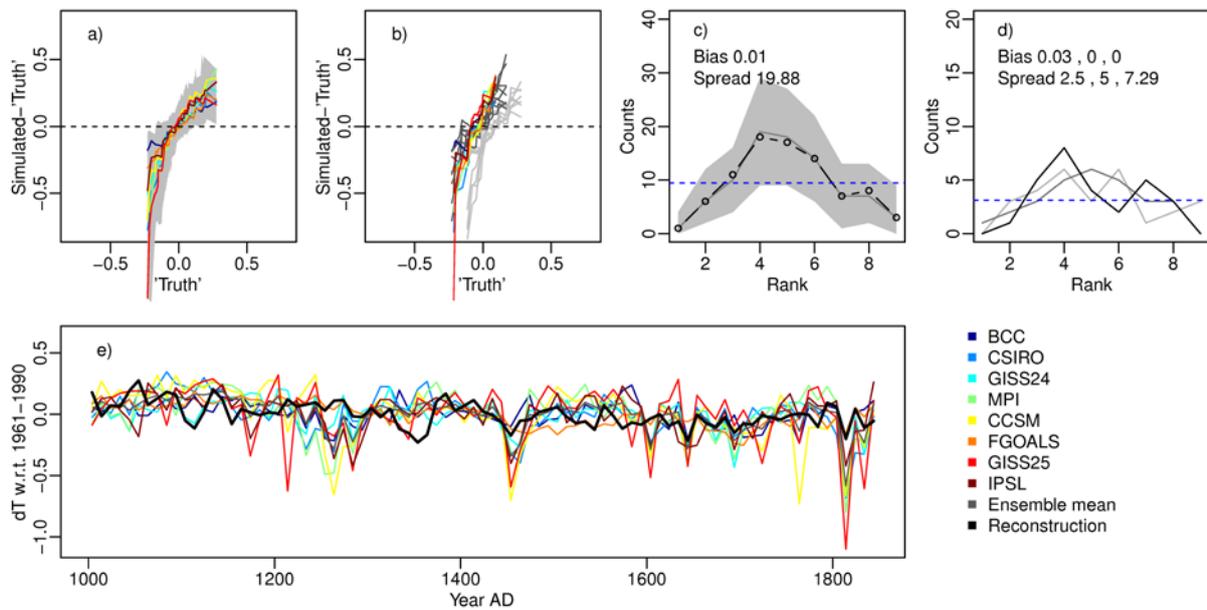


Figure S13: As Figure S9 but for the reconstruction by Mann et al. (2008) recalibrated by Frank et al. (2010) to the period 1920 to 1960.

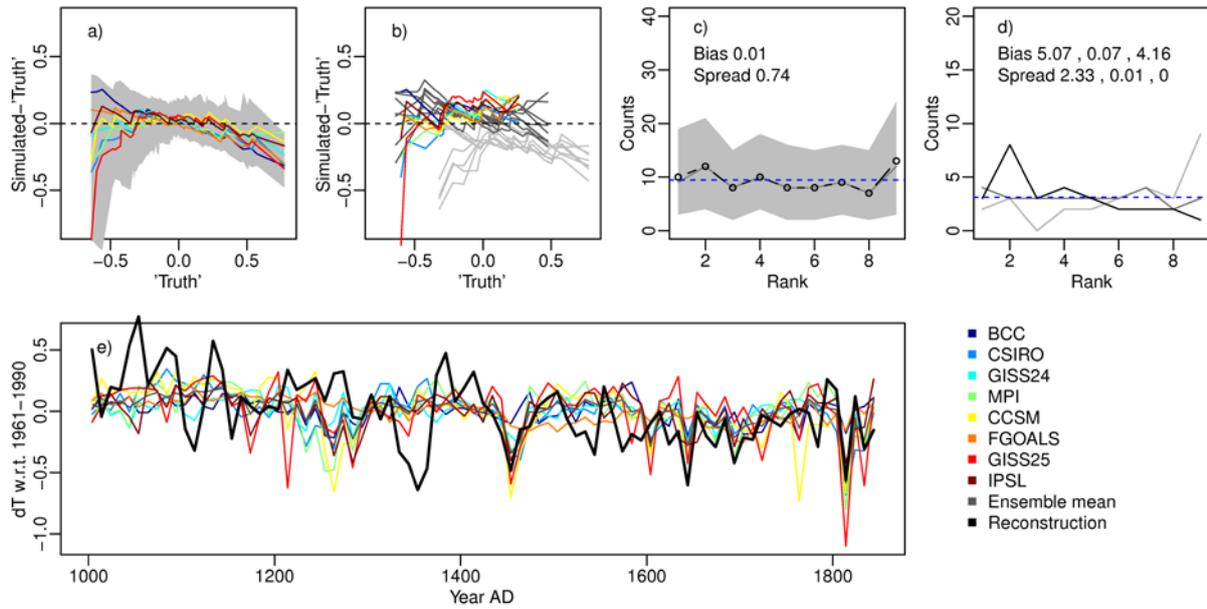


Figure S14: As Figure S9 but for the reconstruction by Mann et al. (2008) recalibrated by Frank et al. (2010) to the period 1950 to 1990.

## References

- Bothe, O., Jungclauss, J. H., Zanchettin, D., and Zorita, E.: Climate of the last millennium: ensemble consistency of simulations and reconstructions, *Clim. Past*, 9, 1089–1110, 10.5194/cp-9-1089-2013, 2013.
- Crowley, T. J., Zielinski, G., Vinther, B., Udisti, R., Kreutz, K., Cole-Dai, J., and Castellano, E.: Volcanism and the little ice age, *Pages News*, 16, 22–23, 2008.
- Frank, D., Esper, J., and Cook, E. R.: Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophys. Res. Lett.*, 34, L16709, doi:10.1029/2007GL030571, 2007.
- Frank, D. C., Esper, J., Raible, C. C., Büntgen, U., Trouet, V., Stocker, B., and Joos, F.: Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, *Nature*, 463, 527–530, doi:10.1038/nature08769, 2010.
- Gao, C., Robock, A., and Ammann, C.: 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:http://dx.doi.org/10.1029/2008JD01023910.1029/2008JD010239, 2008.
- Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures, *Holocene*, 8, 455–471, doi:10.1191/095968398667194956, 1998.
- Ljungqvist, F. C.: A new reconstructions of temperature variability in the extra-tropical northern hemisphere during the last two millennia. *Geografiska Annaler: Series A, Physical Geography*, 92: 339, doi:351. 10.1111/j.1468-0459.2010.00399.x, 2010.
- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *P. Natl. Acad. Sci.*, 105, 13252–13257, doi:10.1073/pnas.0805721105, 2008.

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