Supplement of Clim. Past, 20, 865–890, 2024
https://doi.org/10.5194/cp-20-865-2024-supplement

Climate of the Past
Open Access
EGU

Supplement of

Towards spatio-temporal comparison of simulated and reconstructed sea surface temperatures for the last deglaciation

Nils Weitzel et al.

Correspondence to: Nils Weitzel (nils.weitzel@uni-tuebingen.de)

S1 Introduction

This supplement contains

- Additional information on the ensemble of transient simulations of the last deglaciation;
- Additional information on the proxy record processing;
- A more detailed description of the the model-data comparison algorithm;
- The methodology for constructing the regional stacks in Fig. 6 and Fig. 9 of the main manuscript;
- Exemplary plots of time series decompositions;
- An additional plot for pseudo-proxy experiments with misspecified proxy system model parameters;
- Additional visualizations of the comparison of simulated and reconstructed sea surface temperatures.

S2 Additional information on simulation ensemble

S2.1 MPI-ESM simulations

The MPI-ESM simulations use the coarse resolution version of MPI-ESM v.1.2 (MPI-ESM-CR). The atmospheric component runs at T31 horizontal resolution ($\sim 3.75°$) and 31 vertical levels. The ocean component employs an unstructured grid with a nominal resolution of $3°$ (Kapsch et al., 2022). Orbital parameters follow Berger and Loutre (1991) and greenhouse gas concentrations follow Köhler et al. (2017). Ice sheet topographies, surface topographies, glacier masks, land-sea masks, and bathymetries are updated based on either the GLAC-1D or the ICE-6G histories as described in the main manuscript and Kapsch et al. (2022). Ice sheet topography, glacier mask, land-sea mask, and bathymetry fields are interpolated to a 10-year resolution and updated every 10 simulation years (Kapsch et al., 2022). Accordingly, river pathways are updated every 10 simulation years based on the surface fields. Meltwater fluxes are computed as the temporal derivative of ice sheet thickness in the respective grid boxes (Kapsch et al., 2022). In the simulations with local meltwater input (see Table 1 in the main manuscript), the locations and amounts of meltwater are determined with a dynamically adapting river runoff scheme (Riddick et al., 2018). In MPI_Ice6G_P2_glob, meltwater is distributed equally across all land and ocean grid cells. In MPI_Ice6G_P2_noMWF, meltwater is removed from the system. Vegetation cover changes dynamically as computed by the dynamic global vegetation model JSBACH which is incorporated in MPI-ESM (Kapsch et al., 2022).

S2.2 CCSM3 simulations

The TraCE simulation design is described in detail in He (2011). The simulations are run with CCSM3 using a T31 horizontal resolution ($\sim 3.75°$) with 26 vertical levels for the atmosphere and a variable horizontal resolution with 25 vertical levels for the ocean ($3.6°$ in longitudinal direction, $\sim 0.9°$ in latitudinal direction near the equator but coarser at higher latitudes). Orbital parameters follow Berger (1978). Greenhouse gas concentrations are changed according to Joos and Spahni (2008). Ice sheet masks and topographies are updated approximately every 500 yrs following the ICE-5G history (see He, 2011, for details). Land-sea masks and bathymetries follow the ICE-5G reconstruction but are only adapted at four time steps at 13.1 ka (Barents Sea opens), 12.9 ka (Bering Strait opens), 7.6 ka (Hudson Bay opens), and 6.2 ka (Indonesian Throughflow begins). Location and amounts of meltwater fluxes were adapted manually as described in He (2011). Vegetation cover changes dynamically in the integrated dynamic global vegetation model CLM-DGVM (He, 2011). In TraCE-GHG and TraCE-ORB only greenhouse gas concentrations and orbital parameters are changed respectively, whereas all other boundary conditions are kept at the LGM state (22 ka) of the TraCE-ALL simulation.

S2.3 FAMOUS simulation

FAMOUS is a coarse resolution and simplified version of the HadCM3 climate model (Smith, 2012). The atmosphere module has a horizontal grid spacing of $5° \times 7.5°$ with 11 vertical levels and the ocean module has a horizontal grid spacing of $2.5° \times 3.75°$ with 20 vertical levels (Smith, 2012). In the assessed simulation, orbital parameters are changed transiently following Berger (1978) and greenhouse gas concentrations follow Lüthi et al. (2008). Northern Hemisphere (north of $40°$N) ice sheet topographies and glacier masks are updated according to the ICE-5G ice sheet history, whereas the Antarctic ice sheet topography and glacier mask, the land-sea mask, and the ocean bathymetry are fixed at pre-industrial values (Smith and Gregory, 2012). Due to the unchanged land-sea mask, ice sheet topographies are only prescribed for pre-industrial land grid cells and not over pre-industrial ocean grid cells. Meltwater fluxes are removed from the system. The vegetation cover is fixed at pre-industrial values. All boundary condition changes are applied with a 10x acceleration factor (Smith and Gregory, 2012).

S3 Details of proxy record processing

The PalMod 130k marine paleoclimate data synthesis v1.1.1 contains (near-)surface temperature reconstructions from original publications (Jonkers et al., 2020, 2023). While age models were recomputed and harmonized as described in Jonkers et al. (2020), the SST calibrations proposed by the original authors are used. We build on the published reconstructions, provided

in a harmonized and metadata rich format in the database. The database combines temperature reconstructions from multiple sensors. In particular, the selected records contain (near-)surface temperature reconstructions from Mg/Ca, U^k_{37}, planktonic foraminifera assemblages, TEX_{86}, and diatom assemblages. For some records, multiple calibrations of the same sensor measurements are archived in the database. To avoid biases from including sensor measurements multiple times, a preprocessing of the data provided in the database is required.

This preprocessing consists of five steps. First, we collect all available time series of (near-)surface temperature reconstructions in the database according to the metadata parameter 'surface.temp'. This results in 252 time series from 132 unique marine sediment cores.

Second, we remove records without BACON ages at any depth with available (near-)surface temperature reconstructions. This case can occur because BACON chronologies are only available for the sample depths between the first and last dating point. This step results in 247 time series from 132 unique marine sediment cores.

Third, we harmonize the provided information on recording or calibration season of the sensors/reconstructions. For some reconstructions, in particular many of the planktonic foraminifera assemblage reconstructions, a recording or calibration season is provided in the original data publication. This information is stored either in the ParameterOriginal or the RecordingSeason parameter of the database. We categorize the reconstructed time series into 'annual', 'warm season', 'cold season', or 'unknown' based on the available metadata. We only use the ParameterOriginal if no RecordingSeason is provided. 'unknown' is assigned whenever neither ParameterOriginal nor RecordingSeason contain information on the recording or calibration season.

Fourth, we average warm and cold season temperature reconstructions from the same core and sensor to obtain pseudo-annual temperature reconstructions. In all cases where warm and cold season temperature reconstructions are available, the same number of time series is available for each season. Therefore, we can average them without the need to first create averages for each season. This step results in 205 remaining time series from 132 sediment cores.

Fifth, we average and select records such that we do not include multiple reconstructed time series from the same sensor and core (Mg/Ca temperatures derived from different species are treated as different sensors). We average pseudo-annual, annual, and unknown seasonality reconstructions from the same core and sensor if all sample depths coincide. If the sample depths of time series from the same core and sensor do not coincide, we manually select the time series that provides a better coverage of the period 22-6 ka. Here, the length of the covered interval within that period and the temporal resolution are the determining criteria. In this process, we also remove one time series from core 'SU81_18' which has an unknown sensor in the database. Resolving these special cases results in 186 remaining time series from 132 sediment cores.

Using the record quality criteria described in the main manuscript (Sect. 2.2), we end up with 74 time series from 50 sediment cores which are used in the analyses of this manuscript.

Note that the correlations between time series calibrated from the same measurements are mostly high due to an often similar temporal pattern of orbital- and millennial-scale temperature variations. Therefore, the influence of the averaging on the pattern of orbital- and millennial-scale variations tends to be small. Calibrations for different seasons often vary in the magnitudes of variations. The construction of pseudo-annual records averages between these different magnitudes (since the temporal patterns for the different seasons are often highly correlated). Reconstructions for the same season tend to have similar magnitudes of variations such that the effect of averaging multiple calibrations for the same season is mostly small.

The preprocessing procedure is reproducible using the code provided in support of this manuscript (see section 'Code and data availability' in the main manuscript).

S4 Extended description of the model-data comparison algorithm

S4.1 Enhanced description of the four steps of the algorithm

In this section, we provide technical descriptions of the four steps of the algorithm (see main text for a summary of the steps). When running the algorithm for an ensemble of simulations and a set of proxy records, steps 1 to 3 are performed sequentially for all combinations of proxy records and simulations. Therefore, we describe these steps for one example proxy record and simulation below, while step 4 combines multiple records to a spatially averaged score.

S4.1.1 Step 1: compute forward-modeled proxy time series

We employ a simple PSM that takes simulated 3D (lon $\times$ lat $\times$ time) mean annual SST fields (T_{Sim}) as input and modifies them to resemble a reconstructed SST record (T_{FM}, FM = forward-modeled). The PSM consists of three steps, spatial interpolation (P_{space}), temporal downsampling (P_{time}), and an additive noise process (ε):

$$T_{\text{FM}} = P_{\text{time}}\left(P_{\text{space}}(T_{\text{Sim}})\right) + \varepsilon. \tag{S1}$$

First, we interpolate the spatial SST fields bilinearly to the proxy record location. Given the smoothness of SST fields on long time scales, the influence of the specific interpolation method is mostly negligible. Only in areas with strong local heterogeneities such as coastal upwelling zones, the interpolation method could potentially impact the resulting time series. However, the coarse resolution of deglacial simulations likely limits the accurate representation of small-scale structures more than the interpolation method (see also Sect. 5.2 in the main manuscript). For the downsampling of the simulated time series to the time axis of the proxy record, we draw N Monte Carlo realizations of pairs of simulated and reconstructed time series to quantify chronological uncertainties. For each Monte Carlo realization, we randomly select one iteration of the age-depth model (see Sect. 2.2 in the main manuscript) and downsample the simulated time series to the irregular time axis of the proxy record using blocksampling. The blocksampler cuts the simulated time series into disjoint slices with cutting dates at the midpoints between the sample ages, and assigns the averaged signal of each slice to the date of the corresponding sample. This procedure imitates the limited temporal resolution and integrated nature of the proxy records. The result of the temporal downsampling is a temporally aligned set of N reconstructed SST time series (Fig. 2 in the main manuscript, top left) and N time series of downsampled SST simulations. The blocksampling strategy assumes that gaps in the sampling of records are smaller than the depths over which individual samples average, at least after accounting for smoothing from bioturbation. More detailed reporting of the top and bottom sampling depths of each sample could be used to refine the downsampling procedure and quantify its influence.

In our PSM, we summarize the effects of the inherent uncertainties of SST reconstructions (see Sect. 1) and uncertainties in the formulation of the PSM by a Gaussian additive noise process with a specified signal-to-noise ratio (SNR) and temporal autocorrelation structure. For each of the N time series of downsampled SST simulations, we add a random realization of the additive noise process to the SST time series. We call the N resulting time series 'forward-modeled proxy time series' (Fig. 2 in the main manuscript, top right). We use the additive noise approach because metadata is missing to explicitly model processes that lead to deviations of the reconstructions from mean annual SSTs for all records in the compilation. In addition, climate models do not simulate all variables required to model these processes. For example, the recording season and depth of the sensors are uncertain, insufficiently reported in the literature, and might vary over the last deglaciation due to habitat tracking (Jonkers and Kučera, 2017; Mix, 1987). Therefore, we compare all (near-)surface temperature reconstructions with mean annual SSTs from the simulations. As we only analyze SST changes over time, offsets from mean annual SSTs in the absolute reconstructed temperatures, which stay (nearly) constant over time, are not affecting our results.

S4.1.2 Step 2: decompose time series

In step two, we extract the four components of the deglacial SST evolution outlined above: magnitudes and patterns for both orbital- and millennial-scale variations. We first decompose each of the N time series into three timescales with Gaussian smoothers (Fig. 2 in the main manuscript, second row; see Fig. S2-S9 in the supplemental information for further examples of timescale decompositions). We use Gaussian smoothers because they are a robust method for the analysis of irregularly spaced time series in the time and frequency domain (Rehfeld et al., 2011). The three timescales are orbital-, millennial-, and sub-millennial-scale variations, whose ranges abut one another. We select a smoothing period of 1 kyr to separate sub-millennial from millennial timescales. Since there is no clear scale separation between millennial and orbital variations, we employ three smoothing periods, 4 kyr, 6 kyr, and 8 kyr, and average the respective quantified deviations after step 4.

Next, we isolate the temporal patterns of the variations from their magnitudes. To this purpose, we compute the standard deviations of all reconstructed and forward-modeled proxy time series, which are a measure of the magnitude of variations on a given timescale. As we obtain one estimate from each Monte Carlo realization, this leads to probability distributions for the timescale-dependent magnitudes of variations in reconstructed and forward-modeled proxy time series. We define the

pattern of the respective variations as the normalized, i.e. centered and standardized, time series. We obtain N realizations of normalized time series. Each realization has the same number of time steps, M, and each time step corresponds to the depth of a proxy sample in the sediment record. Thus, the realizations can be interpreted as an empirical, M-dimensional probability distribution with auto-correlation between the time steps. The time series decomposition results in four probability distributions (orbital and millennial magnitudes as well as patterns, Fig. 2 in the main manuscript, third row). Each of the distributions is represented by N Monte Carlo realizations.

S4.1.3 Step 3: quantify deviations between forward-modeled proxy time series and reconstructed SST records

In the third step, we compute the deviation between the simulated forward-modeled proxy time series and reconstructions for each proxy record and each of the four components (Fig. 2 in the main manuscript, bottom row). Each of these deviations is quantified with the integrated quadratic distance (IQD). The IQD is a proper divergence function that has desirable mathematical properties for model selection as it penalizes overly confident or conservative uncertainty estimates compared to the unknown true uncertainties (Thorarinsdottir et al., 2013). The name IQD is motivated by the fact that in one dimension, the IQD is equal to the integral over the squared difference between two cumulative distribution functions. However, the IQD is not just applicable to univariate distributions but also to multivariate probability distributions. Its M-dimensional definition is

$$\mathrm{IQD}(\mathbb{P},\mathbb{Q}) \;=\; \frac{1}{M}\mathbb{E}_{\mathbb{P},\mathbb{Q}}|X-Y| - \frac{1}{2M}\left(\mathbb{E}_{\mathbb{P}}|X-X'| + \mathbb{E}_{\mathbb{Q}}|Y-Y'|\right), \tag{S2}$$

where $\mathbb{P}$ is the probability distribution of forward-modeled proxy time series, $\mathbb{Q}$ is the probability distribution of the reconstructions, M is the dimension of $\mathbb{P}$ and $\mathbb{Q}$, and $\mathbb{E}$ denotes expected values. Further, X and X' are independent random variables distributed according to $\mathbb{P}$, and Y and Y' are independent random variables distributed according to $\mathbb{Q}$. The first term in Eq. (S2) is the expected difference between draws from the distributions of forward-modeled proxy time series ($\mathbb{P}$) and reconstructions ($\mathbb{Q}$). The two last terms quantify the spread of the distributions $\mathbb{P}$ and $\mathbb{Q}$ since $\mathbb{E}_{\mathbb{P}}|X-X'|$ is the expected difference between two random draws from the distribution $\mathbb{P}$.

We compute the IQD using a Monte Carlo approximation of Eq. (S2) with the Monte Carlo realizations from step 2. Thereby, we approximate the analytically intractable distributions $\mathbb{P}$ and $\mathbb{Q}$ by empirical distributions. For the patterns of variations, the N time series realizations are again interpreted as M-dimensional probability distributions, where M corresponds to the number of time steps in a proxy record, and the IQD computes the difference between these distributions. Numerical tests determined that IQD estimates are stable for $N \geq 100$ (see supplemental information). Therefore, we use $N = 100$ for the computationally demanding PPEs and $N = 1000$ for the real-world application. Computational details are provided in the supplemental information (Text S4).

The IQD takes positive values ($\mathrm{IQD}(\mathbb{P},\mathbb{Q}) \geq 0$). It is only zero when $\mathbb{P}$ and $\mathbb{Q}$ are equal ($\mathrm{IQD}(\mathbb{P},\mathbb{P}) = 0$). Smaller IQD values imply a smaller deviation and thus a better agreement of forward-modeled proxy time series and reconstructions. In the absence of age and proxy uncertainties, the IQD reduces to the mean absolute difference between numbers (magnitudes) or time series (patterns). The IQD can be applied to quantities of arbitrary units. In our case, the units are temperature [K] for the comparison of magnitudes, and standard deviations [z] for patterns.

S4.1.4 Step 4: average deviations in space

We analyze IQDs averaged on four spatial scales: locally, regionally (see color-coding of dots in Fig. 1b of the main manuscript for the assignment of proxy records to the regions considered in this study), zonally, and globally. For local IQDs, we treat each proxy record individually, i.e. without averaging proxy records from the same core or nearby locations. Zonal IQDs are obtained by averaging over proxy records within overlapping bands of 20° width that move in 5° steps (Fig. 2 in the main manuscript, bottom row). We only consider latitudinal bands containing at least five proxy records to only incorporate spatial averages where we can assume that a substantial amount of non-climatic influences is averaged out.

S4.2 Approximation of the Integrated Quadratic Distance using Monte Carlo realizations

To compute the Integrated Quadratic Distance (IQD, Sect. 3.1.3 of the main manuscript), we need to approximate

$$\mathrm{IQD}(\mathbb{P},\mathbb{Q}) \;=\; \frac{1}{M}\mathbb{E}_{\mathbb{P},\mathbb{Q}}|X-Y| - \frac{1}{2M}\left(\mathbb{E}_{\mathbb{P}}|X-X'| + \mathbb{E}_{\mathbb{Q}}|Y-Y'|\right), \tag{S3}$$

with the N Monte Carlo (MC) realizations of reconstructed and forward-modeled proxy time series. Remember that $\mathbb{P}$ is the probability distribution of forward-modeled proxy time series, $\mathbb{Q}$ is the probability distribution of the reconstructions, M is the dimension of $\mathbb{P}$ and $\mathbb{Q}$, and $\mathbb{E}$ denotes expected values. Further, X and X' are independent random variables distributed according to $\mathbb{P}$, and Y and Y' are independent random variables distributed according to $\mathbb{Q}$. A standard Monte Carlo approximation of Eq. (S3) is

$$\mathrm{IQD}(\mathbb{P},\mathbb{Q}) \;\approx\; \frac{1}{MN}\sum_{n=1}^{N}|X_n-Y_n| - \frac{1}{2M}\frac{2}{N(N-1)}\left(\sum_{n=2}^{N}\sum_{m<n}|X_n-X_m| + \sum_{n=2}^{N}\sum_{m<n}|Y_n-Y_m|\right), \tag{S4}$$

where $|\cdot|$ denotes the Euclidean distance between one- or multi-dimensional vectors and X_n, X_m, Y_n and Y_m are the respective MC realizations. Eq. (S4) needs $O(N^2)$ operations which becomes computationally expensive for high dimensional vectors such as the pattern time series. Therefore, we employ a less accurate approximation of the last two terms in Eq. (S3), which only needs $O(N)$ operations:

$$\mathrm{IQD}(\mathbb{P},\mathbb{Q}) \;\approx\; \frac{1}{MN}\sum_{n=1}^{N}|X_n-Y_n| - \frac{1}{2MN}\left(\sum_{n=1}^{N-1}|X_n-X_{n+1}| + |X_N-X_1| + \sum_{n=1}^{N-1}|Y_n-Y_{n+1}| + |Y_N-Y_1|\right). \tag{S5}$$

We find that using the less accurate Eq. (S5) with larger N is in total more accurate than using Eq. (S4) with lower N (not shown). Therefore, we use Eq. (S5) to approximate Eq. (S3).

S4.3 Sensitivity to number of Monte Carlo realizations

To find out how many Monte Carlo (MC) realizations are needed to obtain stable IQD approximations, we perform pseudo-proxy experiments (PPEs).

We use four different simulations as reference simulations and employ the proxy system model (PSM) with an AR1 noise using SNR=1.6 and a decorrelation length of 1289 yrs. The four reference simulations are MPI_Glac1D_P3, MPI_Ice6G_P2_glob, TraCE-ALL, and FAMOUS.

To compute the IQD, we use 5, 10, 25, 50, 75, 100, 250, 500, 750, and 1000 MC realizations. To isolate the effect from including additional MC realizations, we always use all MC realizations from the next smaller realization size and then include additional realizations.

In Fig. S1, we plot the mean change in absolute IQD per 10 additional MC realizations. For larger realization numbers, absolute IQD differences per 10 additional MC realizations become smaller. In particular, IQD differences per additional MC realizations become negligible when increasing the MC realization number beyond $N = 100$. This finding holds for all four components of the deglacial temperature evolution and all three spatial averaging scales (global, zonal, and local). We conclude that IQD estimates are robust for MC realization number of $N \geq 100$. Therefore, we employ $N = 100$ in the computationally expensive large sets of PPEs (used in Sect. 4.1.2 and 4.1.3 of the main manuscript), but use $N = 1000$ in the example PPE (Sect. 4.1.1) and the real-world application (Sect. 4.2).

S5 Construction of regional stacks

We constructed regional stacks of reconstructions and forward-modeled proxy time series in Fig. 9 of the manuscript. These are not part of the model-data comparison methodology but only serve to facilitate a better (visual) understanding of the diagnosed model-proxy deviations. The stacks are constructed in four steps:

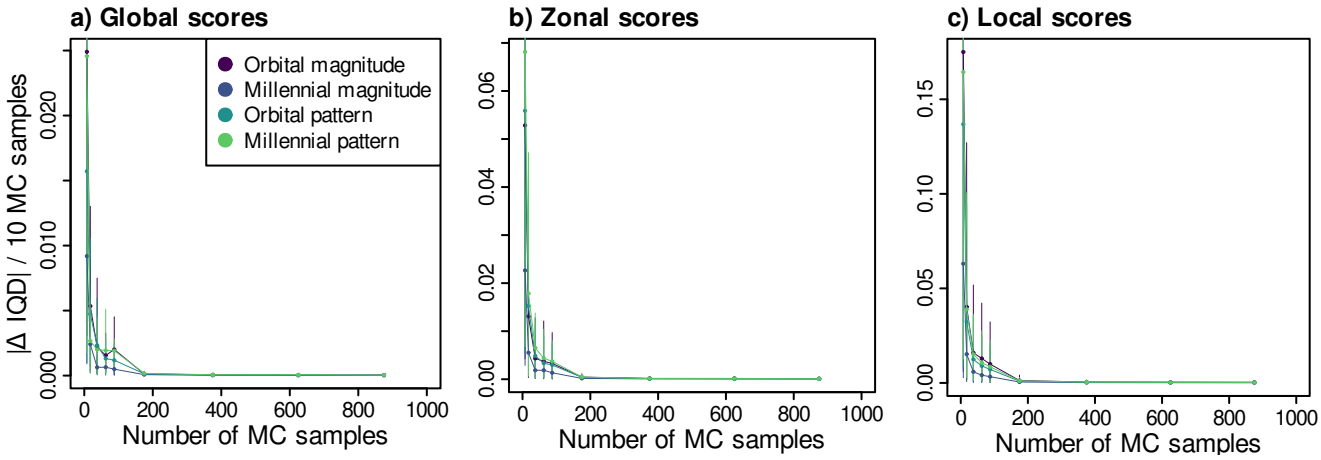

Figure S1. Changes in the absolute IQDs per 10 additional MC realizations for (a) globally averaged IQDs, (b) zonally averaged IQDs, and (c) local IQDs. Values are positive by definition. Connected dots correspond to means over all PPEs with the same number of MC realizations, whereas vertical lines denote 90% confidence intervals across PPEs with the same number of MC realizations (given by the 5th and 95th percentiles of the respective distributions). As each dot corresponds to the difference in IQD between subsequently tested realization numbers, dots are plotted at the mean of the two realization numbers.

1. Select all proxy records in the specified region.

2. Interpolate the reconstructed and proxy-forward modeled time series to a common time axis, which covers the interval 19 ka to 9 ka with a time step of 100 yr. The interpolation is performed for each of the 1000 Monte Carlo realizations that quantify uncertainties from the age-depth models and the stochastic proxy system model. If a time series does not encompass the interval 19-9 ka, the values outside of the first and last age are set to NA.

3. For each of the 1000 Monte Carlo realizations, the average over all records is taken, first for the interpolated reconstructions and then for the interpolated forward-modeled proxy time series. This procedure procedure results in 1000 realizations of the average temperature evolution of the proxy records and proxy-forward modeled time series.

4. Compute the mean and the 0.05 and 0.95 quantiles from the 1000 realizations for each time step. The mean is plotted as thick line (dashed line for sensitivity experiments) and the area between 0.05 and 0.95 quantiles as shaded polygons.

We employ this methodology (1) for the reconstructed and proxy-forward modeled time series before employing the decomposition into different components of the deglacial temperature evolution (Fig. 6 of the main manuscript), (2) for the orbital pattern components (Fig 9a,c,e of the main manuscript), and (3) for the millennial pattern components (Fig. 9b,d,f of the main manuscript).

Note that the purpose of the stacks is not to provide a good estimate of the regional mean temperature evolution but just to provide an impression of the average behavior at a given set of proxy locations. If the temporal evolution at these locations is differing from the regional mean behavior, the stacks will also deviate from the regional mean evolution. Therefore, we do not apply typical techniques to account for uncertainties from sampling limitations such as bootstrapping over the proxy records.

S6 Additional time series decompositions

We plot additional examples of time series decompositions into orbital, millennial, and sub-millennial contributions (compare with Fig. 2 of the main manuscript). For each of the selected 74 proxy records from the PalMod 130k marine paleoclimate data synthesis v1.1.1 (Jonkers et al., 2020, 2023) (Sect. 2.2 of the main manuscript), we randomly select an age-depth history

and a realization of the PSM (SNR = 1.6, AR1 noise with decorrelation length 1289 yrs) with MPI_Glac1D_P3 as reference simulation. We derive the orbital contributions for three smoothing periods, 4 kyr, 6 kyr, and 8 kyr (this leads to three versions of the orbital variations). Millennial and sub-millennial variations are separated using a smoothing period of 1 kyr. Therefore the sum of orbital and millennial variations is the same in all three decompositions. Fig. S2 - Fig. S9 show this single realization of the reconstructed and the forward-modeled proxy time series (black), the orbital variations (blue), and the sum of orbital and millennial variations (red) for each of the 74 proxy records.

S7 Additional plot for results of the pseudo-proxy experiments

Fig. S10 shows the FPRR for over- or under-estimated SNRs and for over- (power-law) or under-estimated (white noise) temporal persistence of non-climatic processes. The FPRR medians and spreads for orbital magnitudes and patterns vary very little for over- or underestimated SNRs across the whole range of SNRs. Thus, correctly estimating SNRs or the temporal structure of the autocorrelation has very little influence on the reliability and robustness of orbital-scale IQDs. For millennial patterns, results are very stable as long as non-climatic noise is not completely neglected (SNR=Inf). For SNR=Inf, medians and spreads of FPRRs both increase, but the medians are still below the 95th FPRR percentile for the correct SNR. The influence of misspecified SNRs is largest for millennial magnitudes, where we find two opposing trends. On the one hand, the median FPRR stays relatively constant for overestimated SNRs but tends to increase for underestimated SNRs. On the other hand, the spread varies little for underestimated SNRs but increases for overestimated SNRs. This suggests that the reliability for millennial magnitudes decreases when the SNR is underestimated whereas the robustness is lower when the SNR is overestimated.

S8 Additional plots for the comparison of simulated and reconstructed SSTs

Fig. S11 - Fig. S17 show supplemental visualizations of the results presented in Sect. 4.2 of the main manuscript.

Since magnitude and pattern scores provide different information on the model-proxy agreement, we combine these two aspects in biplots for a more integrative assessment of performances on a given timescales. The best performing models are located in the lower left corner. Fig. S11 shows biplots with all ten simulations, whereas Fig. S12 is a version without the three 'sensitivity experiments', TraCE-ORB, TraCE-GHG, and FAMOUS. The latter plot improves the visualization of differences between the seven simulations in our main set of simulations in cases where the performance of at least one of the sensitivity experiments differs strongly from the six MPI-ESM simulations and TraCE-ALL.

We compute rank statistics to compare the IQDs of TraCE-ALL and the six MPI-ESM simulations across regions and components of the deglacial temperature evolution. Rankings are computed for each proxy record and each of the four components. We provide a description and interpretation of the rank histograms in Sect. 5.2 of the main manuscript.

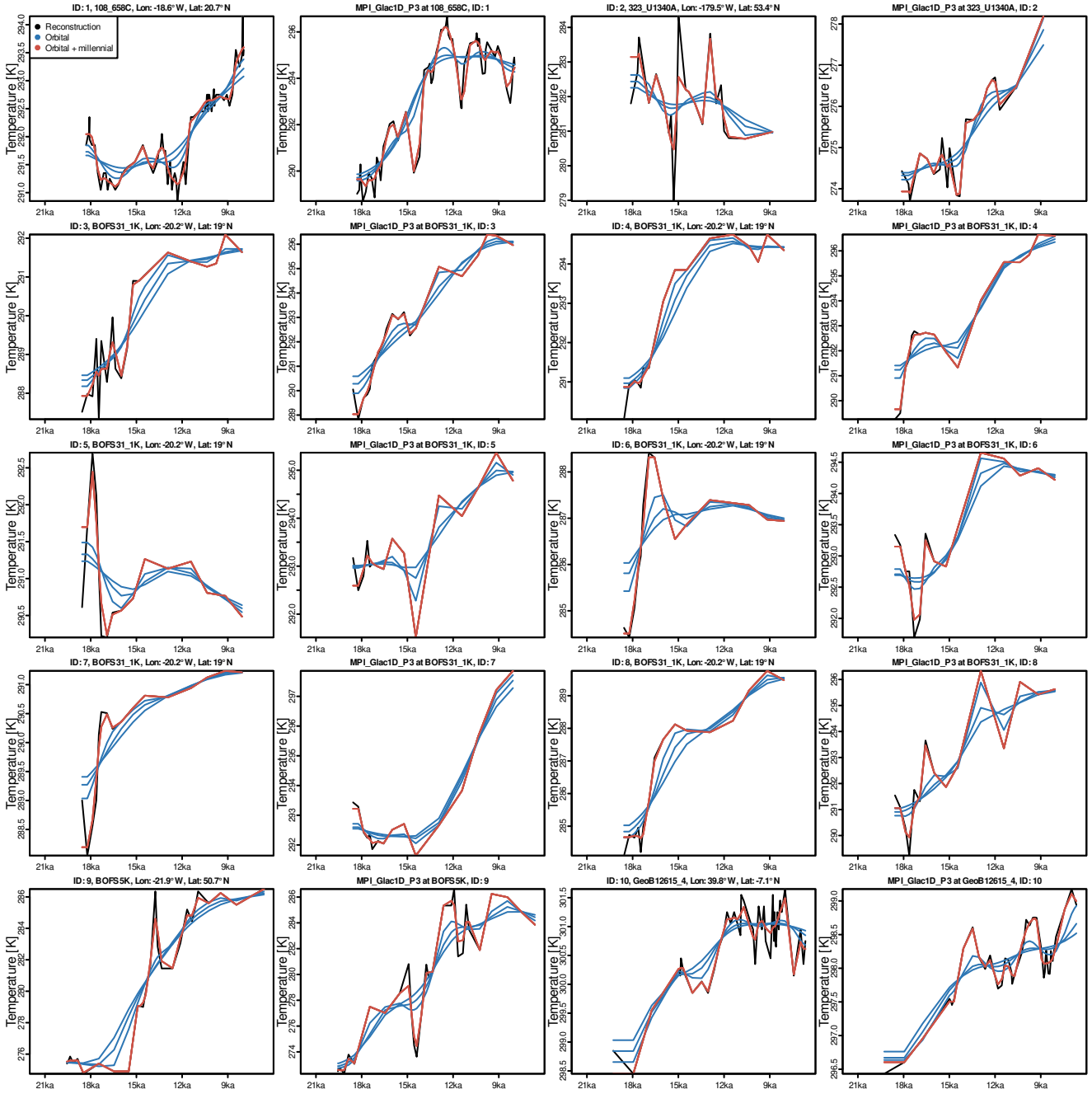

Figure S2. Single realization of the reconstructed and the forward-modeled proxy time series (black), the orbital variations (blue), and the sum of orbital and millennial variations (red) for each of the 74 proxy records (see Sect. S6 for details). Orbital and millennial variations are separated using three different smoothing periods, 4 kyr, 6 kyr, and 8 kyr, leading to three versions of the orbital variations with different degrees of smoothing.

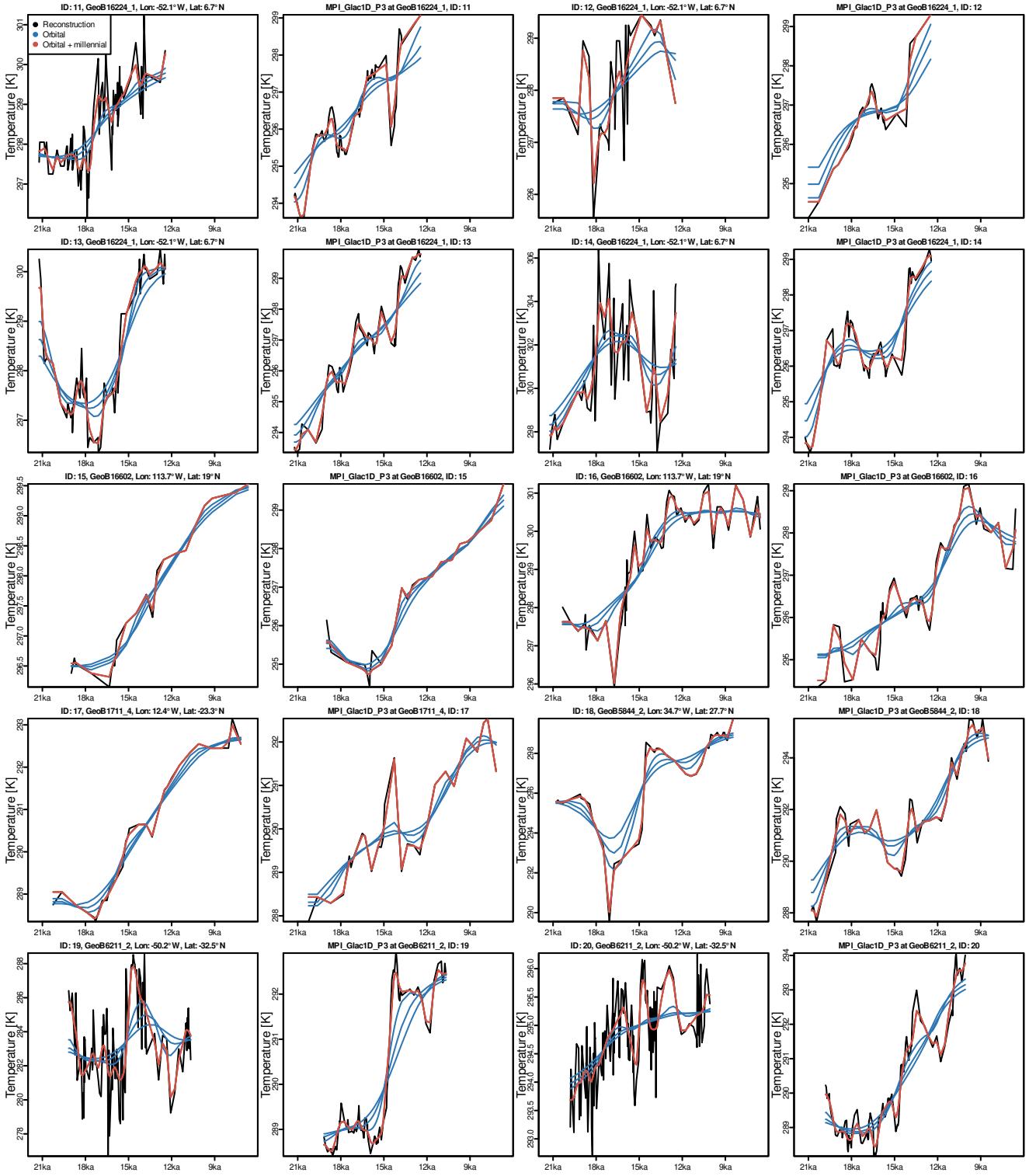

Figure S3. As Fig. S2 but for other proxy records.

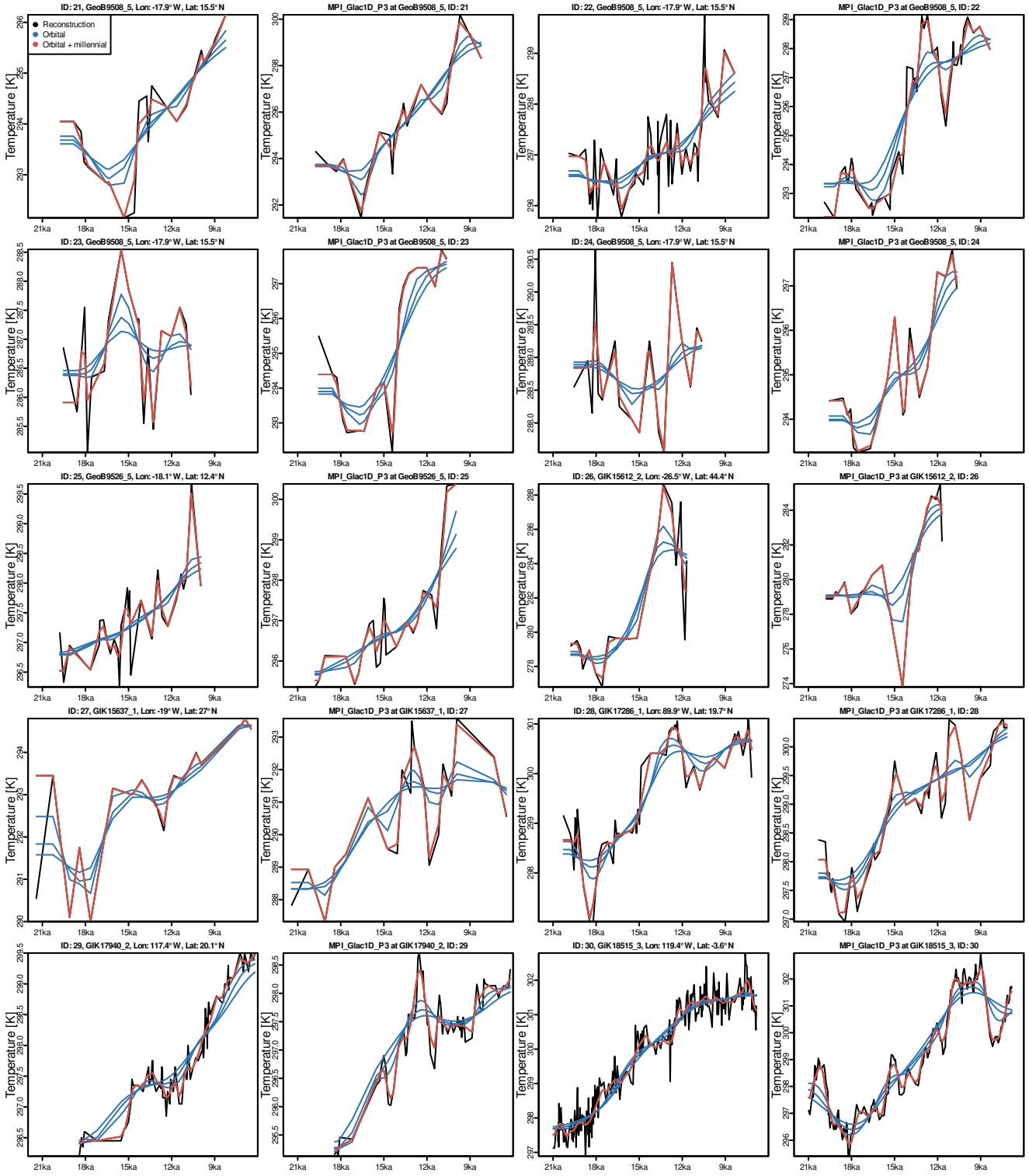

Figure S4. As Fig. S2 but for other proxy records.

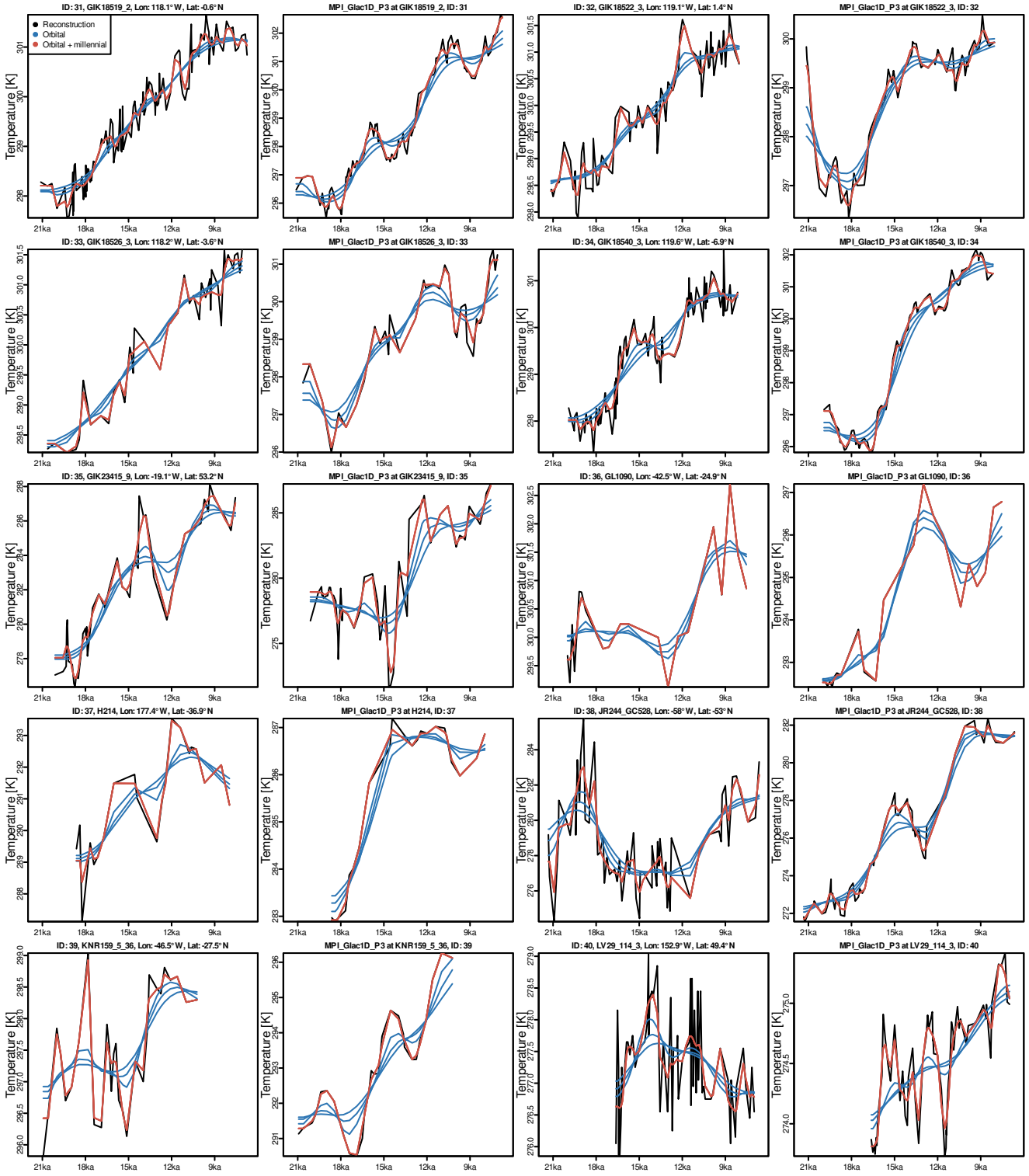

Figure S5. As Fig. S2 but for other proxy records.

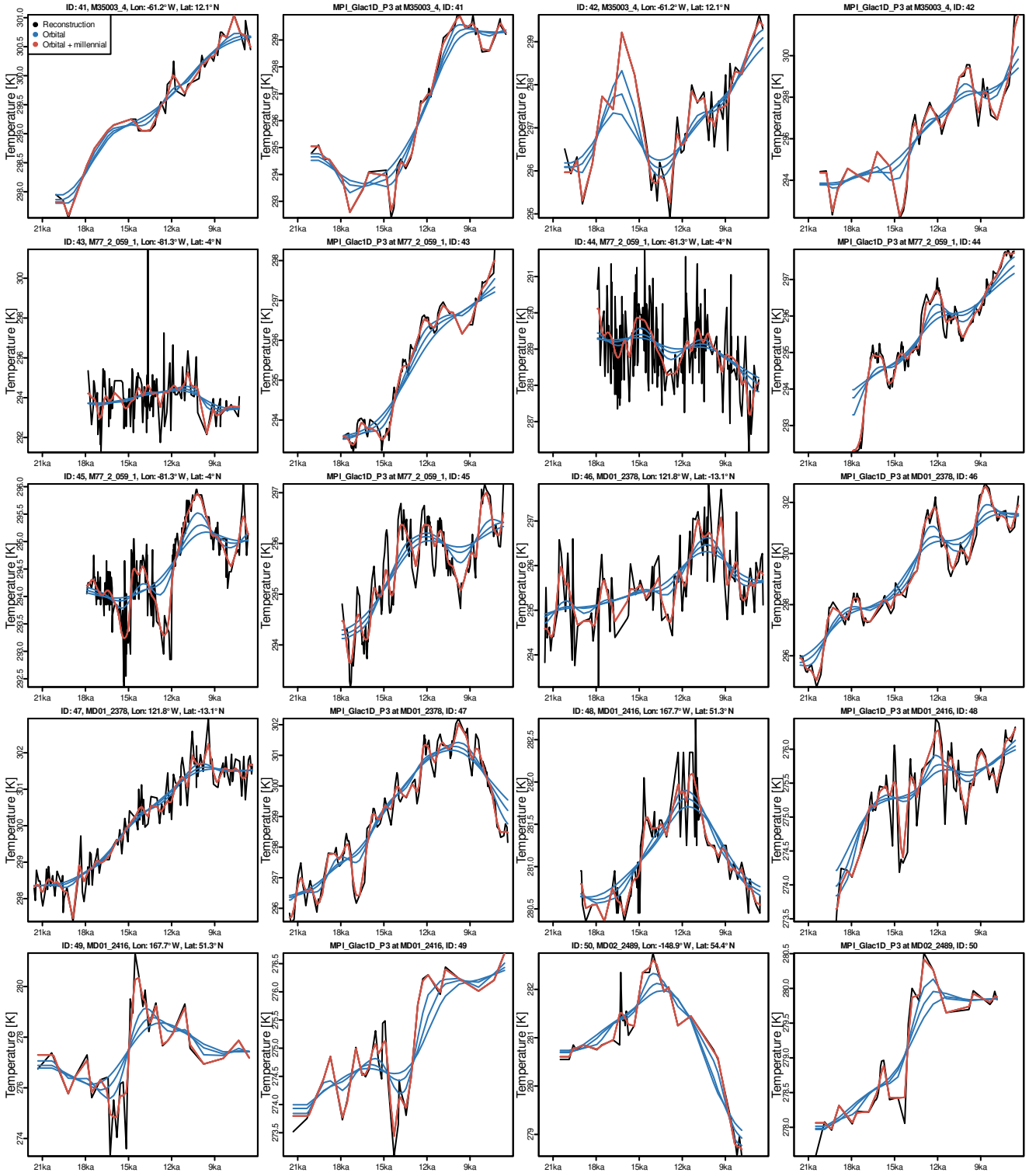

Figure S6. As Fig. S2 but for other proxy records.

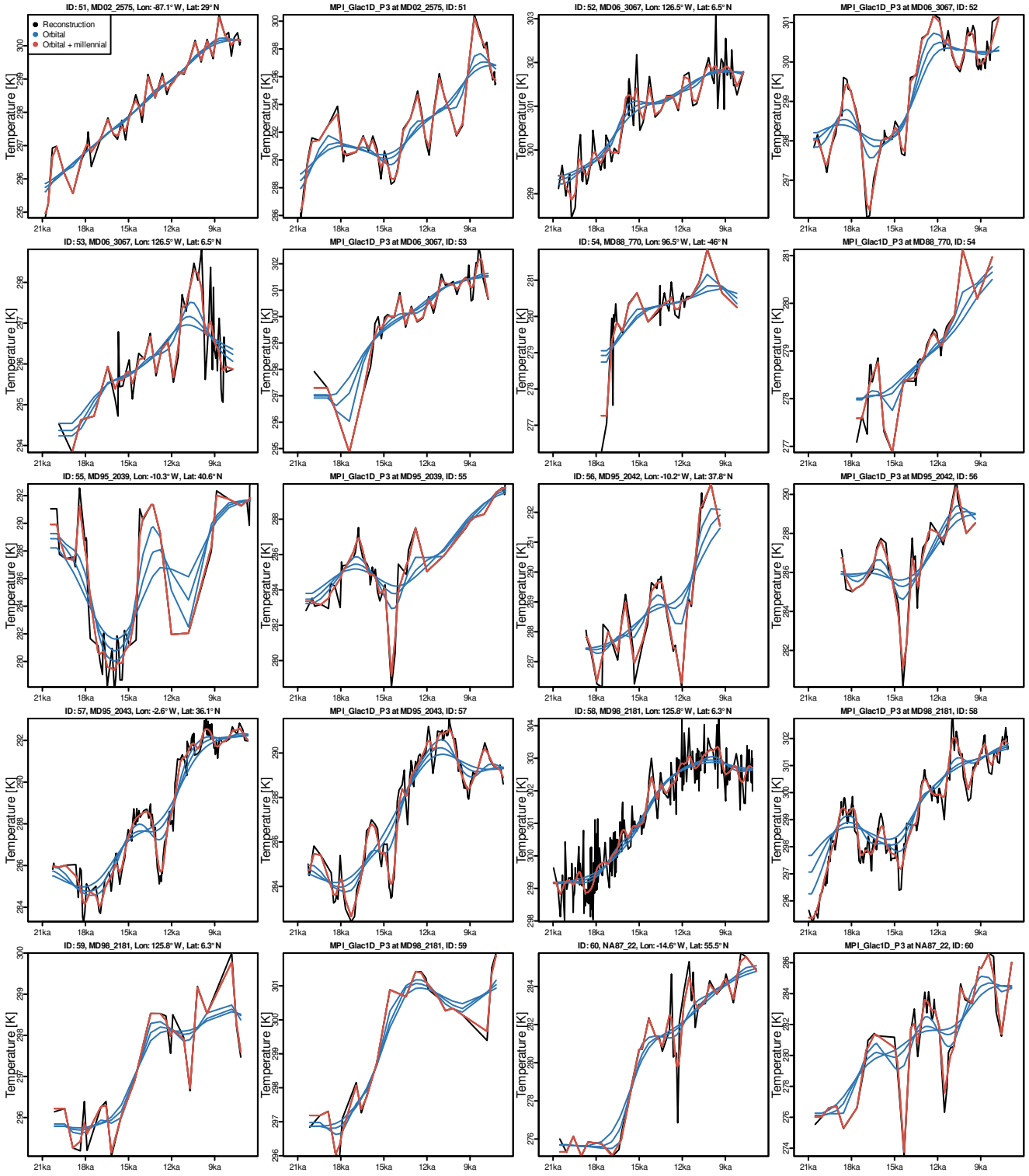

Figure S7. As Fig. S2 but for other proxy records.

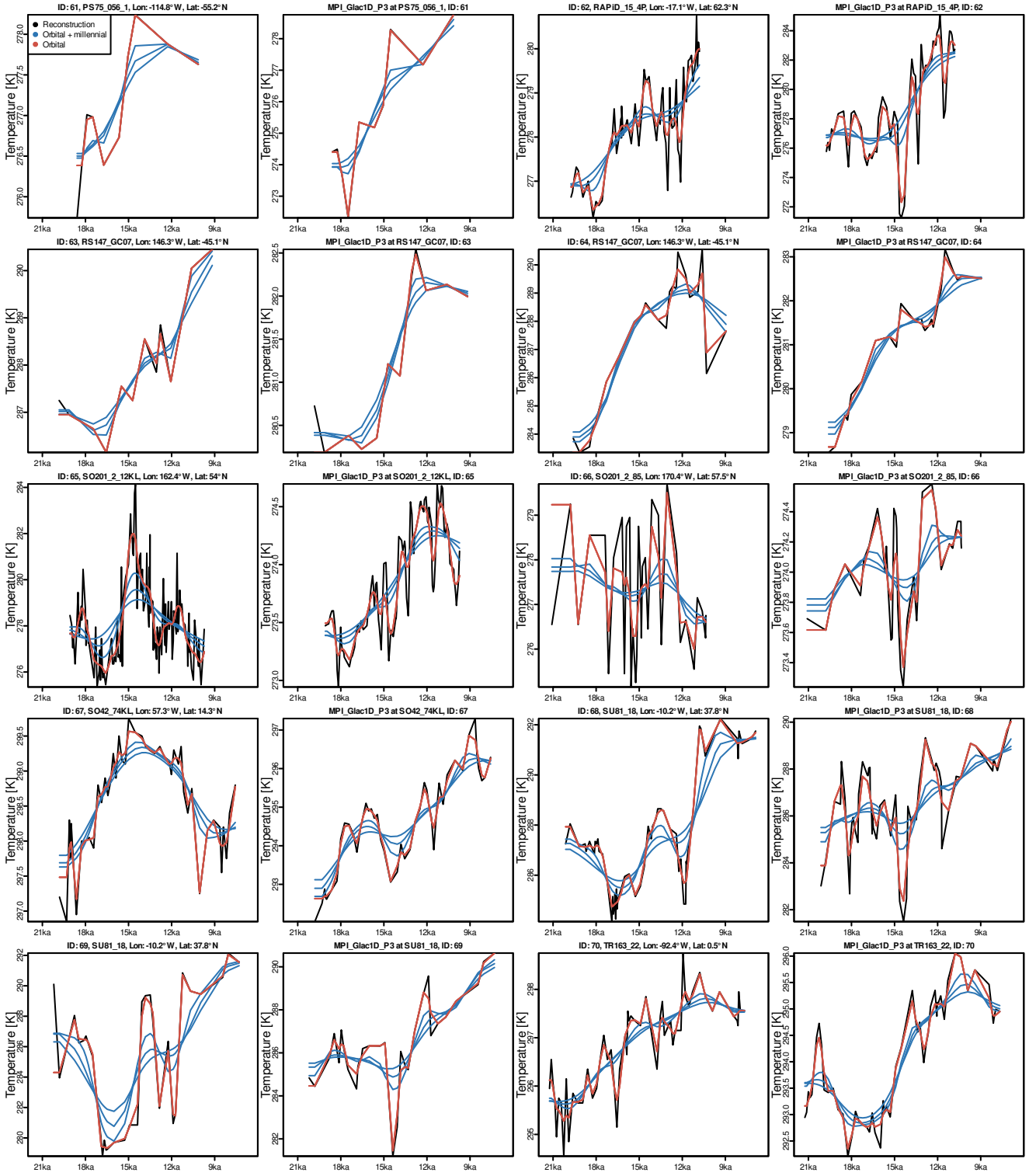

Figure S8. As Fig. S2 but for other proxy records.

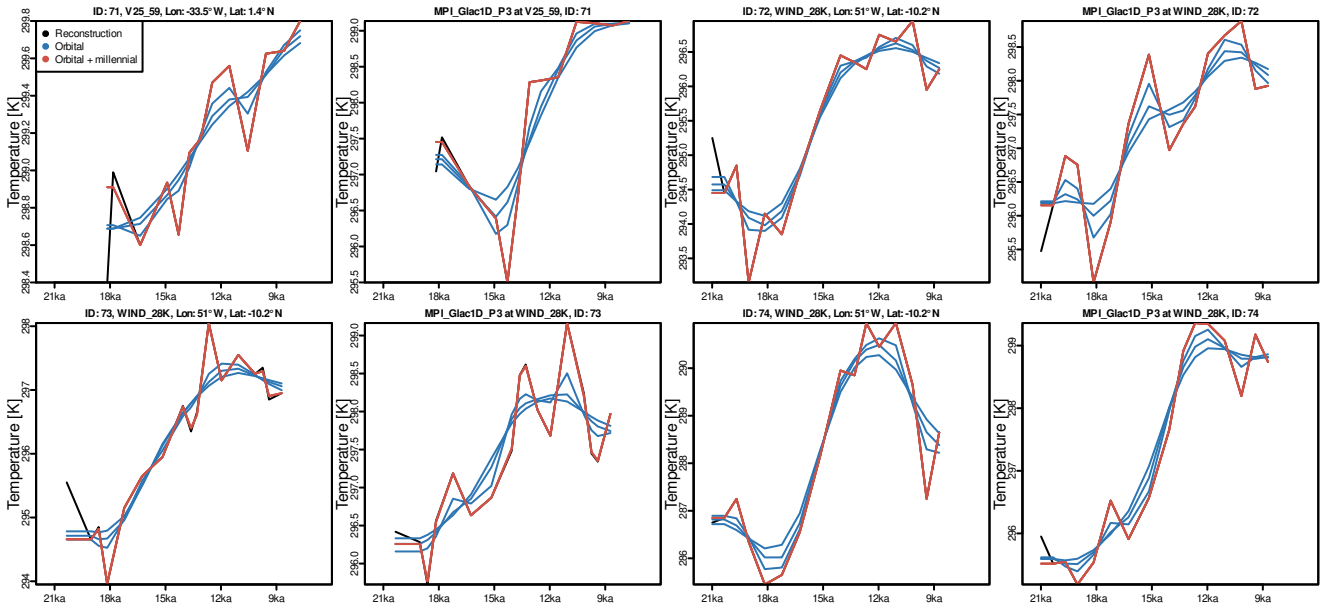

Figure S9. As Fig. S2 but for other proxy records.

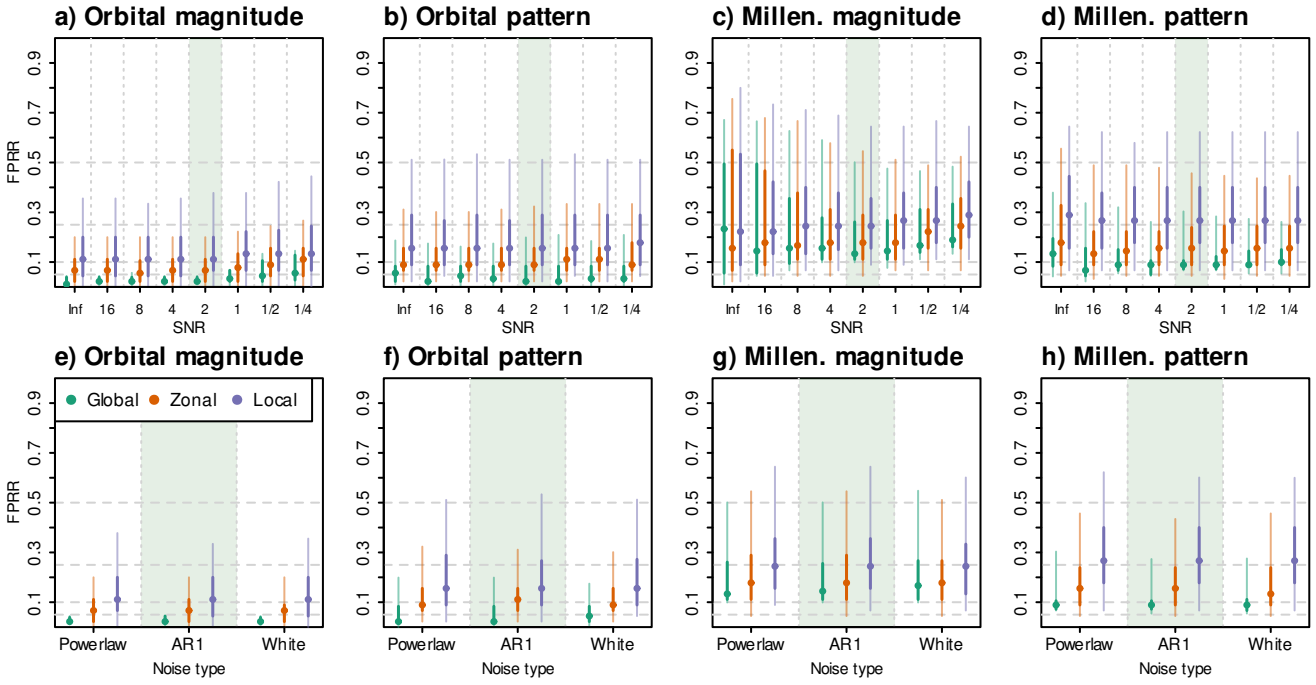

Figure S10. Effects of misspecified SNRs and noise types on simulation rankings in PPEs. The reference configuration of the pseudo-proxies in all PPEs is SNR=2 and an AR1 noise with a decorrelation length of 1000 yrs. (a) - (d) show results for the four different components of the deglacial temperature evolution in PPEs with varying SNRs of the forward-modeled proxy time series. (e) - (h) show results for PPEs in which the noise type of the forward-modeled proxy time series varies. Green shaded rectangles indicate the PPEs in which the same noise configuration (SNR=2, AR1 noise) is used for the reference pseudo-proxies and the forward-modeled proxy time series. Dots depict the medians across all PPEs with (a-d) a given SNR (n=10 for each SNR) or (e-h) a given noise type (n=10 for each noise type). Bars show the spread across PPEs. Darker colors depict the 25th to 75th percentiles, whereas lighter colors depict the 5th to 95th percentiles. Dashed horizontal lines indicate FPRRs of 0.05, 0.1, 0.25, and 0.5. FPRRs above 0.5 are worse than expected for a randomized ranking.

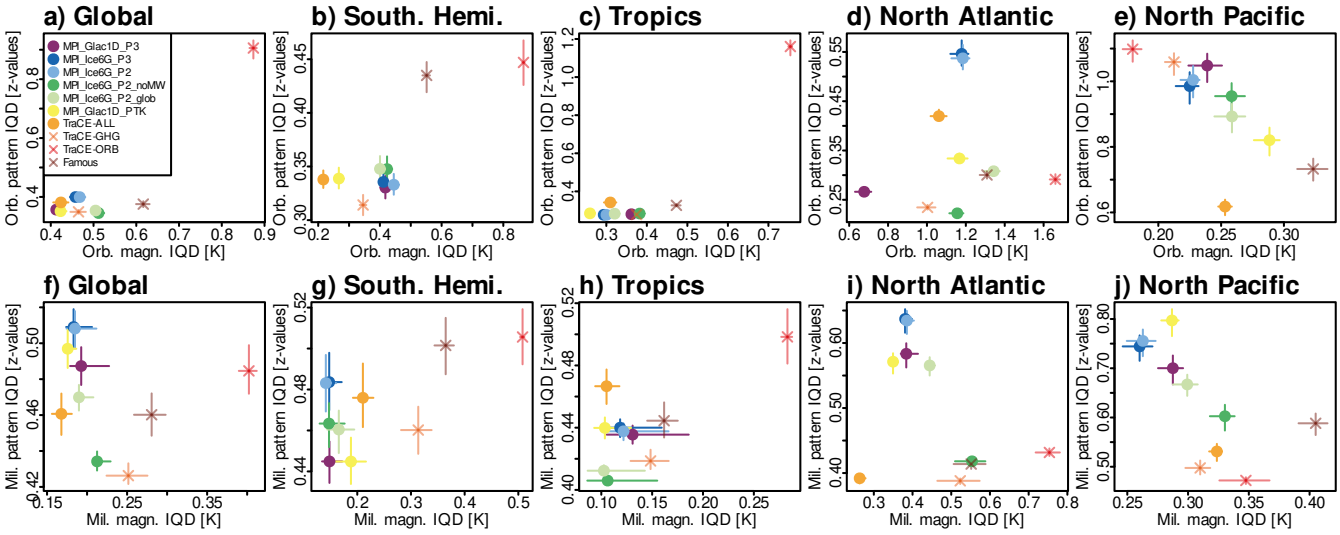

Figure S11. Biplots of IQDs for orbital-scale and millennial-scale variations. The magnitude IQDs of variations are plotted on the x-axes and the pattern IQDs on the y-axes. Lines indicate uncertainties from varying the PSM parameters. Dots in the lower left corner indicate simulations with the highest model-proxy agreement for magnitudes and patterns.

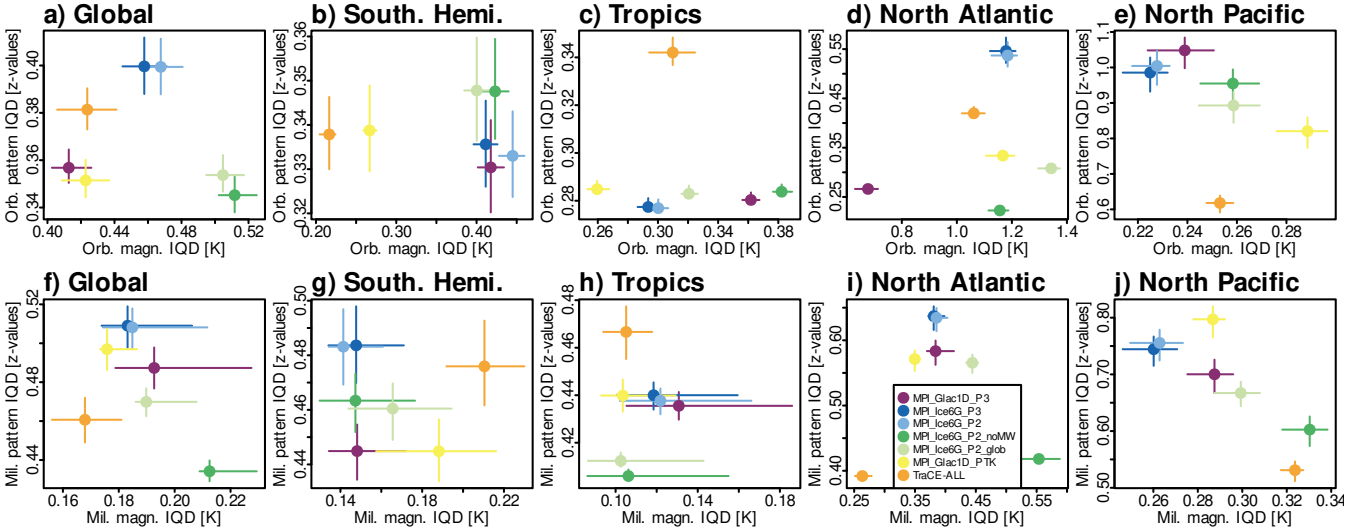

Figure S12. As Fig. S11 but without the three sensitivity experiments TraCE-ORB, TraCE-GHG, and FAMOUS.

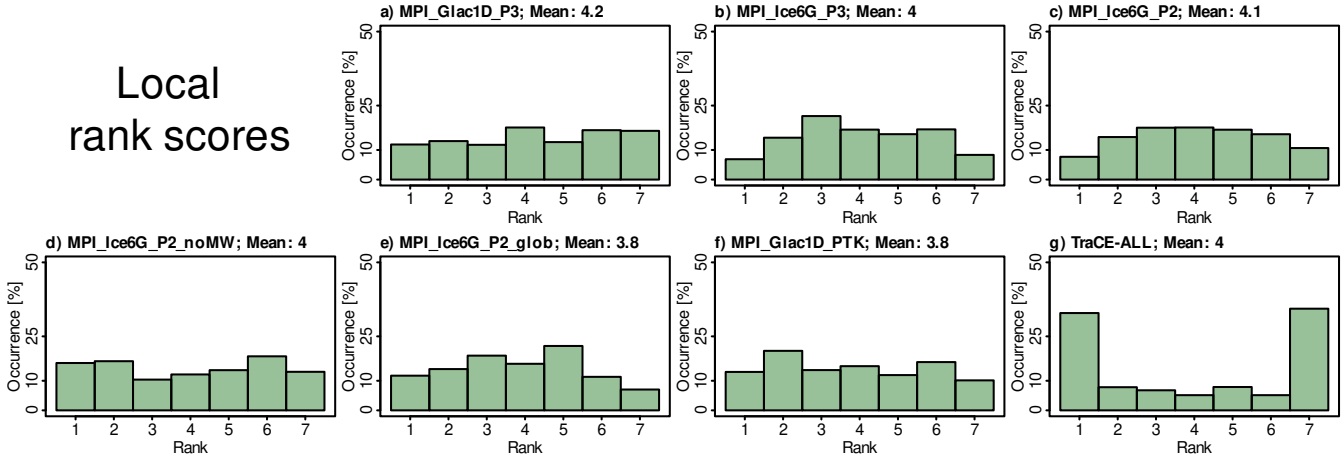

Figure S13. Rank scores for the ensemble of the six MPI-ESM simulations and TraCE-ALL. Rankings are computed for each proxy record and each of the four components. The bars depict the occurrence percentages of the ranks.

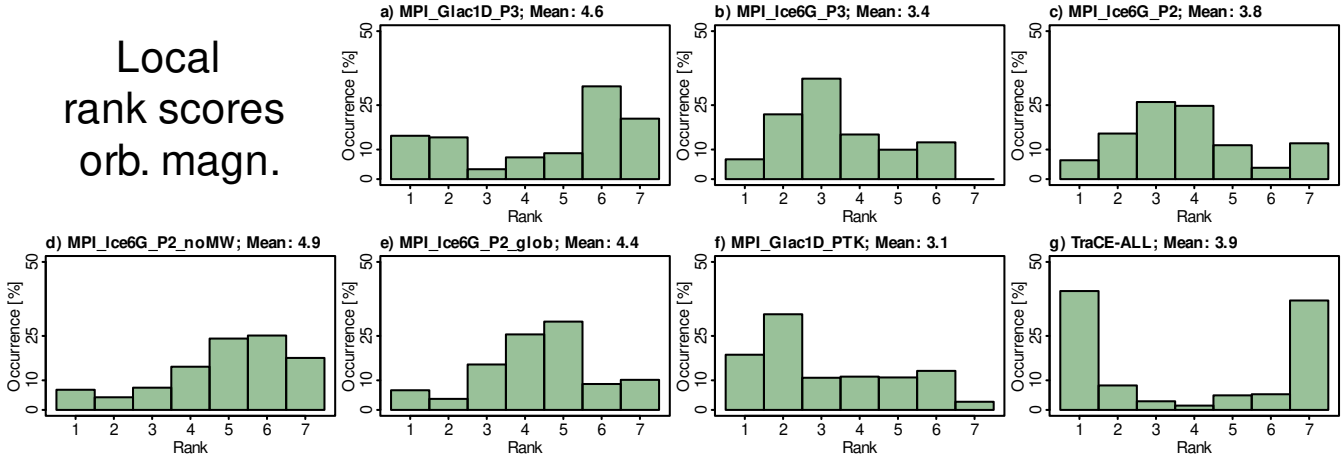

Figure S14. As Fig. S13, but restricted to orbital magnitude IQDs.

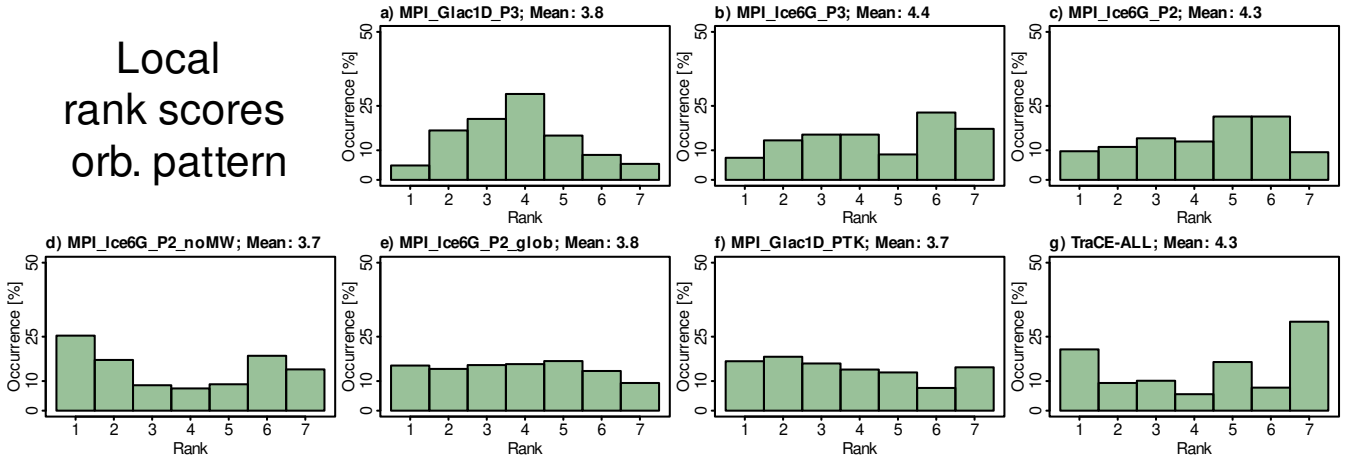

Figure S15. As Fig. S13, but restricted to orbital pattern IQDs.

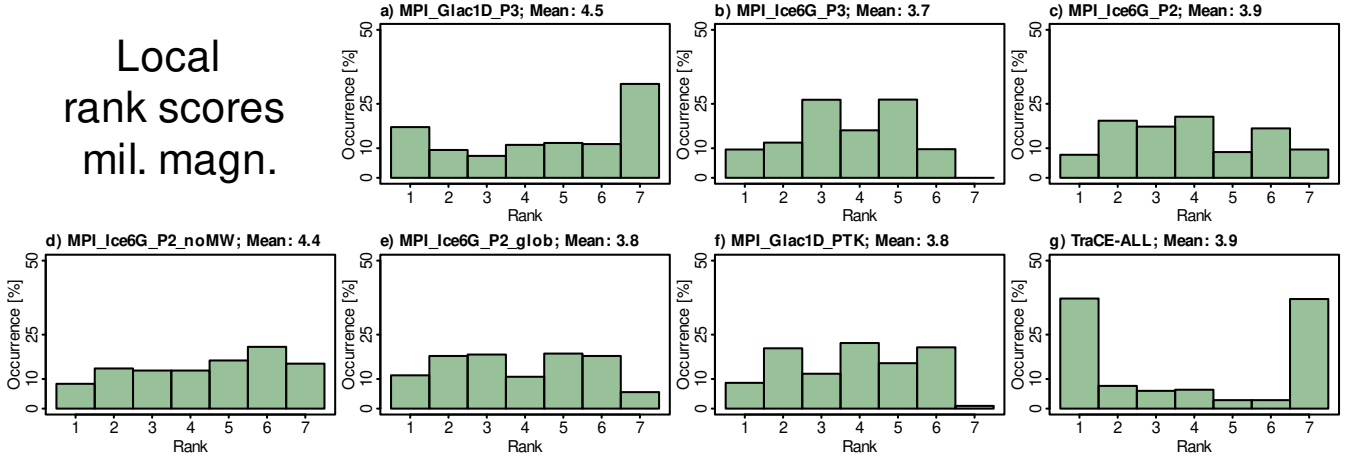

Figure S16. As Fig. S13, but restricted to millennial magnitude IQDs.

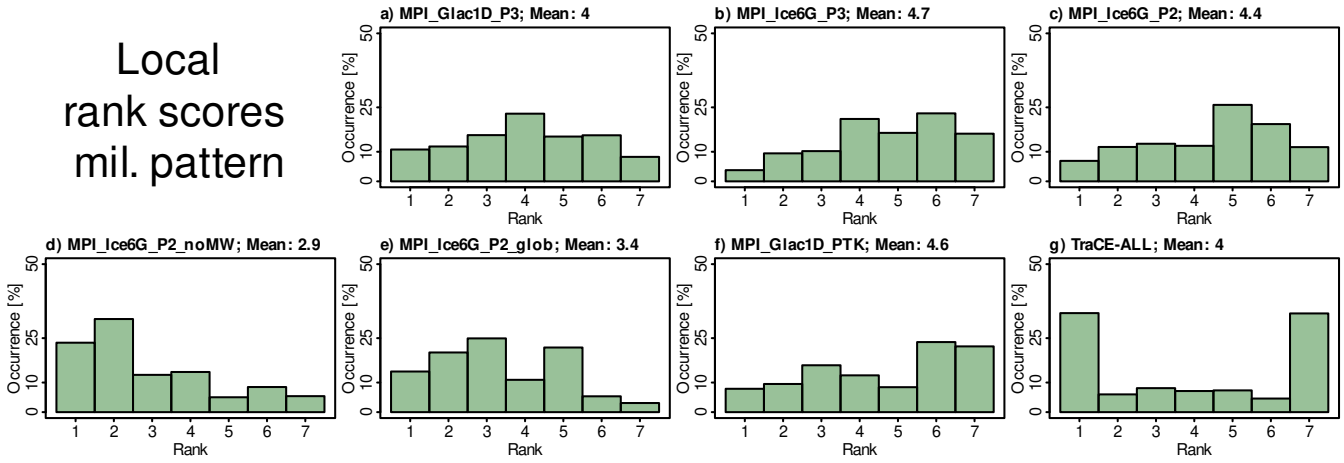

Figure S17. As Fig. S13, but restricted to millennial pattern IQDs.

References

Berger, A.: Long-Term Variations of Daily Insolation and Quaternary Climatic Changes, J. Atmos. Sci., 35, 2362–2367, https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2, 1978.

Berger, A. and Loutre, M.: Insolation values for the climate of the last 10 million years, Quaternary Science Reviews, 10, 297–317, https://doi.org/10.1016/0277-3791(91)90033-Q, 1991.

He, F.: Simulating Transient Climate Evolution of the Last Deglaciation with CCSM3, Dissertation, University of Wisconsin-Madison, 2011.

Jonkers, L. and Kučera, M.: Quantifying the effect of seasonal and vertical habitat tracking on planktonic foraminifera proxies, Clim. Past, 13, 573–586, https://doi.org/10.5194/cp-13-573-2017, 2017.

Jonkers, L., Cartapanis, O., Langner, M., McKay, N., Mulitza, S., Strack, A., and Kucera, M.: Integrating palaeoclimate time series with rich metadata for uncertainty modelling: strategy and documentation of the PalMod 130k marine palaeoclimate data synthesis, Earth Syst. Sci. Data, 12, 1053–1081, https://doi.org/10.5194/essd-12-1053-2020, 2020.

Jonkers, L., Cartapanis, O., Langner, M., McKay, N., Mulitza, S., Strack, A., and Kucera, M.: PalMod 130k marine palaeoclimate data synthesis version 1.1.1, https://doi.org/10.5281/ZENODO.7785766, 2023.

Joos, F. and Spahni, R.: Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years, Proc. Natl. Acad. Sci. U.S.A., 105, 1425–1430, https://doi.org/10.1073/pnas.0707386105, 2008.

Kapsch, M., Mikolajewicz, U., Ziemen, F., and Schannwell, C.: Ocean Response in Transient Simulations of the Last Deglaciation Dominated by Underlying Ice-Sheet Reconstruction and Method of Meltwater Distribution, Geophys. Res. Lett., 49, https://doi.org/10.1029/2021GL096767, 2022.

Köhler, P., Nehrbass-Ahles, C., Schmitt, J., Stocker, T. F., and Fischer, H.: A 156 kyr smoothed history of the atmospheric greenhouse gases CO_2, CH_4, and N_2O and their radiative forcing, Earth Syst. Sci. Data, 9, 363–387, https://doi.org/10.5194/essd-9-363-2017, 2017.

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J.-M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., and Stocker, T. F.: High-resolution carbon dioxide concentration record 650,000–800,000 years before present, Nature, 453, 379–382, https://doi.org/10.1038/nature06949, 2008.

Mix, A.: Chapter 6 - The oxygen-isotope record of glaciation, pp. 111–135, 1987.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlin. Processes Geophys., 18, 389–404, https://doi.org/10.5194/npg-18-389-2011, 2011.

Riddick, T., Brovkin, V., Hagemann, S., and Mikolajewicz, U.: Dynamic hydrological discharge modelling for coupled climate model simulations of the last glacial cycle: the MPI-DynamicHD model version 3.0, Geosci. Model Dev., 11, 4291–4316, https://doi.org/10.5194/gmd-11-4291-2018, 2018.

Smith, R. S.: The FAMOUS climate model (versions XFXWB and XFHCC): description update to version XDBUA, Geosci. Model Dev., 5, 269–276, https://doi.org/10.5194/gmd-5-269-2012, 2012.

Smith, R. S. and Gregory, J.: The last glacial cycle: transient simulations with an AOGCM, Clim Dyn, 38, 1545–1559, https://doi.org/10.1007/s00382-011-1283-y, 2012.

Thorarinsdottir, T. L., Gneiting, T., and Gissibl, N.: Using Proper Divergence Functions to Evaluate Climate Models, SIAM/ASA J. Uncertainty Quantification, 1, 522–534, https://doi.org/10.1137/130907550, 2013.