

Supplement of *Clim. Past*, 11, 1673–1699, 2015
<http://www.clim-past.net/11/1673/2015/>
doi:10.5194/cp-11-1673-2015-supplement
© Author(s) 2015. CC Attribution 3.0 License.



Climate
of the Past

Open Access

The EGU logo features the letters 'EGU' in a bold, sans-serif font, with a stylized gear or circular arrow element behind the 'G'.

Supplement of

Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium

Correspondence to: H. Goosse

The copyright of individual parts of the supplement might differ from the CC-BY 3.0 licence.

1 ***Sect. S1: processing of the model time series***

2 Model data were processed to correspond to the temporal and area-weighted spatial
3 averaging of the PAGES 2k temperature reconstructions (PAGES 2k Consortium, 2013) for
4 each region as follows:

- 5 1. Antarctica (annual): 90° S-60° S ; 180° W-180° E
- 6 2. Arctic (annual): 60° N-90° N ; 180° W-180° E
- 7 3. Asia (June-August): 23.5° N-55° N ; 60° E-160° E
- 8 4. Australasia (September-February): 50° S-0° S ; 110° E-180° E
- 9 5. Europe (June-August): 35° N:70° N ; 10° W-40° E
- 10 6. North America (annual)*: 30° N:55° N ; 130° W-75° W
- 11 7. South America (December-February): 65° S-20° S ; 75° W-30° W

12 Note that all simulations for North America were bilinearly interpolated to the
13 HadCRUT3V latitude-longitude grid before the grid box centred at 52.5° N; 77.5° W was
14 removed from the averaging calculation to match processing of the instrumental predictand
15 (PAGES 2k Consortium, 2013).

16 The ocean was masked for Antarctica, Asia, Europe and South America regions using
17 the respective binary (1, 0) land masks for each simulation. No significant difference was
18 found between the use of fractional land masking (proportion between and including 0 and 1)
19 and the binary mask common to all simulations for the land-only PAGES 2k reconstruction
20 equivalent.

21 As the GISS-E2-R control simulation is known to contain a drift from non-equilibrated
22 initial conditions (Schmidt et al., 2014), the transient as well as the control simulation from
23 that model have been detrended by subtracting a low-frequency loess fit that has been
24 estimated from the corresponding time period of the control simulation from each time series.

25

26 ***Sect. S2: specific implementation of some methods***

27 **S2.1 Probabilistic and climatological consistency**

28 The two concepts of probabilistic and climatological consistency can be seen as two
29 alternative ways to evaluate biases and spreads in simulations and reconstructions. In the
30 current study, however, the analyses use temperature anomalies from long-term averages, and
31 hence the bias is always zero by construction. Thus, our analysis mainly assesses two different
32 aspects of spread in the distributions of the regional temperature reconstructions and climate
33 model simulations.

34 The quantile-quantile (r-q-q) plots displayed to analyse the climatological consistency
35 show the difference between the simulated and the target quantiles. Residuals should
36 approach zero for a consistent simulation (a flat line in the plots). Offsets relative to $y=0$ on
37 the quantile-quantile plot indicate biases between the simulation and the target. Slopes in the
38 residuals indicate underestimation or overestimation of the variance, i.e. excessively narrow
39 or wide distributions. We refer to such cases as being under- or over-dispersive. Negative
40 slopes occur if the simulated variance is smaller than that of the target and positive ones for
41 larger simulated variance. If a simulation and a reconstruction are consistent, the difference in
42 their quantiles should be close to zero. Consequently, the plot should be approximately flat. If
43 for low values of the reconstruction the residual quantile simulated minus reconstruction are
44 always negative and always positive for positive value of the reconstructions (positive slope),

45 the simulated ensemble is too broad and thus the simulations have larger simulated variance
46 compared to the reconstruction.

47 To assess probabilistic consistency, we test whether the occurrence frequencies of the
48 simulation ensemble agree with those of the verification target, within limits of uncertainty. At
49 each time step, we identify the rank of the temperature reconstructions within the set formed
50 by the combination of the simulation ensemble and those temperature reconstructions
51 (Anderson, 1996). Flatness of the histograms is thus a necessary condition for our simulation
52 ensemble to be considered as a reliable representation of the target. The histograms visually
53 highlight biases (meaning here an offset in mean between the target and the ensemble) and
54 differences in ensemble variance. Over-dispersion (ensembles that are too wide) and under-
55 dispersion (ensembles that are too narrow) are identified by dome- or U-shaped histograms,
56 respectively. Such shapes imply that the target data are too often close to the central rank or
57 too often on the outer ranks (i.e., far from the mean of the ensemble of simulations). Slopes in
58 the histograms reveal biases, with positive (negative) slopes suggesting the target data are
59 ranked high (low) too often.

60 S2.2. Superposed epoch analysis

61 The response to volcanic aerosol forcing is evaluated at interannual and multidecadal
62 time scales for two different external forcing estimates (Gao et al., 2008; Crowley and
63 Unterman, 2013) that have been used as last-millennium boundary conditions in the PMIP3-
64 CMIP5 simulations (Schmidt et al., 2011, 2012).

65 The volcanic composite at interannual timescales is generated by first selecting the 12
66 strongest volcanic events. The mean from 5 years before to 10 years after the date of the peak
67 eruption is then computed for the forcing sequence as well as for the simulated and
68 reconstructed temperature sequences. In the case of the multidecadal composites, the time
69 series are first filtered with a 40-year low pass filter using least-squares coefficients
70 (Bloomfield, 1976). For the multidecadal composites, the 5 strongest events are selected and
71 the means from 40 years before to 40 years after the eruption are calculated, following
72 Masson-Delmotte et al. (2013). All the events are individually selected for each of the PAGES
73 2k regions making use of the latitudinal discretization of the volcanic forcing.

74 The composites for the strongest multidecadal changes in the solar forcing are based on
75 low solar forcing periods selected to be the same as in Fig. 5.8 of the IPCC AR5 (Masson-
76 Delmotte et al., 2013) for the sake of a better comparison. This corresponds to seven 80-year
77 time windows centred on the years 1044, 1177, 1451, 1539, 1673, 1801 and 1905.

78 S2.3 Framework for evaluation of climate model simulations: U_R and U_T statistics

79 The statistical model underlying the framework developed by Sundberg et al. (2012,
80 henceforth SUN12), Hind et al. (2012) and Moberg et al. (2015) has similar components to
81 the one used in detection and attribution studies (see section 5.3), but there are some
82 differences. An important similarity is the idea that temperature variations can be expressed as
83 a sum of forced and unforced variability. The two frameworks explicitly distinguish internal
84 variability in simulations and in observations, which can consist of instrumental observations
85 or, as in this study, proxy-based climate reconstructions.

86 The SUN12 framework also explicitly accounts for error variance in the observations,
87 such as non-climatic noise in proxy data. It even allows this type of error to vary with time, if

88 such information is available. Despite similarities in the underlying assumptions, the main
89 purposes of the SUN12 and detection and attribution approaches differ. While detection and
90 attribution studies seek to identify the forced response in observations, the SUN12 framework
91 was developed as a tool for evaluating forced simulations, with the aim of testing if one
92 simulation significantly fits observations better than another simulation or to rank a set of
93 plausible simulations. In the current study, this framework is mainly used to investigate the
94 common behaviour of all simulations by means of how well they agree with the different
95 regional reconstructions.

96 U_R and U_T are calculated here for each forced simulation, using PAGES 2k regional
97 temperature reconstructions as the observational basis and a time resolution of non-
98 overlapping 15-year averages. Three types of calculations have been done: separately for each
99 region, combining information from all seven regions and combining regions only within each
100 hemisphere, using equal regional weights (see Moberg et al., 2015). Whenever a certain
101 control simulation is not sufficiently long, its data sequence is extended by repetition and
102 concatenation. For the COSMOS ensembles with high and low solar forcing, and for the GISS
103 ensemble, metrics are calculated for each individual simulation and for the entire ensembles,
104 following Moberg et al. (2015).

105 The statistical framework by Sundberg et al. (2012) requires that all proxy-based
106 temperature reconstruction time series are re-calibrated against instrumental records to suit
107 certain assumptions. Therefore, such a re-calibration was done here, but note that this is
108 specific for the calculation of U_R and U_T statistics and is not applied for any other diagnostics.

109 To obtain appropriate calibration target data series, gridded instrumental temperature
110 data were used and averaged over exactly the same regions and seasons as explained in Sect.
111 S1 for the models. To comply with the different boundaries and land/sea masks used, the
112 respective instrumental series were derived from CRUTEM4 (Jones et al., 2012) for regions 5
113 and 6, HadCRUT4 (Morice et al., 2012) for regions 2, 4 and 6 and CRU TS3 (Harris et al.,
114 2014; as updated and available on the KNMI Climate Explorer, <http://climexp.knmi.nl/>, on
115 April 23, 2014) for region 3. For region 1, we used the same instrumental target series as the
116 PAGES 2k Consortium (2013). Re-calibration was made for the same calibration periods as
117 used by the PAGES 2k Consortium. Each instrumental target series was arbitrarily assumed to
118 contain 10% noise variance. Sensitivity experiments were also made with assumptions of 5%
119 and 15% noise. This had no effect on any main conclusions.

120 For simplicity, it is assumed here that each proxy record has the same statistical
121 precision over its entire length, despite the fact that their precision typically decreases back in
122 time as the number of contributing local proxy series decreases. Therefore, the derived
123 measures are only approximate values, but a more accurate treatment would require detailed
124 work far beyond the scope of this study. As for some other methods in this study, the U_R and
125 U_T analysis uses anomalies from long-term averages to avoid systematic climatological bias
126 influencing the results.

127 S2.4 Detection and attribution

128 Detection and attribution techniques provide an estimate of the magnitude of the
129 forced response in a reconstruction, with an uncertainty estimate. These techniques can be
130 used to determine the relative contribution by different forcings simultaneously to a period or
131 climatic event, with uncertainty estimates reflecting if the contribution of different forcings

132 can be separated from each other and from climate variability (see Bindoff et al., 2013; Hegerl
133 and Zwiers, 2011). To estimate the different contributions from several individual forcings, it
134 is necessary to have access to separately forced simulations with each individual forcing. This
135 was not possible in this study however, because we are only using models driven by all
136 forcings together; hence we focus here on estimating the magnitude of the overall forced
137 response.

138 Detection and attribution studies rely on a multiple regression of reconstructions onto
139 the response expected by different individual contributing forcings. This assumes that climate
140 models approximately capture the response to individual forcings in shape (e.g., pattern in
141 time or spatial pattern of the response), but may misrepresent the magnitude of the overall
142 response. This is a reasonable assumption since the magnitude of the response to forcings is
143 affected by uncertainty in the transient climate sensitivity. Moreover, the magnitude of
144 forcings itself is also often uncertain, such as for the low-frequency component of solar
145 forcing (see e.g., Schmidt et al., 2011, 2012). A difficulty in the application of detection and
146 attribution methods to the last millennium is accounting for uncertainty in both
147 reconstructions and forcings. This can be addressed to some extent by using multiple
148 reconstructions and forcing estimates (e.g., Schurer et al., 2014), but a more systematic
149 approach is desirable.

150 The detection and attribution framework applied here has been extensively used for
151 instrumental data (Bindoff et al., 2013) and to some extent for paleoclimatic reconstructions
152 (see Hegerl et al., 2007; Schurer et al., 2014). This approach calculates a possible scaling
153 range for the response to the external forcing in the reconstruction (equation S1) based on
154 total least squares regression (Allen and Stott, 2003):

$$155 \quad Y(t) = \sum_{i=1}^m \beta_i (X_i(t) - \gamma_i(t)) + \gamma_0(t) \quad (S1)$$

156 where Y , the reconstructed temperature, is equal to a linear combination of m different model
157 fingerprints X_i (where m in this analysis is always equal to 1 as only the response to all the
158 forcings together is analysed here) multiplied by a scaling factor β_i . Each model simulation
159 has associated internal variability γ_i and the reconstructions contain a realization of internal
160 variability γ_0 . The scaling factors β_i determine the amplitude of the fingerprints in the
161 reconstructions. A range of scaling factors is obtained using samples of internal variability
162 taken from model simulations. A forcing is said to be detected if a scaling value of zero is
163 rejected at some significance level, for example, the 5% level. To evaluate the self-
164 consistency of the regression results, the residual of the fit is checked against estimates of
165 model-based internal variability. This is the same method as used in Schurer et al. (2013).

166

167 ***Sect. S3: correlation between simulated and reconstructed time series.***

168 Figure S1 displays the correlation between the 23-year Hamming filtered model
169 simulation results and temperature reconstructions for individual regions. This illustrates the
170 agreement between the contribution of radiative forcing on observed temperatures and in the
171 model simulations. The highest correlation values are obtained for the Arctic region in most
172 simulations (Figure S1a). Correlations for the North American pollen-based reconstruction
173 and for Australasia and Europe tend to be highly significant. Correlations tend to be non-

174 significant for the North American tree reconstruction, and for the South American and
175 Antarctic reconstructions.

176 If we consider the available single-model ensembles (Figure S1b; COSMOS and GISS),
177 the correlation of the ensemble mean with the regional temperature reconstructions is always
178 higher than the average of all individual member correlations. The ensemble averaging
179 reduces the internal variability present in the simulated series in favour of the response to the
180 external forcing common in simulation results and reconstructed temperature.

181

182 ***Sect. S4: EOF analysis for GISS and COSMOS ensembles.***

183 Investigating the variability in the GISS and the COSMOS ensemble simulations
184 provides insights into the intra-model spread. The (detrended) GISS simulations show a very
185 coherent picture with similar loadings and variance explained by the leading EOF (~80-90 %)
186 for the different regions within the single ensemble members (Fig. S7). However, the
187 COSMOS simulations have a larger spread of the variance explained by the leading EOF. The
188 larger heterogeneity in the COSMOS simulations might be indicative of a larger amount of
189 internal variability and hence less externally forced spatial coherence among the regions. For
190 the ensemble with the larger scaling of the solar forcing (COSMOS high, Fig. S7) the amount
191 of variance represented by the leading EOF is larger compared to the weaker scaling
192 (COSMOS low, Fig. S7), indicating a larger common forced signal in the different ensemble
193 members.

194

195 ***Sect. S5: correlation between hemispheres.***

196 An analysis of coherence between hemispheric temperatures, calculated herein simply
197 by weighting the individual regions according to their area (Fig. S10), confirms the results of
198 Neukom et al. (2014) in the sense that the two hemispheres are significantly correlated during
199 most of the last millennium in model simulations. The control simulations indicate a natural
200 tendency for inter-hemispheric correlation in models. Nevertheless, the correlation during
201 periods with strong external forcing clearly exceeds the range derived from control
202 simulations (not shown). In the reconstructions, the Southern Hemisphere experienced
203 temperature anomalies opposite to the ones in the Northern Hemisphere during long periods
204 of the first half of the millennium, indicative of non-coherence between the two hemispheres
205 and potentially unforced variability.

206

207 ***Sect. S6: superposed epoch analysis for the Gao et al. (2008) forcing.***

208 The simulated response to the Gao et al. (2008) forcing has an amplitude of the order of
209 -1 to -0.5 °C in all regions (Fig. S11). This means that the simulated response is, as for the
210 Crowley and Untermann (2012) forcing (Fig. 8), larger than the reconstructed one, in
211 particular for Australia and Antarctica.

212

213

214

215 **Additional references**

216 Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H: Updated high-resolution grids of
217 monthly climatic observations – the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34, 623-642,
218 doi:10.1002/joc.3711. 2014

219

220 Jolliffe, I. T. and Primo, C.: Evaluating rank histograms using decompositions of the
221 chi-square test statistic, *Mon. Weather Rev.*,136, 2133–2139, doi: 10.1175/2007MWR2219.1,
222 2008.

223

224 Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.,
225 Hemispheric and large-scale land surface air temperature variations: An extensive revision
226 and an update to 2010, *J. Geophys. Res.*, 117, D05127, doi:10.1029/2011JD017139, 2012

227

228 Priestley, M. B. *Spectral analysis and time series*, Academic Press, 1982.

229

230

231

232 **Supplementary table S1: additional information on model simulation sources**

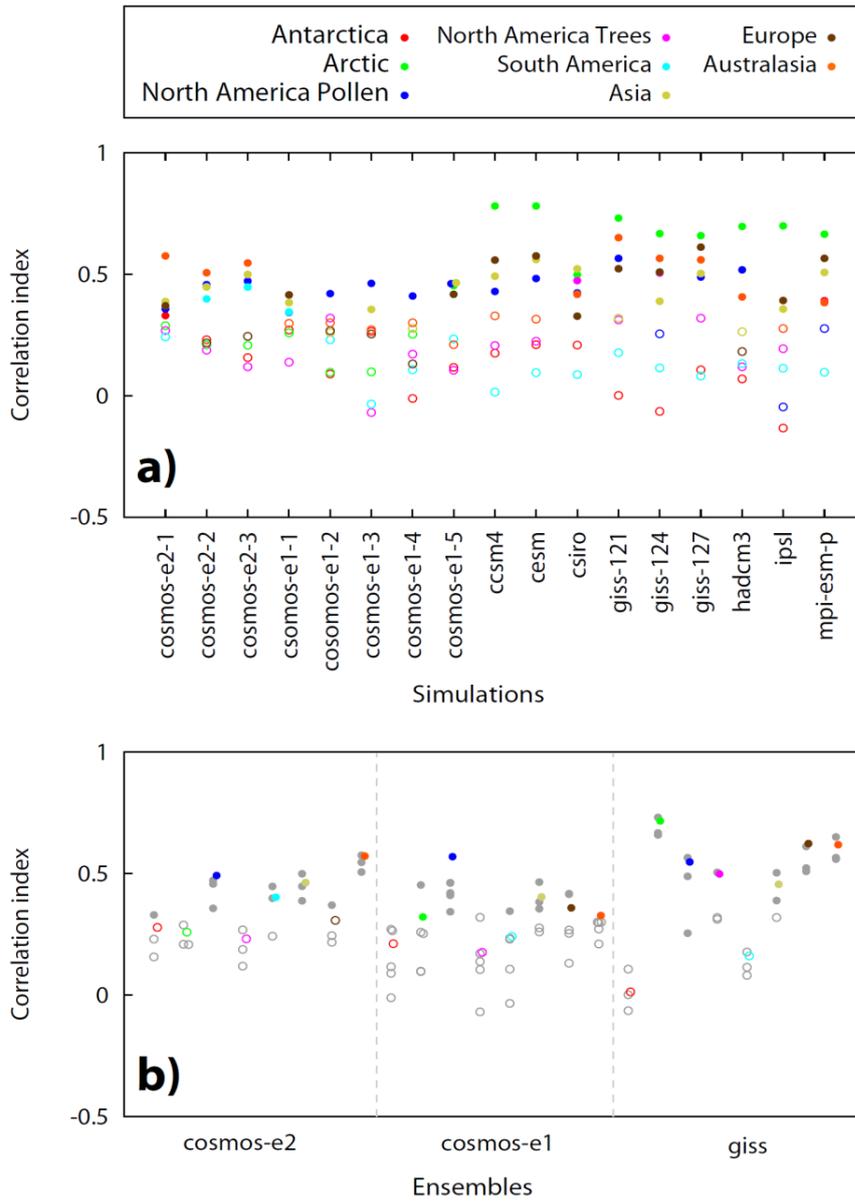
Model	past1000	historical extension	Source past1000	Source historical extension	Reference
CCSM4	r1i1p1	r1i2p2	PMIP3 ¹	PMIP3	Landrum et al. (2013)
CESM1	Single continuous simulation (not contained in the CMIP5/PMIP3 database)		Flavio Lehner (lehner@climate.unibe.ch)	Flavio Lehner (lehner@climate.unibe.ch)	Lehner et al. (submitted.)
CSIRO-Mk3L-1-2	r1i1p1	r1i1p1	PMIP3	PMIP3	Phipps et al. (2013)
GISS-E2-R	r1i1p12[1,4,7]	r1i1p12[1,4,7]	PMIP3	PMIP3	Schmidt et al. (2013)
HadCM3	r1i1p1	exists, but is not in CMIP5/PMIP3 database	PMIP3 ²	Andrew Schurer (aschurer@staffmail.ed.ac.uk)	Schurer et al. (2013)
IPSL-CM5A-LR	r1i1p1	exists, but is not in CMIP5/PMIP3 database	PMIP3	PMIP3	Dufresne et al. (2013)
MPI-ESM-P	r1i1p1	r1i1p1	PMIP3	PMIP3	Jungclaus et al. (2014)
ECHAM5/MPIOM	continuous simulations (pre-PMIP3, no rip code)		CERA ³	CERA	Jungclaus et al. (2010)

233

234 ¹ <http://pcmdi9.llnl.gov/esgf-web-fe>235 ² <http://badc.nerc.ac.uk/browse/badc/euroclim500/data/ALL/r1>236 ³ <http://cera-www.dkrz.de/CERA>

237

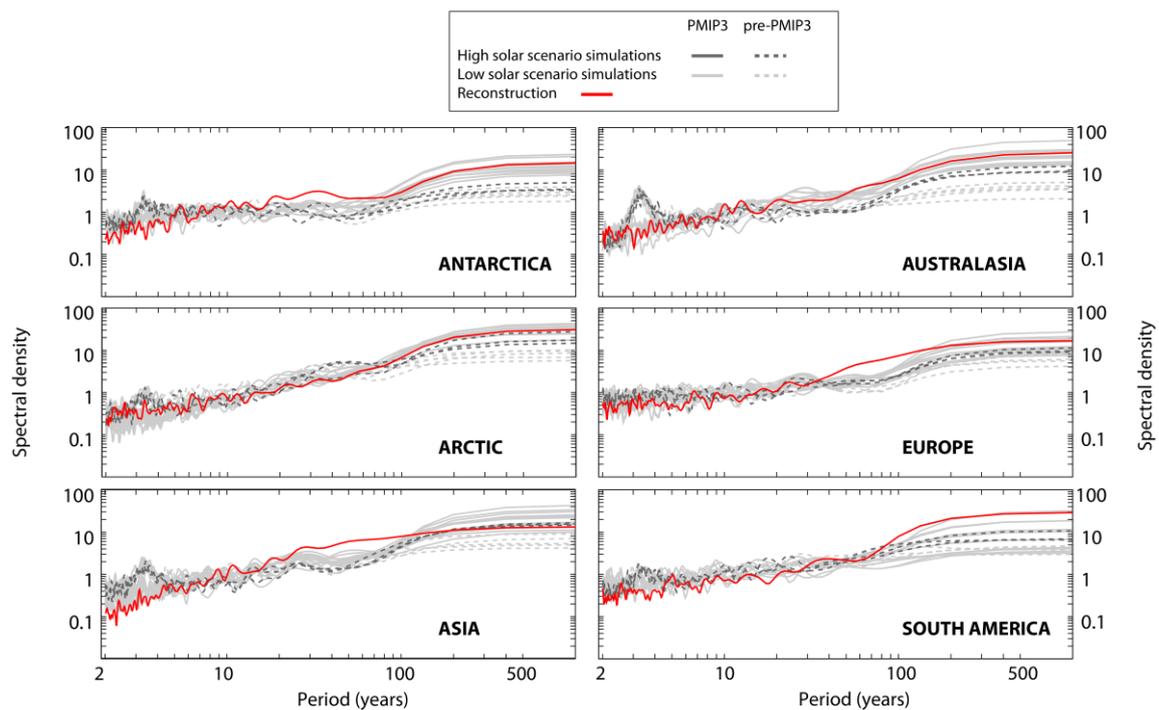
238



239

240 Figure S1. a) Correlations between 23-year Hamming filtered PAGES2k temperature
 241 reconstructions and climate model simulations. Dots represent the correlation between each
 242 regional reconstruction (see legend for region-colour) and the simulation averaged over the
 243 corresponding domain. Filled (unfilled) circles stand for significant (non-significant)
 244 correlation values. b) As in a), but focused on the models with ensembles of simulations;
 245 individual ensemble member correlations are shown in grey and the ensemble average in
 246 colour.

247

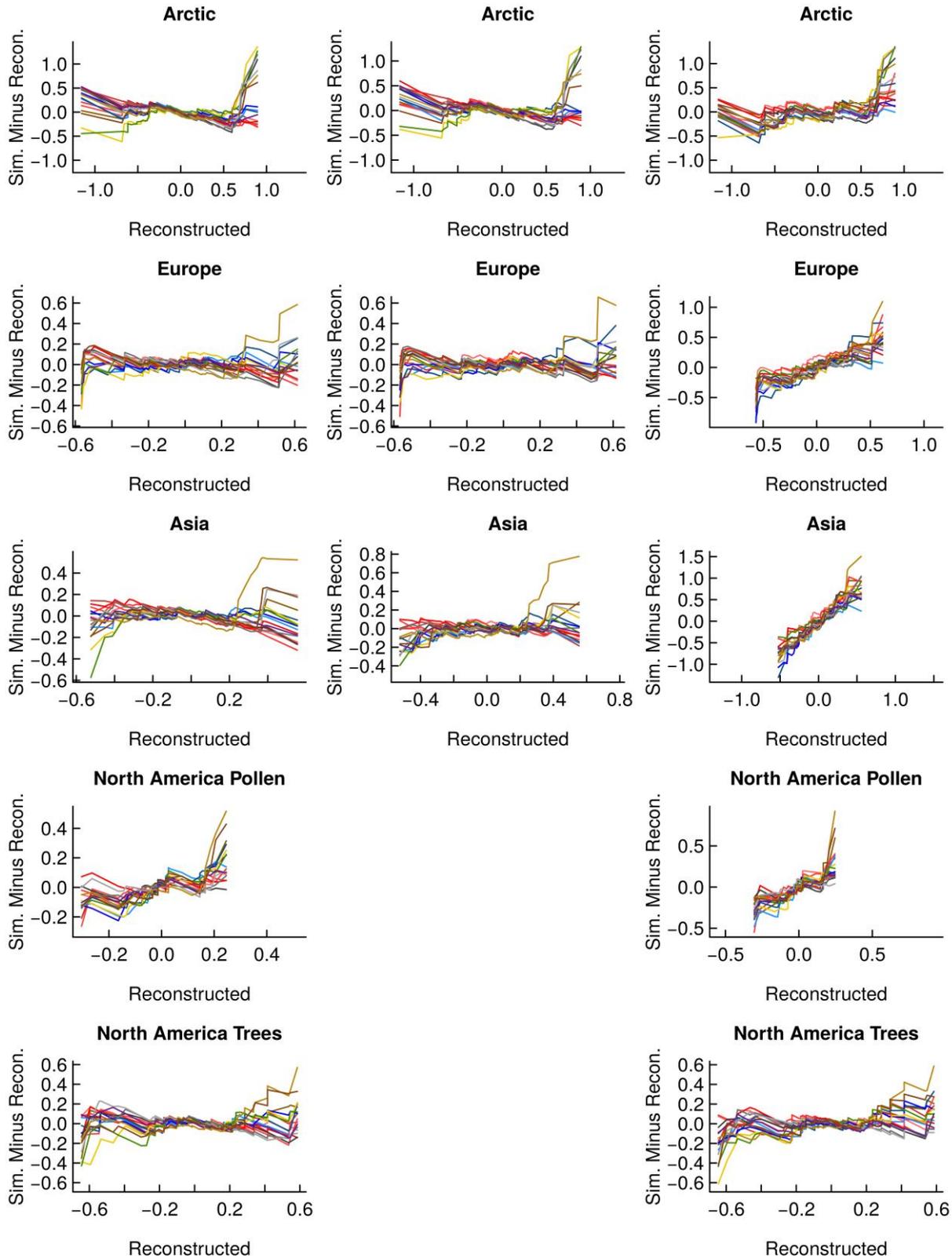


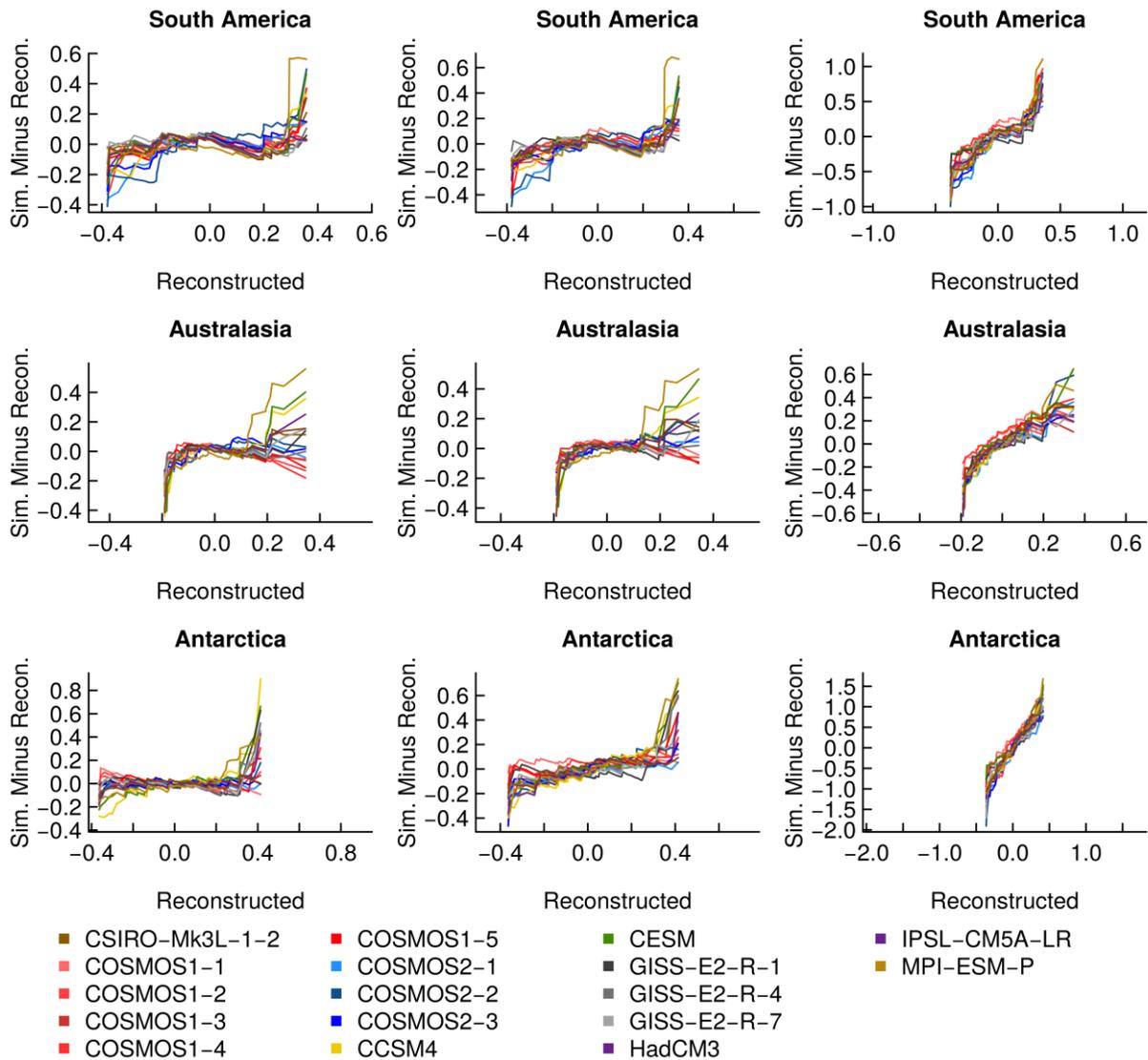
249

250 Figure S2. Normalized spectra of pre-PMIP3 (dashed) and PMIP3 (solid) simulations (grey)
 251 and reconstructions (red) for six PAGES 2k regional reconstructions for the period 850 to
 252 2000 CE. The spectra were computed from the normalized reconstructed and simulated
 253 regional temperatures using a 100 years Tukey-Hanning filter (Priestley, 1982). The
 254 simulations using solar forcing with higher (lower) variability are also highlighted in dark
 255 (light) grey.

256

257

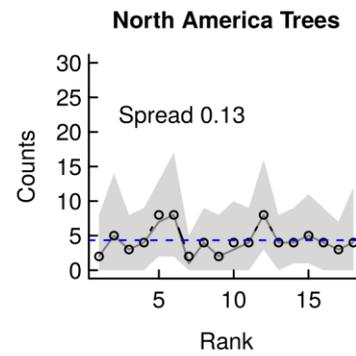
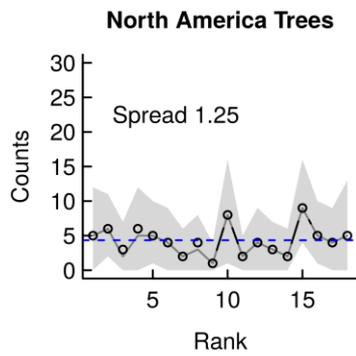
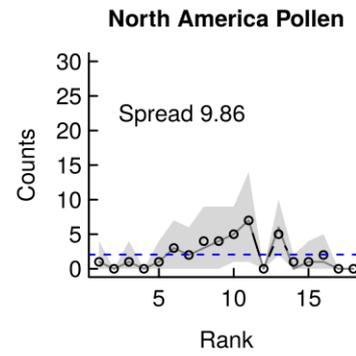
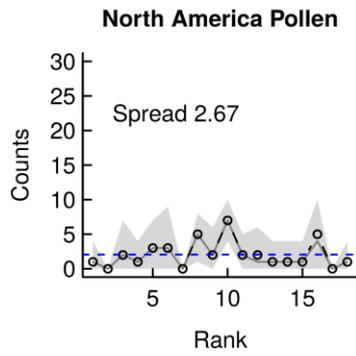
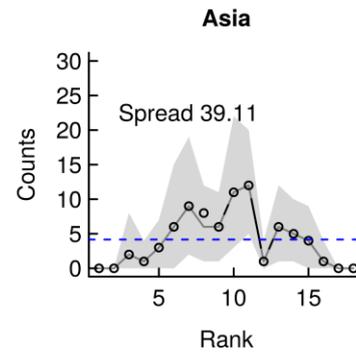
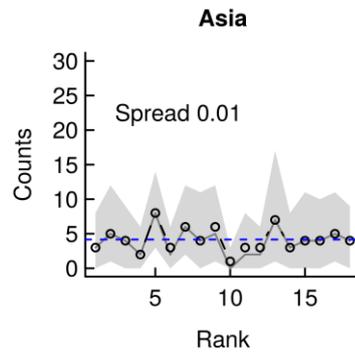
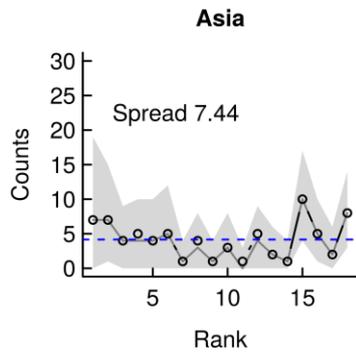
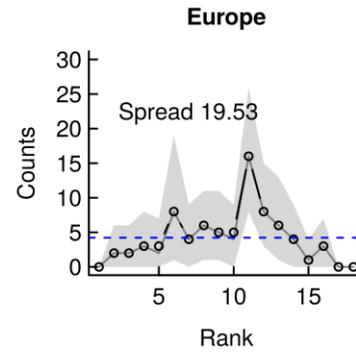
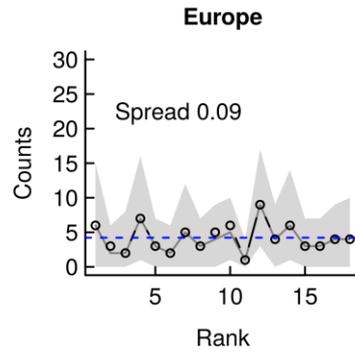
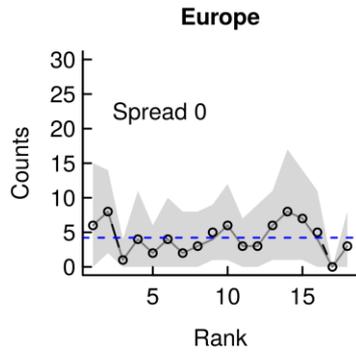
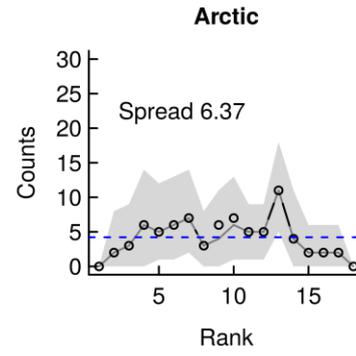
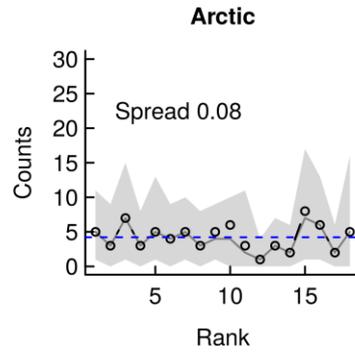
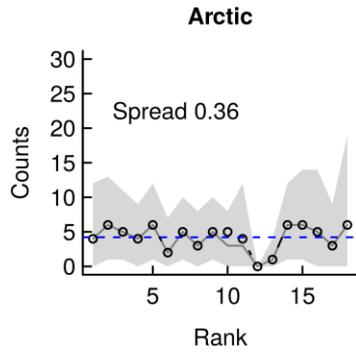


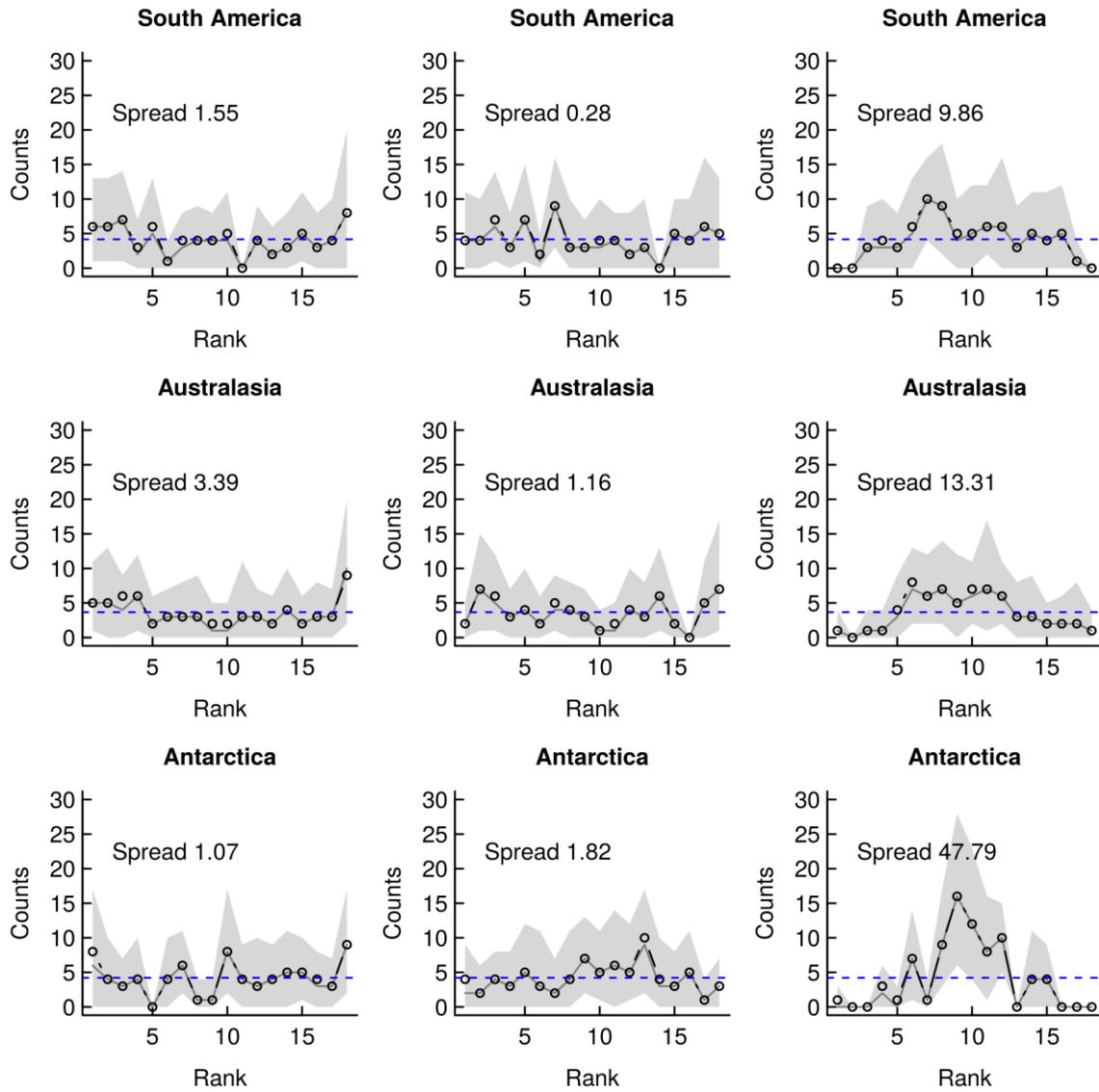


259

260 Figure S3. Climatological consistency: residual quantile-quantile plots for the full period for
 261 all the regions. In the left column, the uncertainty is neglected in the computations, in the
 262 middle column the original uncertainty divided by a factor $\sqrt{15}$ is used to take into account
 263 the smoothing while the original uncertainty is applied for the right column. There is no
 264 middle column for North American reconstructions because of their resolution. Positive and
 265 negative slopes or large differences from 0 emphasize lack of consistency.

266

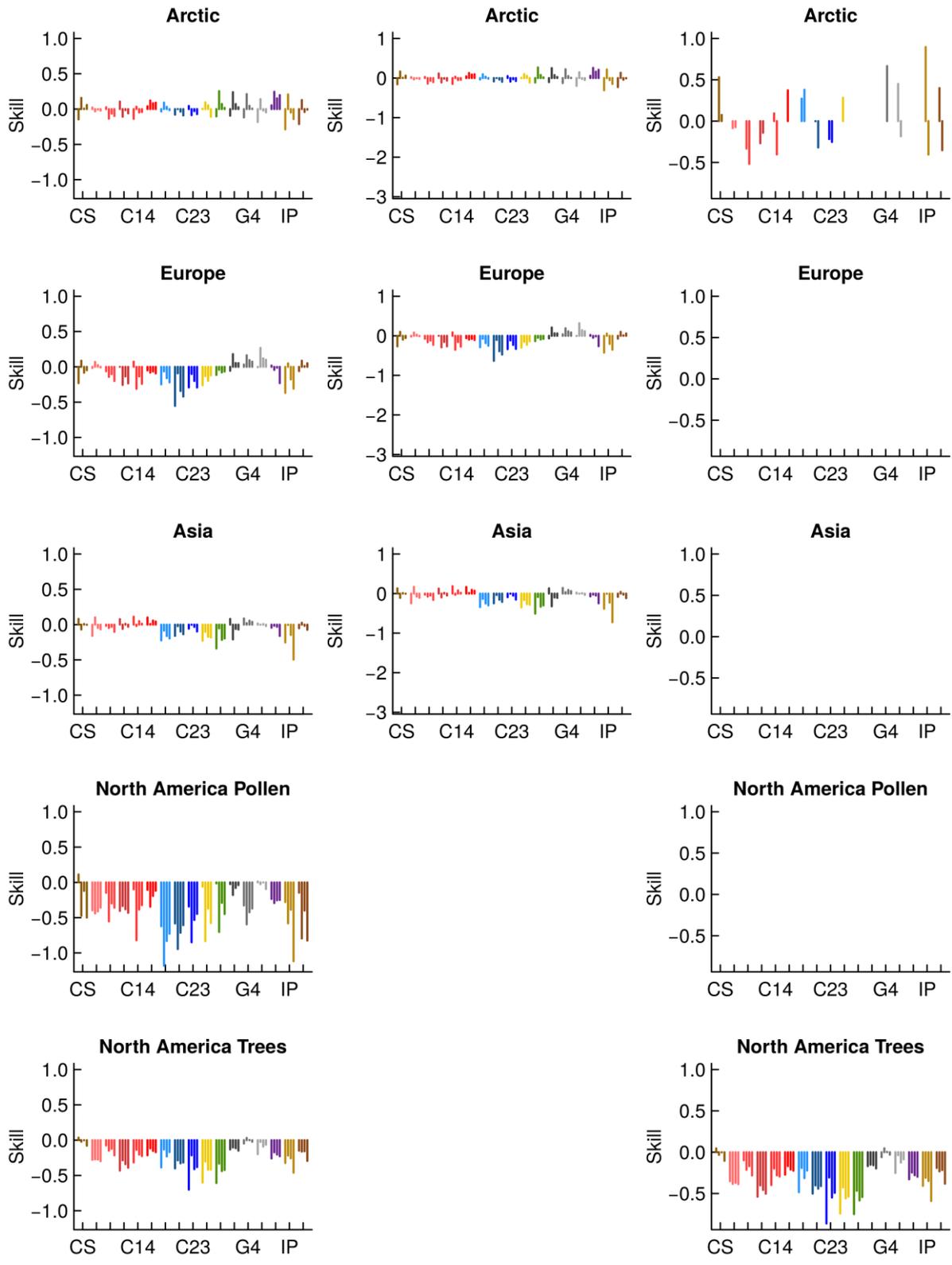


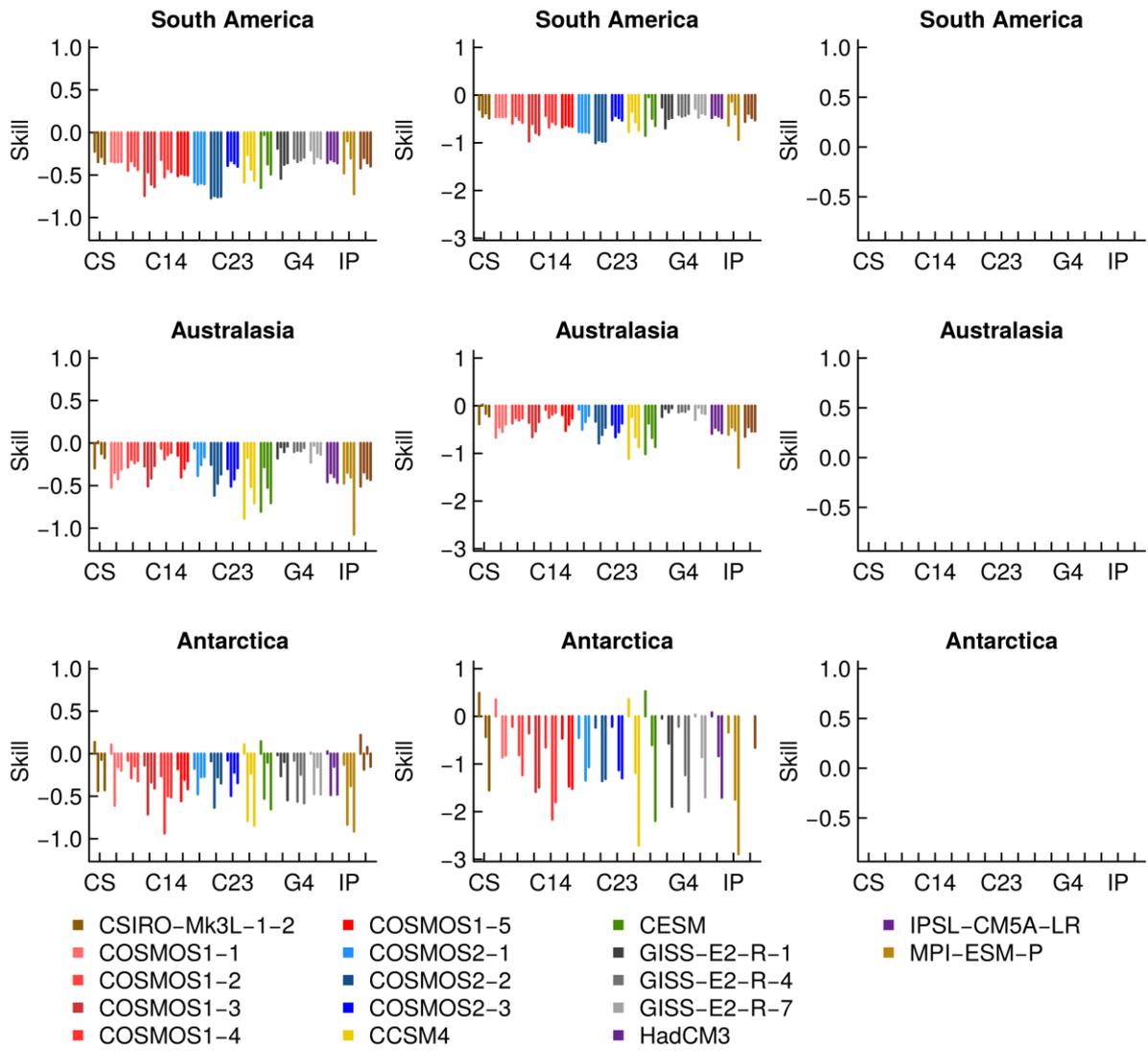


268

269 Figure S4. Probabilistic consistency for all the regions. The χ^2 goodness-of-fit statistic is
 270 applied to evaluate the consistency between observed rank count and the flat null hypothesis.
 271 The statistic can be decomposed to test for individual deviations like bias or spread (Jolliffe
 272 and Primo, 2008), as in Bothe et al. (2013a, b). In the left column, the uncertainty is neglected
 273 in the computations, in the middle column the original uncertainty divided by a factor $\sqrt{15}$ is
 274 used to take into account the smoothing while the original uncertainty is applied for the right
 275 column. There is no middle column for North American reconstructions because of their
 276 resolution. U- or dome-shaped features highlight lack of consistency.

277



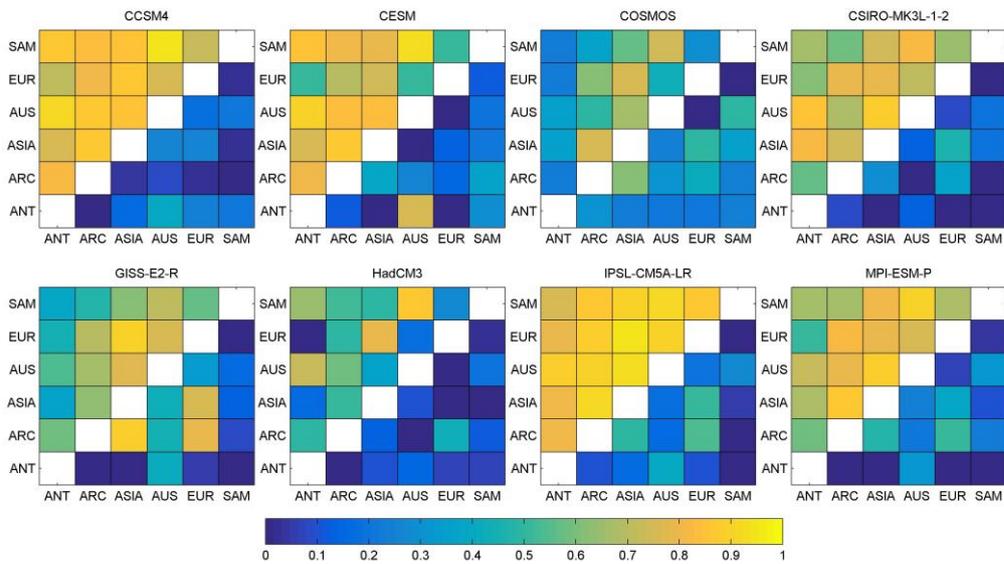


279

280 Figure S5. Skill metric for the individual models for all periods (bars from left to right: 850-
 281 1350, 1350-1850, 850-1850, 850-2000 CE). In the left column, the uncertainty is neglected in
 282 the computations, in the middle column the original uncertainty divided by a factor $\sqrt{15}$
 283 is used to take into account the smoothing while the original uncertainty is applied for the right
 284 column. There is no middle column for North American reconstructions because of their
 285 resolution. When the skill is undefined no bar is shown. Positive values indicate skill in this
 286 simple evaluation.

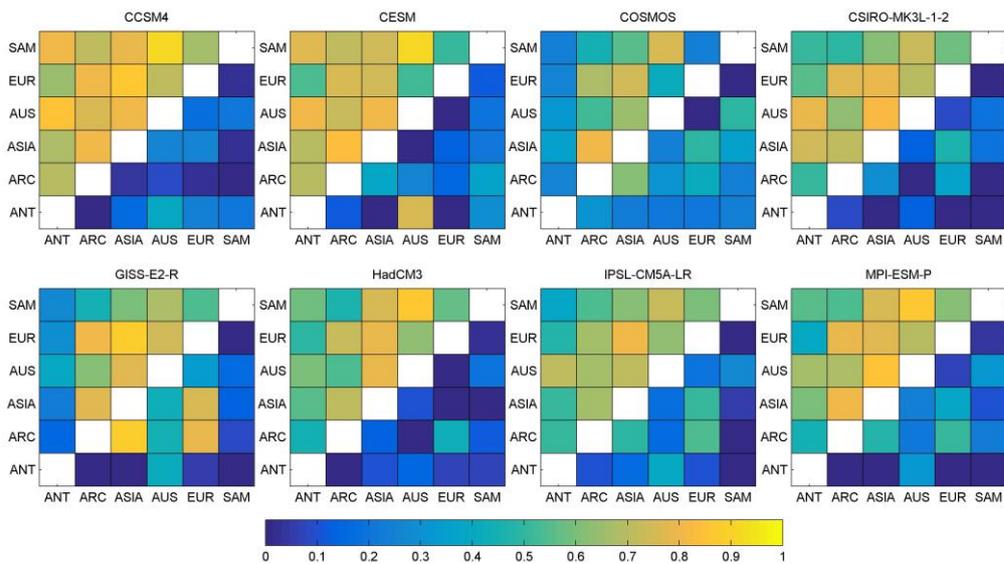
287

288 a) Correlation for the entire period



289

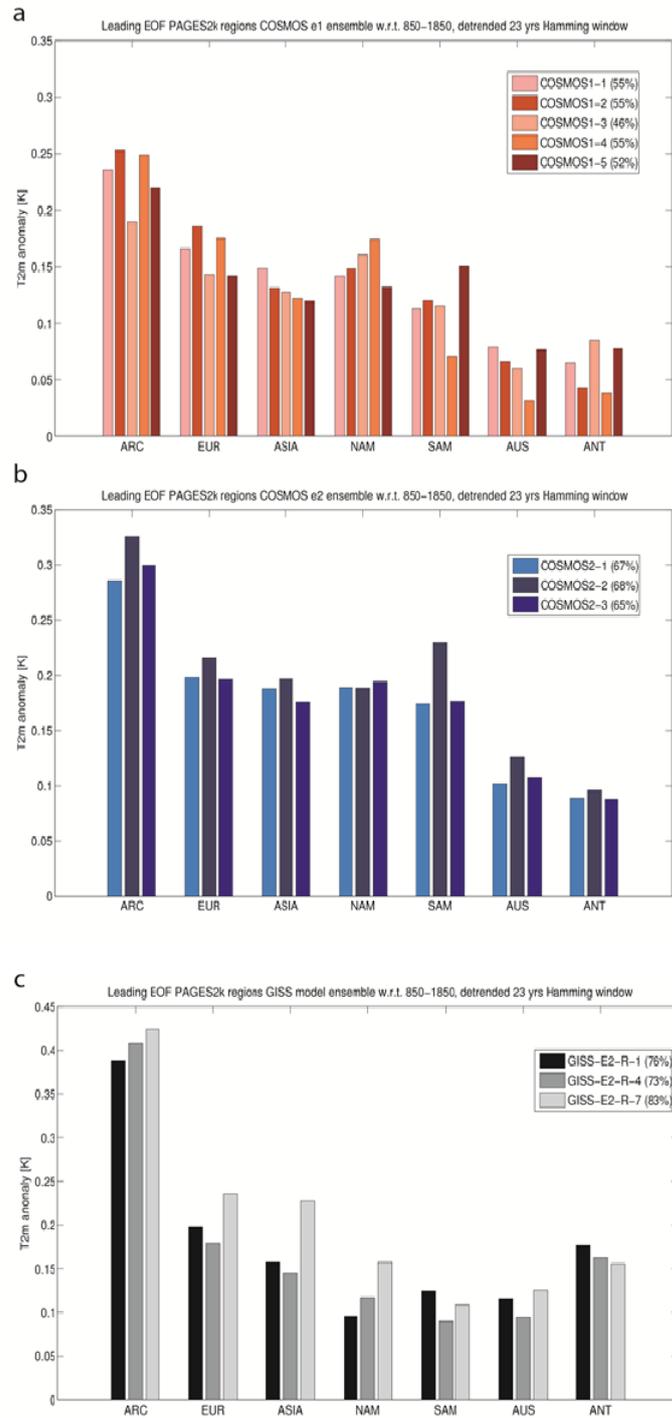
290 b) Correlation for the preindustrial period



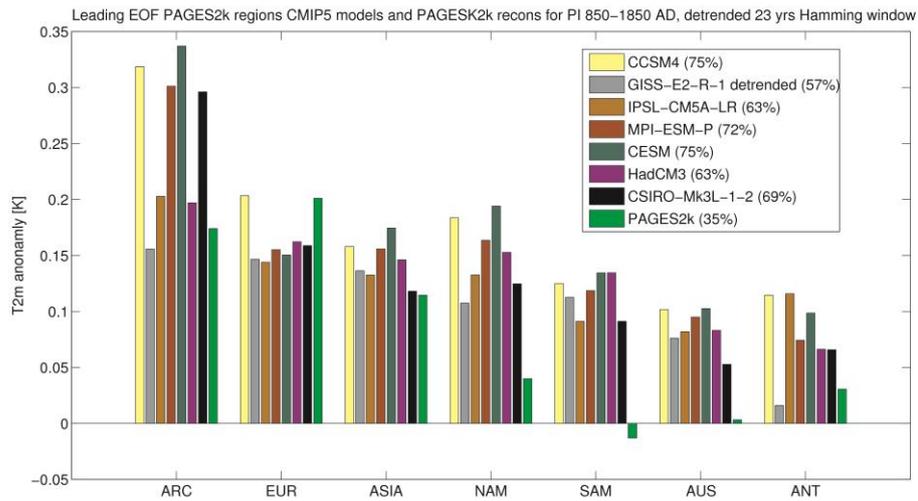
291

292 Figure S6. Correlations among the PAGES 2k regions for the different models using
 293 detrended time series filtered with a 23-year Hamming filter. a) full period 1012 CE – 1978
 294 CE, b) pre-industrial period 1012 CE -1850 CE. The upper left triangle represents the
 295 correlations for the forced simulations while the lower right triangle represents the
 296 correlations for the control runs (based on the full length of the control runs).

297



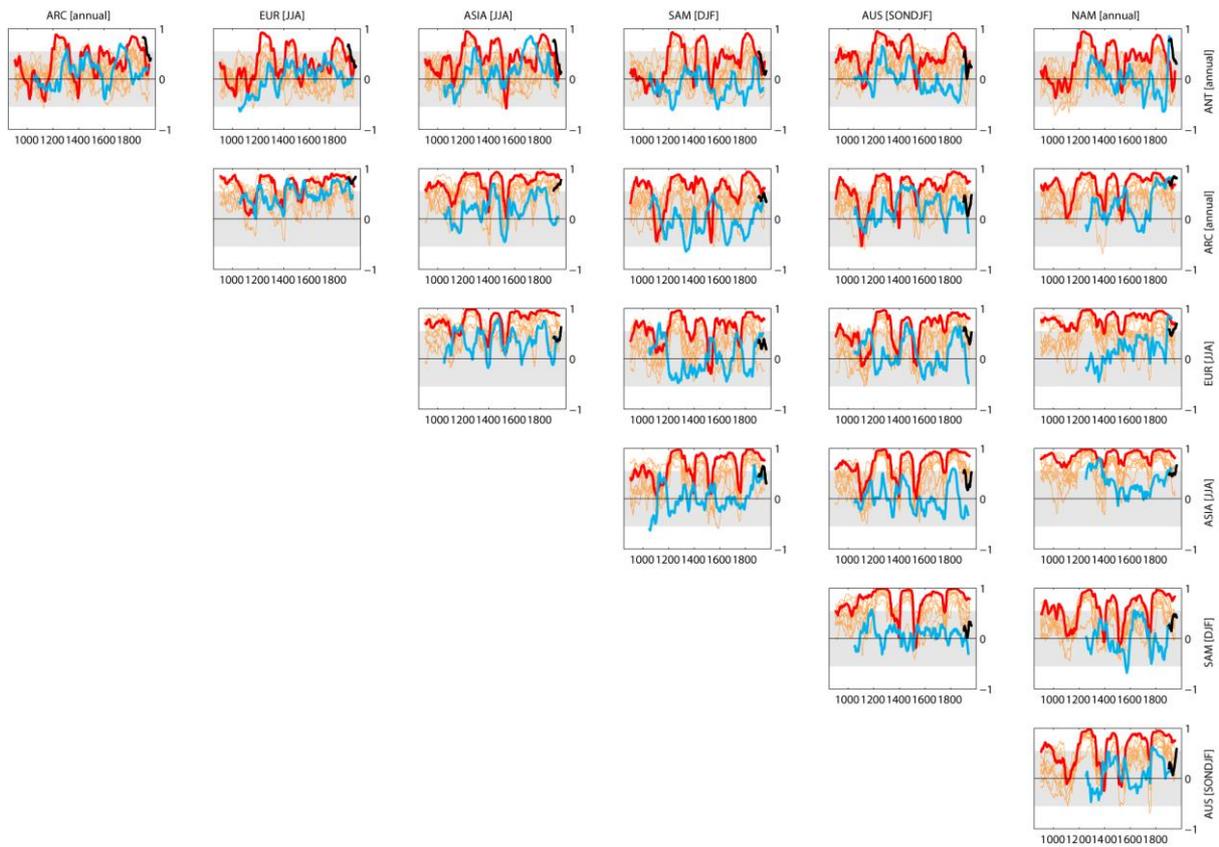
298
 299 Figure S7. Same as Figure 6a for the leading EOF of the COSMOS ensemble with low (a) and
 300 high (b) solar activity changes and the GISS ensemble (c) models over the period 850–2004
 301 AD. The figure shows the spread among the single members for those models with multiple
 302 realizations. The eigenvectors are based on the covariance matrix with respect to temperature
 303 anomalies for the period 850–1850. Values in parentheses relate to the amount of variance
 304 represented by the leading EOF. The time series were filtered with a 23-year Hamming filter
 305 and were linearly detrended afterwards. Within their specific experimental setup (COSMOS
 306 low, COSMOS high, GISS) the individual simulation members show similarities related to
 307 both the amplitude of the temperature anomalies and the variance represented by the leading
 308 EOF.



310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321

Figure S8. Leading EOFs of the near-surface temperature simulated by each CMIP5/PMIP3 model and in reconstructions over the period 850–1850 CE. The time series were filtered with a 23-year Hamming filter and were linearly detrended before the covariance matrix was calculated. Values in parentheses relate to the amount of variance represented by the leading EOF. The difference to Fig. 6a in the main text relates to a different basis for the calculation of the EOFs corresponding to the pre-industrial period. The pre-industrial EOF pattern is similar to Fig. 6a, albeit with differences in the amplitude of temperature anomalies in individual regions and the amount of variance represented by the leading EOF.

322



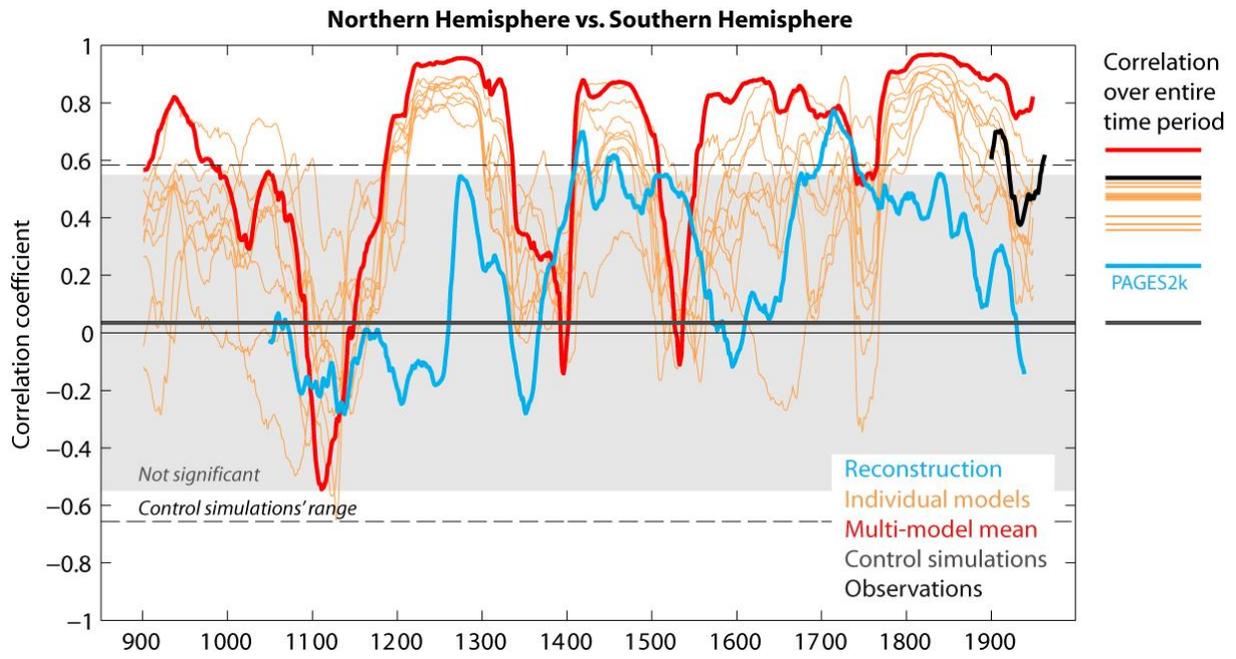
323

324

325 Figure S9. 100-year moving Tukey window correlations between all PAGES 2k regions for
326 the PAGES 2K reconstructions (blue) and PMIP3 models (8 models in orange, multi-model
327 mean in red) and observations from HadCRUT4 (black). Each 100-year segment is linearly
328 detrended beforehand. Grey shading illustrates not significant correlation at the 5% level.

329

330



331

332

333 Figure S10. 100-year moving Tukey window correlations between hemispheric averages for
 334 the PAGES 2k reconstructions (blue) and PMIP3 models (8 models in orange, multi-model
 335 mean in red) and observations from HadCRUT4 (black). Each 100-year segment is linearly
 336 detrended beforehand. Grey shading illustrates not significant correlation at the 5% level.

337

338

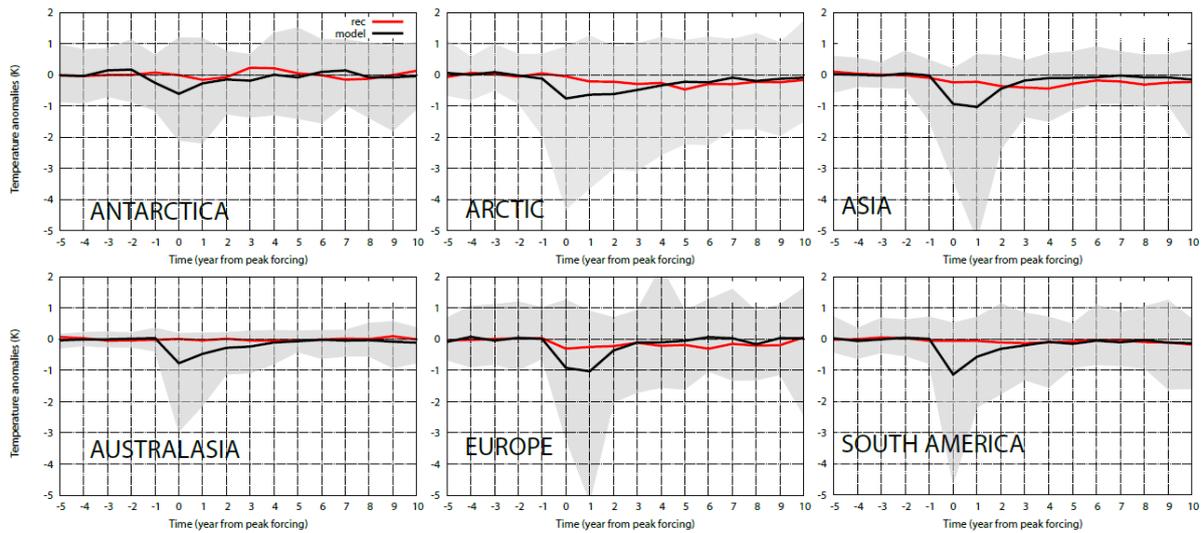
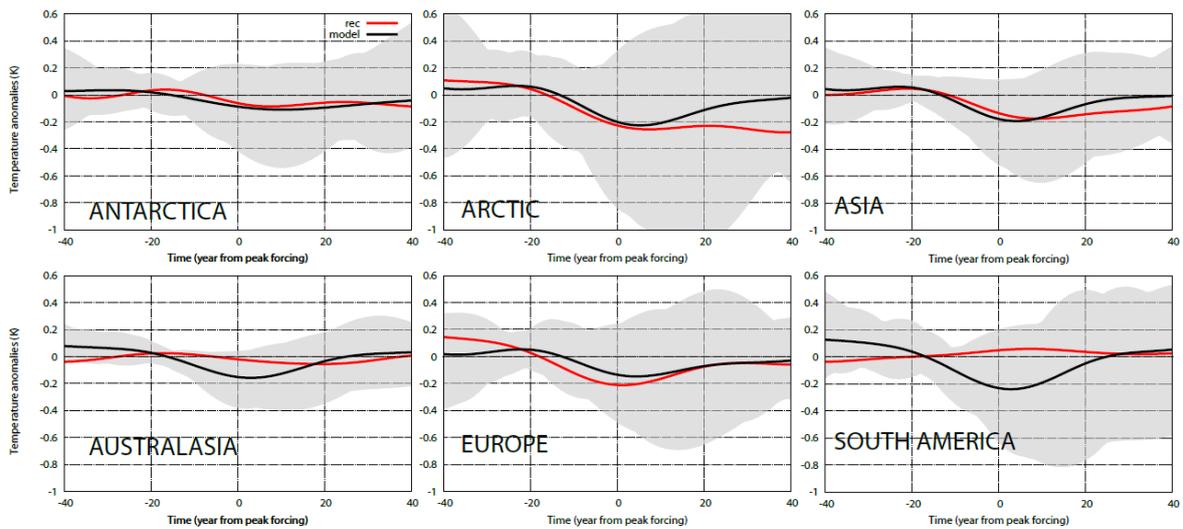


Figure S11. Same as Figure 8 but for the events selected in the Gao et al. (2008) reconstruction.

339

340



341 Figure S12. Superposed Epoch Analysis of the impact of the volcanic activity at multidecadal
342 timescales in the reconstructed and simulated temperatures. Superposed composites of
343 temperature responses during time intervals in which the years with peak negative forcing in
344 the Crowley and Unterman (2012) volcanic reconstruction are aligned. The composite is
345 produced by selecting the 5 strongest volcanic events, and a composite of the 30-year low
346 pass filtered temperature series from 40 years before to 40 years after the date of the peak
347 eruption. All the other elements are the same as in Figure 8.

348

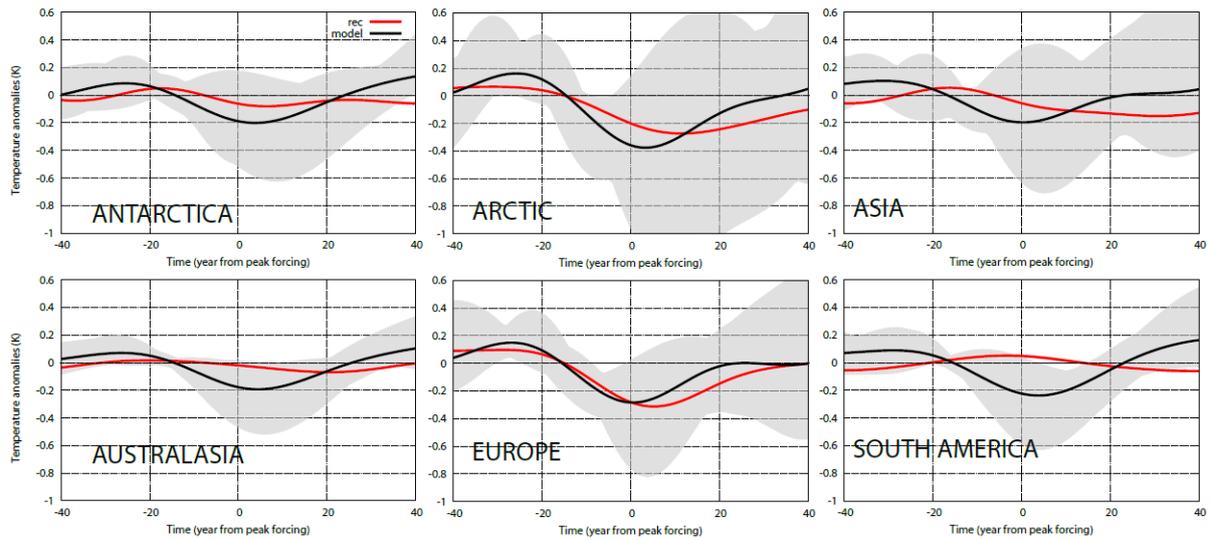


Figure S13. Same as Figure S12 but for the Gao et al. (2008) forcing.

349

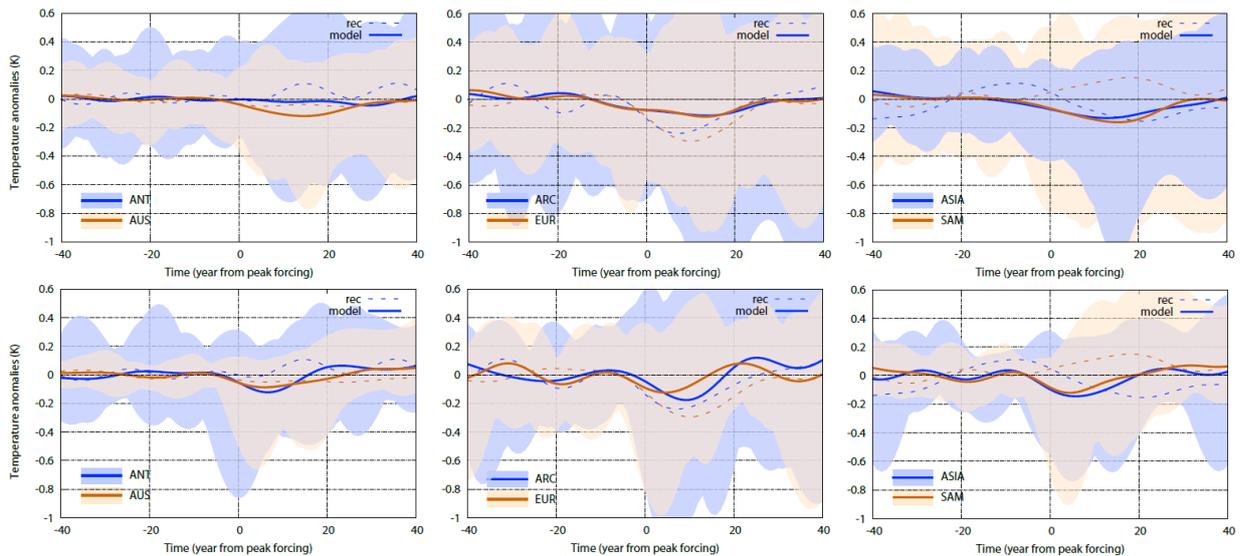
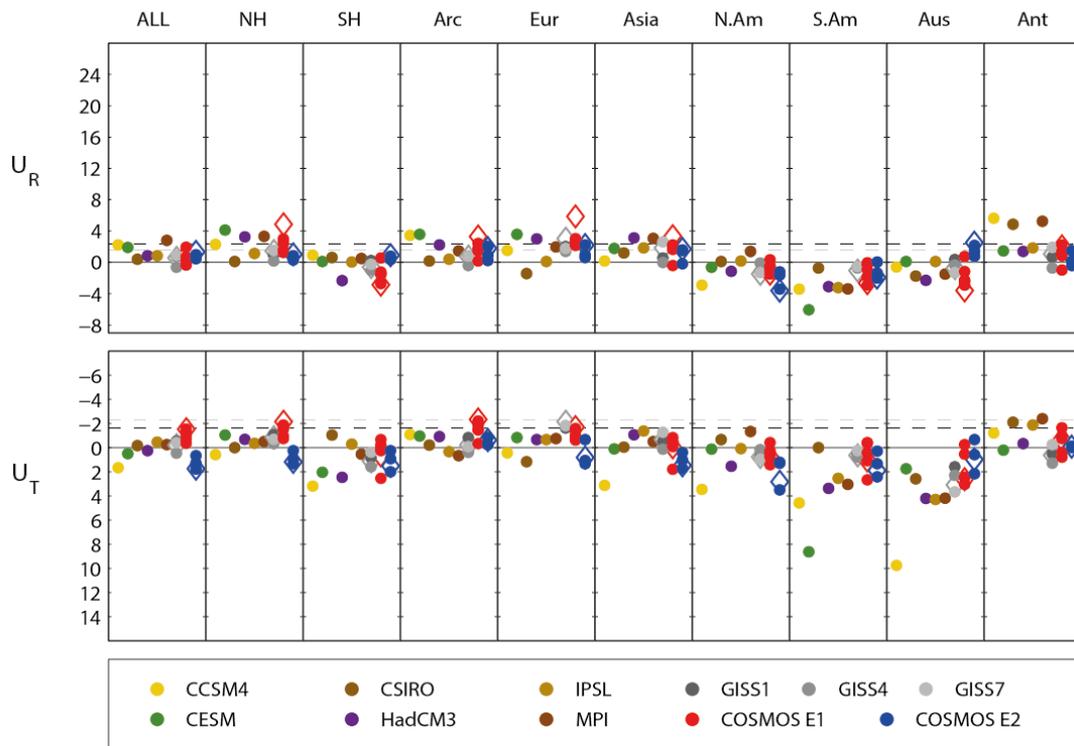


Figure S14. Superposed Epoch Analysis of the impact of the solar activity at multidecadal timescales in the reconstructed and simulated temperatures. Superposed composites of the temperature response during selected periods in which the solar forcing was lowest were performed (see text for details). Panels show results for reconstructions in six PAGES 2k regions and for model experiments performed using the volcanic forcing by: (top) the Crowley and Unterman (2012); and (bottom) Gao et al. (2008). Each panel indicates the reconstructed (dashed lines) and simulated (solid) composites of the temperature response for the same events for two different regions (see each panel for legend). The colour shading indicates the complete range of simulated temperature responses.

350

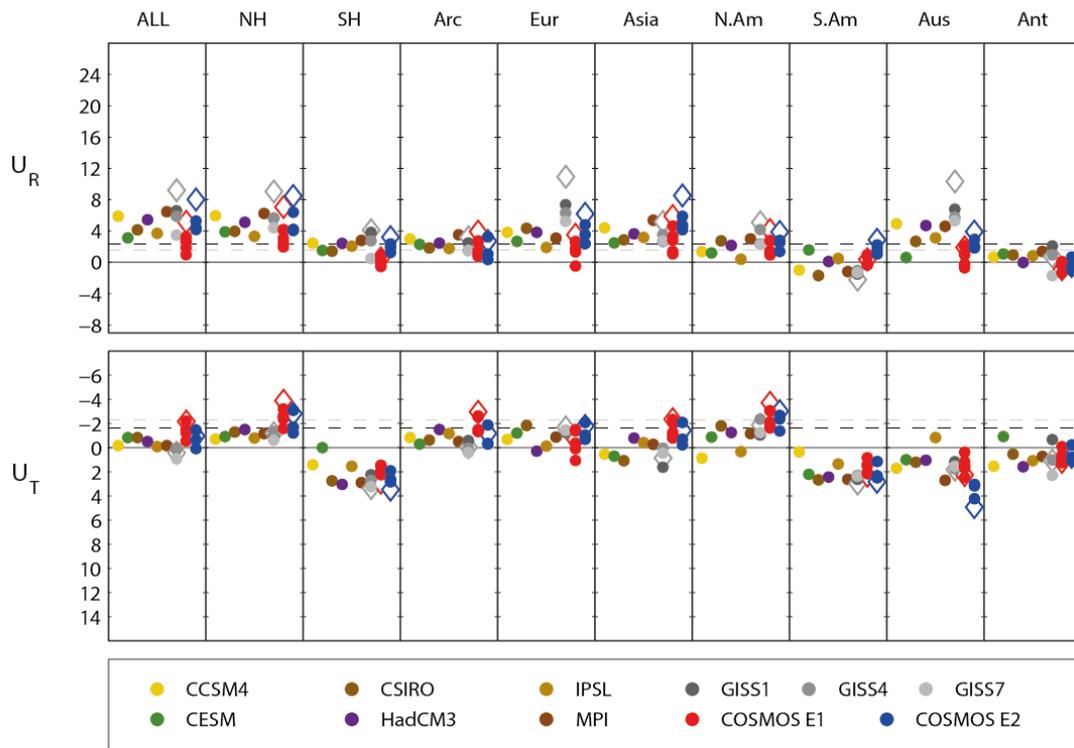
351



352

353 Figure S15: Correlation (U_R) and distance (U_T) statistics for PAGES 2k regions, with
 354 hemispheric and global combinations of all regional data, in the period 856–1350 CE. Positive
 355 U_R indicates that simulations and reconstructions have a positive correlation and that they
 356 share an effect of temporal changes in external forcings. Negative U_T indicates that a forced
 357 simulation is closer to the observed temperature variations than its own control simulation.
 358 Coloured dots: individual simulations. Diamonds: ensemble-mean results for COSMOS and
 359 GISS models. Dashed lines show one-sided 5% and 1% significance levels. Note the reversed
 360 vertical axis in the U_T graphs.

361

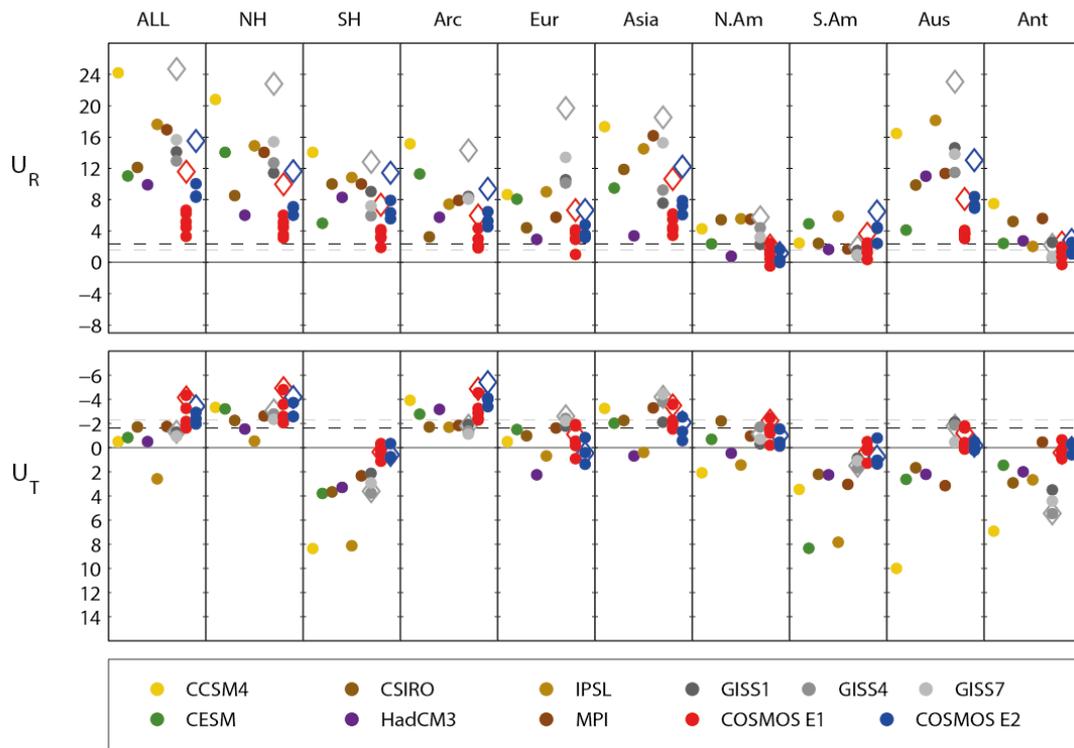


362

363 Figure S16: Same as Figure S15 but for the period 1356–1850.

364

365

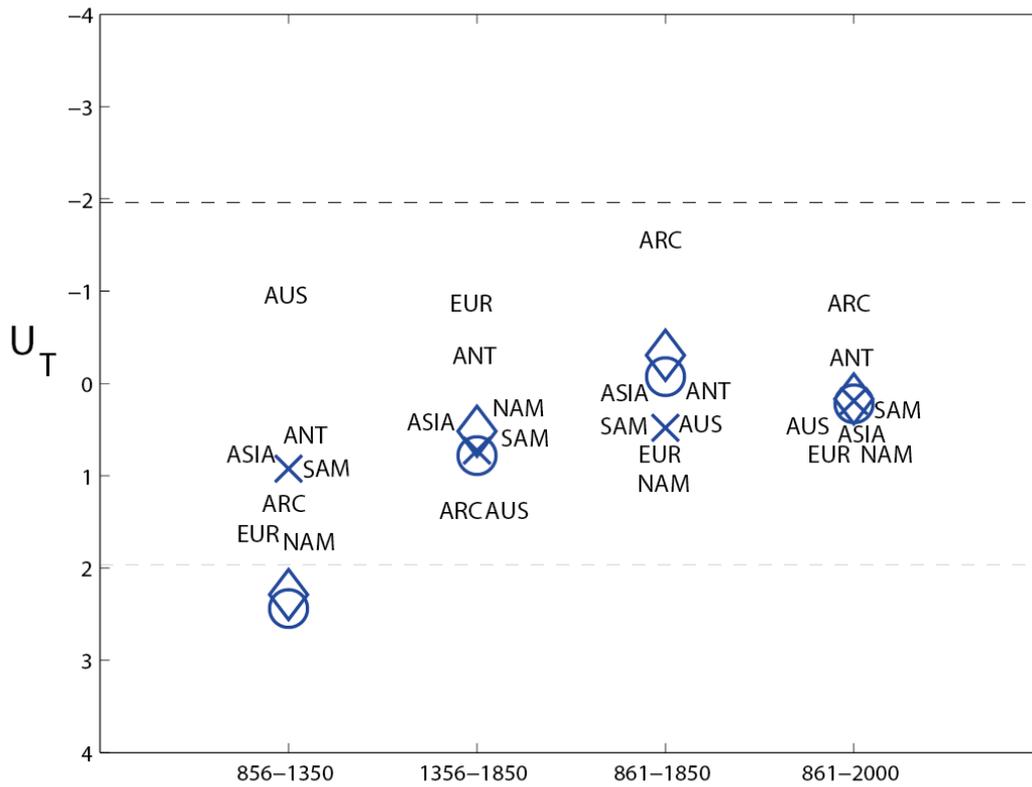


366

367 Figure S17: Same as Figure S15 but for the period 861–2000.

368

369



370

371 Figure S18: Distance (U_T) statistics computed for a direct comparison between the high vs.
 372 low solar COSMOS simulation ensembles, using the method of Moberg et al. (2015,
 373 Appendix B4), for PAGES 2k regions and four different analysis periods. A negative U_T
 374 (upwards in the graph) indicates that the high solar simulation ensemble is closer to the
 375 observed temperature variations than the low solar ensemble. Dashed lines show two-sided
 376 5% significance levels for the null hypothesis that the two simulations are equivalent. Results
 377 for each region are indicated with their abbreviated names. Results where regions are
 378 combined are shown with blue symbols: All regions (circle), Northern Hemisphere regions
 379 (diamond), Southern Hemisphere regions (cross).