https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Using ocean surface paleo-density to evaluate PMIP3 and PMIP4 Last Glacial Maximum climate simulations.

Barathieu Héloïse[1], Caley Thibaut[1], Kageyama Masa[2], Swingedouw Didier[1], Braconnot Pascale[2]

[1]Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, 33600 Pessac, France

[2]Laboratoire des Sciences du Climat et de l'Environnement – Institut Pierre-Simon Laplace (LSCE-IPSL), UMR CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France.

Correspondence to: Héloïse Barathieu (*heloise.barathieu@u-bordeaux.fr*)

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Abstract

Quantitative reconstruction of ocean surface density during the Last Glacial Maximum (LGM) offers valuable insights into the ability of climate models to simulate past climate conditions, when global temperatures were about 4.5°C to 6°C colder than today. We assess the performance of the LGM climate simulations, as part of the 3rd and 4th phase of the Paleoclimate Modeling Intercomparisons Project, using a recent ocean surface density reconstruction based on the $\delta^{18}O$ of foraminiferal calcite ($\delta^{18}Oc$). We consider the differences between the LGM and the preindustrial climates and each period separately, at both global and regional scales. Because surface density reflects the combined effects of temperature and salinity, we also examined sea surface temperature (SST) to better identify the processes underlying model–data differences.

Surface density reconstructions show greater variability than simulated surface density. Models therefore struggle to reproduce the spatial variability of the density difference (LGM – pre-industrial (PI)), but part of the mismatch may arise from the uneven spatial distribution of reconstructions, which are mostly located near coastal areas.

Density anomaly (LGM – PI) differences between data and models are largely controlled by sea surface salinity (SSS), with SST contributing to a lesser extent. This influence of SSS is directly linked to the reduction in tropical precipitation during the LGM: models that best match the large-scale density anomalies also simulate the strongest reductions in reconstructed low-latitude precipitation during the LGM, highlighting the key role of hydrological cycle changes in shaping surface density.

On a global scale, 100% of model simulations show a statistically significant relationship with surface density reconstructions, looking at LGM and PI separately. However, on a regional scale, some features are poorly simulated, leading to weaker agreement between data and model simulations, particularly in the North Indian and Southern Oceans. Our analysis concludes with a focus in the Indo-Pacific Warm Pool. Past reconstructions indicate a LGM weakened Indian ocean west–east surface density gradient, but only 7 out of 14 models (50%) reproduce this feature. These results highlight the need to better constrain regional hydrological cycle changes in models, as improving their representation is crucial to reduce uncertainties in both paleoclimate simulations and future climate projections.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

1. Introduction

Past surface seawater density is a key property for studying ocean dynamics, as it reflects the combined influence of surface temperature and salinity and is directly linked to circulation changes through geostrophic balance. In this study, we focus specifically on seawater surface density, providing a novel perspective in model–data comparisons for the Last Glacial Maximum (LGM), a variable that has not been explored in previous assessments.

To simulate future climate change, scientists rely on coupled general circulation models (GCMs). Yet these models differ in their representation of Sea Surface Temperature (SST), Sea Surface Salinity (SSS), and in the processes that control density. Differences in changes of SST and SSS in the future is therefore leading to large uncertainties in the simulation of future ocean dynamics (Flato et al., 2014 – IPCC AR5; Eyring et al., 2021 – IPCC AR6).

Since climate models are developed based on present-day conditions and used to project future climates that may be very different from the present one, it is important to also evaluate them with reconstructed very different climate conditions from the past, to gain confidence in their projections. One way to test the response of these models to various external forcings is to use paleoclimate simulations, which provide an independent evaluation of model performance (Harrison et al., 2014; Kageyama et al., 2024) against available reconstructions. This allows us to benchmark models using evidence from past climates, which is essential for strengthening confidence in their future projections. The Paleoclimate Model Intercomparison Project (PMIP; Joussaume and Taylor, 1995; Kageyama et al., 2018) tests the ability of models to simulate paleoclimate reconstructions. Currently in its fourth phase (PMIP4), with a fifth in preparation, PMIP plays a critical role in evaluating how well models reproduce past climates.

One of the PMIP reference periods is the Last Glacial Maximum, which occurred between 19,000 and 23,000 years ago, when the ice sheets reached their maximum global volume (Mix et al., 2001). During this period, the climate was markedly different from pre-industrial conditions, with significantly colder temperatures (from −4.5 ± 0.9 °C according to Annan et al. 2022 to −6.1 ± 0.4 °C according to Tierney et al. 2020) and altered hydrological cycles, making it an interesting benchmark period for evaluating climate models (MARGO project, 2009; Braconnot et al., 2012).

Evaluating numerical climate models using a data-model comparison allows us to assess their robustness in simulating key variables such as sea surface temperature (SST) and precipitation (Brierley et al., 2023). Recent intercomparison studies focusing on SST reconstructions at the LGM (Tierney et al., 2020; Kageyama et al., 2021) show that models generally capture large-scale cooling patterns but still often exhibit regional biases and differences between PMIP3 and PMIP4 simulations.

To enable quantitative evaluation of ocean surface density, a new method has been developed to reconstruct annual seawater surface densities in the past (Caley et al., 2025), providing a novel tool for model assessment. However, until now, quantitative evaluations of surface density remain unexplored, which limits our understanding of how well models capture the combined effects of temperature and salinity on ocean circulation.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

In this study, we make use of this new surface density reconstruction to evaluate the PMIP3 and PMIP4 simulations in terms of annual ocean surface density, both on a global scale (excluding the Nordic Seas region, Caley et al., 2025) and regionally. We consider simulations of the Last Glacial Maximum (LGM) and the pre-industrial period (piControl). We focus on evaluating model performance in simulating surface density, and where there are large discrepancies between models and past reconstructions, we further investigate SST and, in combination with density, qualitative changes in SSS. This approach provides a complementary perspective to previous studies, offering new insights into the coupled role of temperature and salinity in shaping ocean surface density and allowing for a more integrated evaluation of model performance.

Our analysis is structured as follows: we first assess model simulations against past reconstructions at the global scale in Sect. 3. We then examine regional differences in Sect. 4, identifying areas where models perform better or worse, with a particular focus on the Indian Ocean as a case study of regional variability (Sect. 4.3).

2. Material and methods

2.1. Climate reconstructions

2.1.1. Surface ocean density

To evaluate model simulations on ocean surface density, we use the quantitative past density reconstruction dataset proposed by Caley et al. (2025). They developed a new Bayesian calibration model to calculate the annual surface ocean density using the δ^{18}Oc measurements of several foraminiferal species. Briefly, this probabilistic approach explicitly accounts for inter-species differences and calibration uncertainties, allowing quantitative density reconstructions. New and published δ^{18}Oc datasets were compiled to create an extended database of 474 density reconstructions distributed across all oceanic regions. For each marine sediment core, reconstructions are available for both the LGM and the Late Holocene (LH) (Caley et al., 2025). Analyses from the northern region > 40°N of the Atlantic Ocean were rejected due to potential errors when applying the calibration to the LGM time period (Caley et al., 2025). We thus also exclude this region for the model-data comparison. Surface density is expressed in kg/m³. Throughout this work, values are expressed as anomalies relative to 1000 kg/m³.

Concerning the instrumental observations, we used the version 4.2.2 (analyses.g10, downloaded in 2024) of the EN dataset from the Met Office Hadley Centre (Good et al., 2013), commonly referred to as EN4. This dataset provides quality-controlled ocean temperature and salinity profiles globally, as well as monthly gridded fields derived from objective analyses, covering the period from 1900 to 2022. EN4 is a compilation of temperature and, when available, salinity measurements from various ocean data sources. Surface seawater density, which is a non-linear function of temperature and salinity, was calculated from EN4 temperature and salinity fields using the GSW (Gibbs Seawater) formulation implemented in the gsw Python package (Roquet et al., 2015).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

2.1.2. Sea surface temperature

Since sea surface density and SST are linked, we also performed a data/model comparison in terms of surface temperature. This provides an additional way to investigate the drivers of surface density changes. To do this, we used two previously published SST databases (MARGO Project, 2009; Tierney et al., 2020). These two databases were not combined, as the Tierney dataset includes some MARGO data but with more recent calibrations. Notably, Tierney et al. (2020) also recalibrated all age models using the Marine13 radiocarbon calibration curve and the BACON age modelling software, ensuring a better chronological consistency across records. The temporal periods differ slightly: Tierney's data refer to the LGM and the Late Holocene (LH), while the MARGO data include LGM and "pre-industrial" measurements from the WOA1998 dataset (NODC, Silver Springs, 1998). Following Tierney et al. (2020), we make here the approximation that the Late Holocene is considered representative of the pre-industrial climate state.

The MARGO database (MARGO Project, 2009) contains 821 SST reconstructions based on a diverse range of proxies, including Mg/Ca ratios, $U_k'_{37}$ indices, radiolarians, diatoms, foraminiferal transfer functions, and the tetraether index TEX_{86}. In contrast, Tierney's dataset (Tierney et al., 2020) comprises 244 SST records derived exclusively from Mg/Ca, TEX_{86}, and $U_k'_{37}$ proxies. We do not use the assimilated SST product developed by Tierney et al. (2020), but only the raw SST proxy database published in association with their study. This selection reflects a deliberate choice by Tierney et al. (2020) to exclude assemblage-based proxies such as foraminiferal transfer functions, due to concerns over "no-analogue" assemblages and the lack of Bayesian calibration models, which are central to their probabilistic framework. SST values inferred from $\delta^{18}Oc$ were also excluded from our analysis, as they were already incorporated into our density reconstructions. Finally, both SST and density datasets were re-gridded onto a common 1° × 1° spatial grid, matching the reference grid to which the model simulations were also re-gridded, allowing for a direct comparison.

2.2 Climate model simulations

For this study, we used LGM and pre-industrial (hereafter piControl) climate model simulations from a total of sixteen simulations, including seven from PMIP3 (Braconnot et al., 2012) and nine from PMIP4 (Kageyama et al., 2018) (see Table 1). Two more simulations were excluded due to inconsistencies in salinity data (e.g. unit or formatting issues), making them unsuitable for analysis. Some model simulations share the same piControl but differ by their imposed ice-sheet reconstructions for LGM (e.g., HadCM3-ICE6GC vs. HadCM3-GLAC1D and iLOVECLIM1-1-1-GLAC1D vs iLOVECLIM-1-1-1-ICE-6G-C). We tested whether simulations from the same model (within a PMIP phase or across PMIP3/PMIP4 versions) were too similar to each other. Our analysis showed that all simulations differed in at least one basin and for at least one of the variables (SSS, SST, density). Based on this, we retained all simulations.

The pre-industrial control simulation uses the constant boundary conditions established for 1850 CE (Eyring et al., 2016). It aims to produce a stable quasi-equilibrium climate under 1850 conditions, characterised by the annual cycle (mean and seasonality) and internal variability arising from interactions between Earth system components. This simulation serves as a baseline from which changes in all other experiments are calculated.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

The LGM experimental protocol (Kageyama et al., 2017) considered as boundary conditions the large continental ice sheets, the associated land-sea mask changes, adjustments to ocean salinity (as ice sheets store large volumes of freshwater), and reductions in greenhouse gases. This makes it a challenging experiment for climate models, which explains why only a limited number of modelling groups have performed it. In two simulations (MIROC-ESM and IPSL-CM5A2), the salinity field was not initialized with the +1 psu offset prescribed in the protocol to account for freshwater stored in ice sheets. To ensure comparability across models, we added +1 psu to the LGM salinity of these two simulations before calculating absolute density. Simulations performed with the iLOVECLIM model found the dynamical effect of that + 1 psu to be very small, supporting this direct correction (Caley et al., 2025). For the calculation of surface density changes due to the hydrographic changes in SST and SSS, i.e. corrected for mean ocean density changes related to ocean volume, we removed this +1 psu from the LGM simulations and applied a −0.77 kg/m^3 density correction to the reconstructions, following Caley et al. (2025).

We analysed annual mean SSS and SST from these simulations. All outputs were regridded to a common 1°×1° grid, and monthly data were averaged to obtain climatological annual means. These means were computed over the full duration of each simulation, which ranges from 100 to 1100 years for the piControl experiments and up to 500 years for the LGM experiments.

The variables used here are salinity (so) and temperature (thetao). To study surface density, we used salinity and temperature from the first ocean layer. Seawater density was calculated using the gsw (Gibbs Seawater) Python package, which is based on the TEOS-10 thermodynamic framework and provides thermodynamically consistent equations for seawater properties. The gsw package requires Absolute Salinity (SA) and Conservative Temperature (CT) as inputs. CT is a more accurate and thermodynamically sound analog of Potential Temperature (θ), and SA is derived from Practical Salinity (SP), which is unitless and not directly usable in thermodynamic calculations. Since the model simulations originally provided SP and θ, we first converted them to SA and CT using the functions SA_from_SP and CT_from_pt, respectively. Finally, in-situ seawater density was computed with gsw.rho(SA, CT, p), using a computationally efficient expression for specific volume as a function of SA, CT, and pressure (Roquet et al., 2015).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Model simulation	PMIP/CMIP - Spin-up phase and rip (f)	References
CNRM-CM5	PMIP3-CMIP5 r1i1p1	Voldoire et al. (2013)
CCSM4	PMIP3-CMIP5 r1i1p1	Gent et al. (2011) and Brady et al. (2013)
GISS-E2-R	PMIP3-CMIP5 r1i1p150	Schmidt et al. (2014) and Ullman et al. (2014)
IPSL-CM5A-LR	PMIP3-CMIP5 r1i1p1	Dufresne et al. (2013)
MIROC-ESM	PMIP3-CMIP5 r1i1p1	Sueyoshi et al. (2013)
MPI-ESM-P	PMIP3-CMIP5 r1i1p1	Adloff et al. (2018)
MRI-CGCM3	PMIP3-CMIP5 r1i1p1	Yukimoto et al. (2012)
HadCM3-GLAC1D	PMIP4-CMIP6	Izumi et al. (2023)
HadCM3-ICE6GC	PMIP4-CMIP6	Izumi et al. (2023)
HadCM3-PMIP3	PMIP4-CMIP6	Izumi et al. (2023)
iLOVECLIM1-1-1-GLAC1D	PMIP4-CMIP6	Lhardy et al. (2021) and Bouttes et al. (2023)
iLOVECLIM1-1-1-ICE-6G-C	PMIP4-CMIP6	Lhardy et al. (2021) and Bouttes et al. (2023)
IPSL-CM5A2	PMIP4-CMIP6	Sepulchre et al. (2020)
MIROC-ES2L	PMIP4-CMIP6 r1i1p1f2	Hajima et al. (2020)
MPI-ESM1.2-LR	PMIP4-CMIP6 r1i1p1f1	Mauritsen et al. (2019)
CESM1-2	PMIP4-CMIP6	Tierney et al. (2020)

Table 1: Model simulations available for this study, with LGM and piControl simulations. Model simulations from PMIP3 (blue) and PMIP4 (pink).

2.3 Statistical analysis:

Model–data comparisons were performed on a 1° × 1° grid, restricted to locations where proxy reconstructions provide valid values. Model anomalies were extracted exactly at the proxy sites to ensure strict spatial consistency.

Distributional agreement between models and reconstructions was evaluated using three complementary criteria. First, interquartile range (IQR) overlap was assessed to check distributional consistency while limiting sensitivity to extreme outliers. Second, the Kolmogorov–Smirnov (KS) statistic, which quantifies the maximal difference between cumulative distributions, was compared to basin-specific thresholds to account for differences in sample size: South Indian 0.240 ($n = 35$), North Indian 0.254 ($n = 27$), South Atlantic 0.259 ($n = 26$), North Atlantic 0.224 ($n = 36$), Pacific 0.154 ($n = 75$), and Southern Ocean 0.483 ($n = 7$). Exceeding these thresholds indicates a substantial difference between model and data distributions. Third, a two-sample KS test p-value < 0.05 indicates a statistically significant difference between distributions: a value above 0.05 indicates that the null hypothesis (distributions are similar) cannot be rejected, while $p < 0.05$ indicates a significant difference. For visualization, kernel density estimates (KDEs) provided smoothed, continuous representations of the distributions, highlighting central tendencies and the most frequent values.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

For linear relationships, regression analyses were performed separately for PI and LGM periods. The coefficient of determination (R^2) quantifies the proportion of variance in reconstructions explained by the models, while the slope measures the amplitude of the model response relative to observations. To account for uncertainties in the reconstructions, a Monte Carlo procedure (10,000 iterations) added Gaussian noise to the observations, derived from the 95% confidence intervals. Distributions of R^2 and slope were then analyzed, and reported values correspond to mean ± standard deviation.

This framework provides a consistent and rigorous evaluation of model–data agreement, encompassing both distributional comparisons and linear relationships, and applies globally as well as regionally across ocean basins.

2.4 Testing spatial representativeness with a pseudo-proxy approach

The spatial distribution of reconstructions is uneven, with a clear concentration of data near coastal areas (Fig. 2). To assess whether the proxy data locations are representative of broader basin-scale conditions, a “pseudo-proxy” approach (Ayache et al., 2018) was performed, in order to compare the mean local values with basin-wide means derived from the models (Fig. 1, Fig. S1). This analysis evaluates the spatial representativeness of proxy locations at the basin scale. However, it does not address differences between coastal and open-ocean environments, since climate models do not explicitly represent coastal processes and therefore cannot accurately simulate nearshore dynamics. Details on the basin definition and spatial masks used for this analysis are provided in Appendix A.

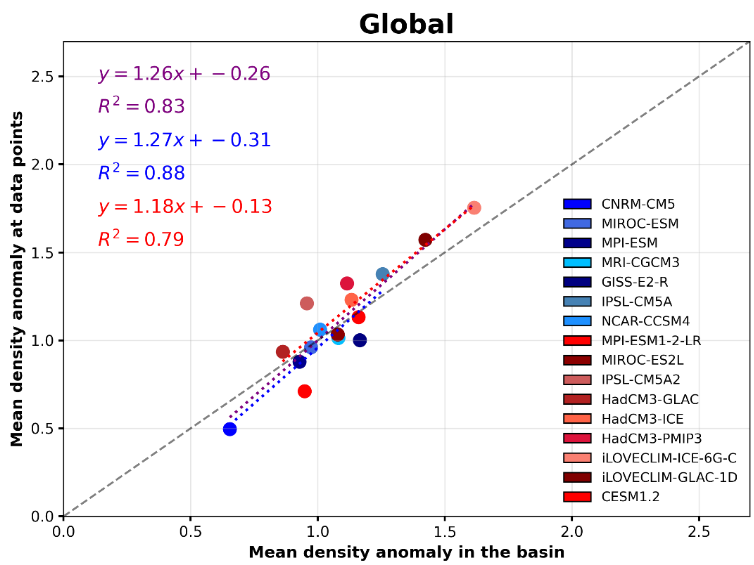


Figure 1: Pseudo-proxy test for global ocean surface density. Comparison between the mean density anomalies (kg/m3) averaged across all model grid points (x-axis) and the mean density anomalies (kg/m3) averaged over proxy reconstruction sites (y-axis). PMIP3 simulations are shown in blue, and PMIP4 simulations in red. The purple regression line and R^2 represent the fit across all model simulations combined (PMIP3 + PMIP4). Results for individual basins are provided in Supplementary Fig.S1.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

For this purpose, we compare the average of density simulated by PMIP3 and PMIP4 model simulations over a given basin with the average density only at locations where proxy data are available (Fig. 1). A near-linear relationship is found at the global scale (Fig.1) and across most basins (Fig. S1), with regression slopes ranging from 0.78 in the Southern Ocean to 1.32 in the South Atlantic, and coefficients of determination (R^2) between 0.40 (South Indian) to 0.98 (North Indian), when compiling PMIP3 and PMIP4. These regressions indicate how well the mean over proxy sites reproduces the true basin-wide mean in the model simulations. The slopes being close to one and the high R^2 values indicate that the mean anomalies at the proxy sites capture the basin-wide means simulated by the models reasonably well. This suggests that, within the models, the uneven proxy distribution does not strongly bias the large-scale signal. However, when analyzing PMIP3 and PMIP4 separately (Fig. S1, individual points in blue and red), significant relationships are not found in the South Indian and Southern Ocean basins. This suggests that the smaller number of models in each ensemble, combined with the limited proxy coverage in these basins, reduces the robustness of the relationship. When pooling PMIP3 and PMIP4 simulations, all basins show statistically significant relationships (p-value < 0.05), confirming that the signal emerges when sample size is increased. These results therefore support using specific proxy locations as representative of larger signal at basin scale.

3. Evaluate model simulations at the global scale

Before investigating regional features, we first assess the large-scale behaviour of model simulations. A global-scale evaluation allows us to assess the overall ability of PMIP3 and PMIP4 simulations to reproduce the reconstructed large-scale signal of surface density changes between the LGM and the pre-industrial period. Evaluating models at this integrated scale also helps reduce the influence of local reconstruction uncertainties and highlights the dominant climatic drivers of density variations, such as global temperature and hydrological cycle changes.

3.1. Sea surface density anomaly (LGM-PI)

We first analyse absolute surface density anomalies between the LGM and the Pre-Industrial period (LGM–PI), in order to reduce the impact of potential systematic model-specific biases that may persist across time periods. All models simulate positive anomalies, i.e. they agree on the sign of the change. However, both the spatial patterns and the amplitude of the simulated anomalies vary considerably from one model simulation to another (Fig. 2). When averaged over space, simulated mean density anomalies typically range from 0.6 to 1.6, with an average value close to 1.0. In contrast, the proxy-based reconstructions exhibit a slightly higher mean, approximately 1.5.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

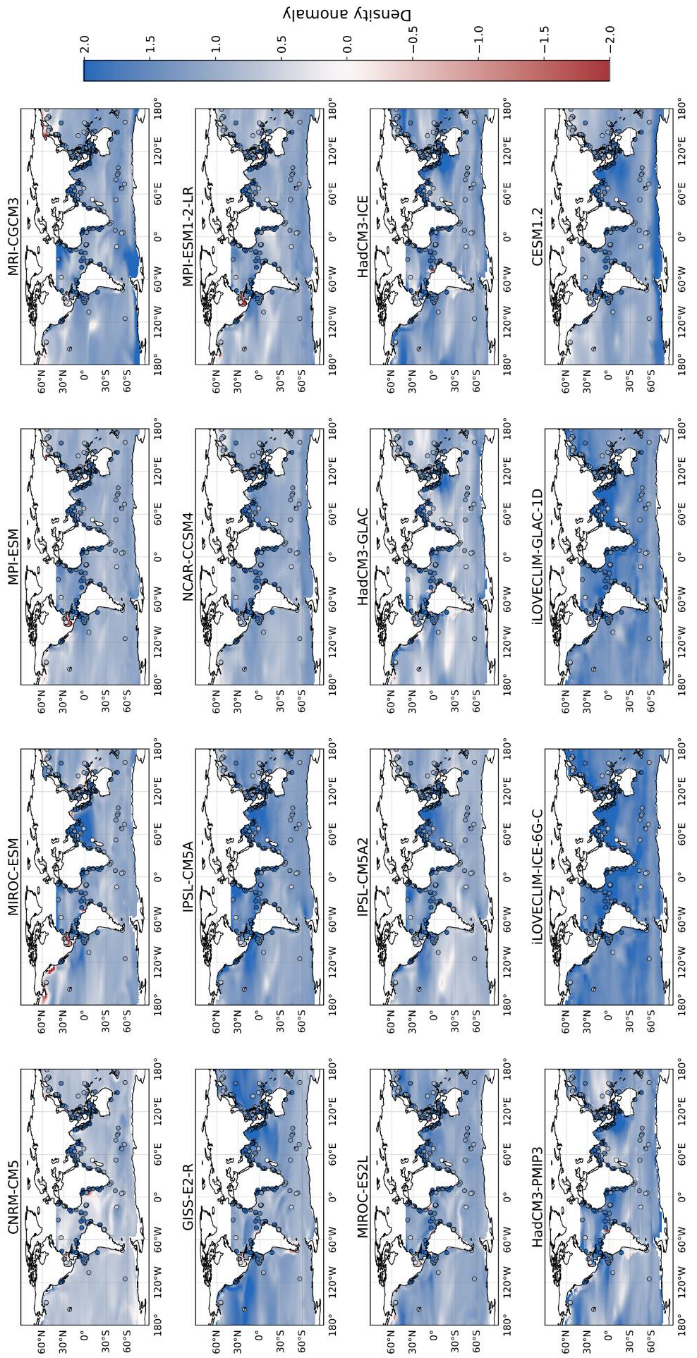
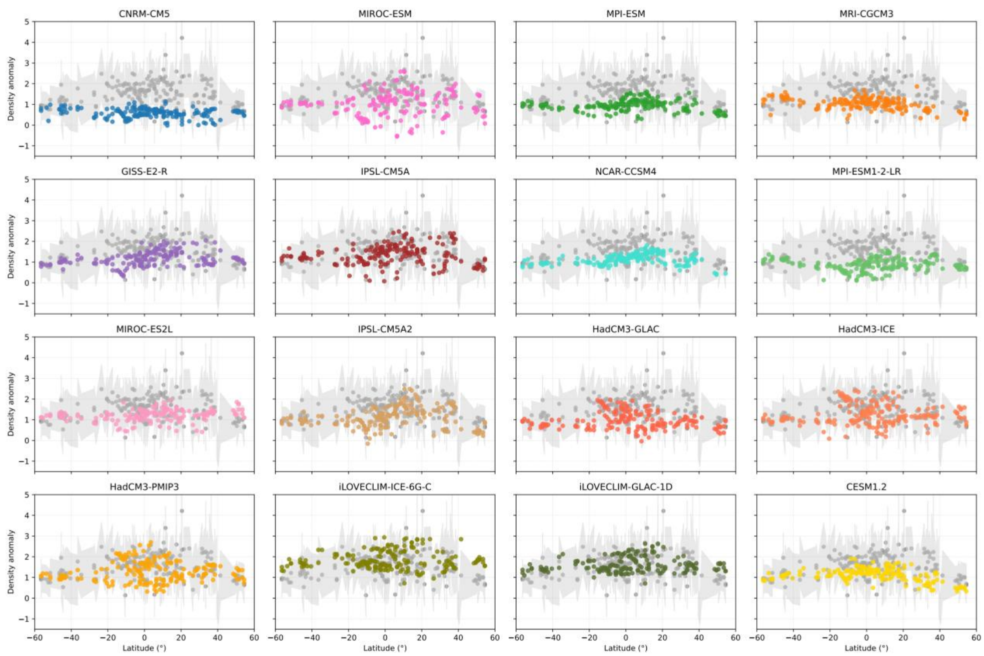


Figure 2: Absolute surface density (kg/m3) anomaly map (LGM - piControl). The dots represent the surface density anomaly database reconstructions (Caley et al., 2025) and the background map is the anomaly mean for each model simulations in the study.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

When zonally averaged, the density anomaly (LGM-PI) in the observations (Fig. 3, grey dots) is stronger in the low latitudes than in the mid-latitudes, as already discussed in Caley et al., 2025. The observed increase in density during the LGM, both in the simulations and in the reconstructions, is consistent with the SST cooling (MARGO project, 2009; Tierney et al., 2020) and with a weaker hydrological cycle at low latitudes, in which precipitation decreased more than evaporation. This reduction in precipitation leads to saltier and denser surface waters (Kageyama et al., 2021), as already discussed in Caley et al., 2025.

Some model simulations fail to reproduce the full latitudinal structure, such as CNRM-CM5 and MPI-ESM-P, while others do not capture the shape but match the density anomaly well at low latitudes, for instance iLOVECLIM-ICE-6G-C and iLOVECLIM-GLAC-1D. Some model simulations, such as IPSL-CM5A and HadCM3-PMIP3, are in good agreement with the data across all latitudes, especially between 0 and 40°N, considering the uncertainties on the reconstructions (Fig. 3). The same type of zonally averaged analysis was performed for SSTs, and the corresponding results are shown in Figs. B2 and B3.



<u>Figure 3:</u> Density anomaly (kg/m3) as a function of latitude for each model simulation (colored dots) compared with the observational data (grey dots) and the 68% confidence interval (grey shading). Model outliers, identified using the interquartile range (IQR) method (values outside 1.5×IQR), were excluded to reduce the influence of extreme values. This filtering highlights the main structure and latitudinal patterns of the modelled density anomalies while retaining all available latitude points.

We next investigated the physical drivers of the density differences between reconstructions and simulations by decomposing the total density anomaly into temperature and salinity components. Model outputs and proxy datasets were collocated at common sampling points without additional interpolation, ensuring a strict one-to-one correspondence between density and SST observations. This procedure yielded 80 common points for the MARGO

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

database (MARGO Project, 2009) and 92 points for the SST dataset from Tierney et al. (2020). No outlier filtering was applied.

Before computing the salinity contribution, the global density related to sea-level–induced salinity increase (−1 g/kg in models, −0.77 kg/m3 in reconstructions; Caley et al., 2025) was removed to isolate density changes linked to hydrographic changes in SST and SSS. Because this correction is applied consistently to both reconstructions and model outputs, and since the analysis is based on model–data differences, it has no effect on the results presented here.

To quantify the relative influence of temperature and salinity on the model–data density anomaly differences, we decomposed the total density anomaly difference (model simulations minus reconstruction) into two components. We first isolated the temperature effect by calculating how density would change keeping salinity fixed to isolate the thermal contribution. The remaining difference, attributed to salinity, was computed as the residual between the total density bias and the temperature-only component.

Figure 4 shows the relative contributions (%) of SST and salinity to the density anomaly for each model. In most models, salinity (red bars) is the dominant contributor to the density anomaly difference between reconstructions and simulations. Only a few simulations, such as HadCM3-GLAC and CESM1.2, exhibit larger SST contributions (orange bars), but these are exceptions rather than the rule. Note that this figure shows relative contributions (%) and does not indicate the absolute magnitude of the total density bias.

Overall, these results demonstrate a clear pattern: SST effects alone are insufficient to explain the observed density anomaly differences between reconstructions and models, whereas SSS differences account for the majority of the model–data discrepancy (Fig. 4) in most simulations. Furthermore, it is important to note that Figure 4 shows an average, which "smooths" the variability observed across latitudes. Looking at Figure B2, we see that salinity anomalies explain the variability in density anomalies. The variability is largely dominated by salinity. Very little variability is observed in SST in model simulations.

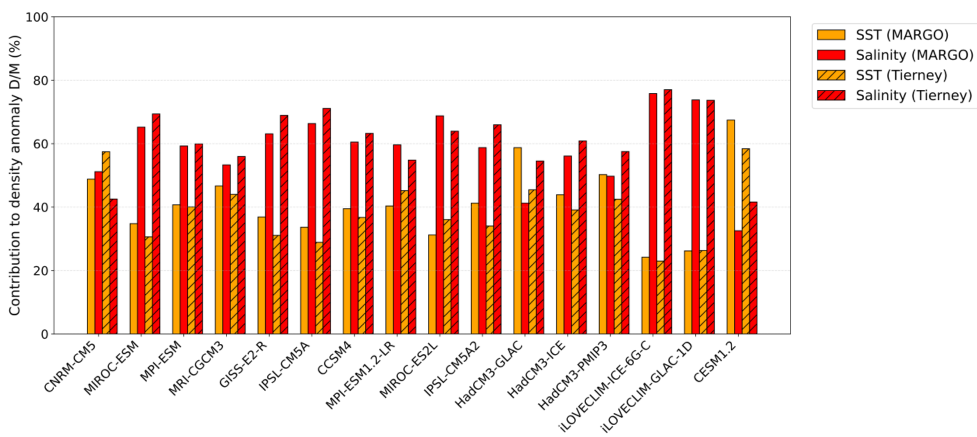


Figure 4: Relative contributions (%) of SST (orange) and SSS (red) to the model-data density anomaly (LGM-PI) differences. Contributions are shown as percentages of the total absolute density difference between model simulations and reconstructions. Solid bars: MARGO SST dataset (MARGO Project, 2009); hatched bars: Tierney et al. (2020) SST dataset. The

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

decomposition was performed by first calculating the temperature-only density anomaly (using model LGM temperatures with PI salinity), then attributing the residual to salinity effects. Density calculations use TEOS-10 GSW routines.

Given that the hydrological cycle influences surface salinity and that salinity anomalies strongly influence the difference in data/model density anomalies, we explored the link between density anomalies and large-scale precipitation changes. Using the PMIP3 and PMIP4 ensembles, Kageyama et al. (2021) showed that nearly all models simulate substantial decreases in precipitation in high-rainfall regions during the LGM, particularly across the tropics and monsoon zones, although the magnitude varies between models. Models that best reproduce the reconstructed density anomalies found in reconstructions also tend to exhibit the largest reductions in tropical precipitation. In contrast, models with smaller precipitation decreases fail to reproduce the observed density structure. Figure 5 illustrates the relationship between density anomalies and mean annual precipitation anomalies for both the tropics (30°S–30°N, circle) and the global ocean (square). A clear linear relationship emerges between density anomalies and precipitation anomalies, with $R^2 = 0.67$ in the tropics and $R^2 = 0.64$ globally, highlighting the robustness of this connection (p-value < 0.05). This analysis emphasizes that accurately representing low-latitude hydrological feedbacks is critical for capturing the full magnitude of glacial ocean density changes.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

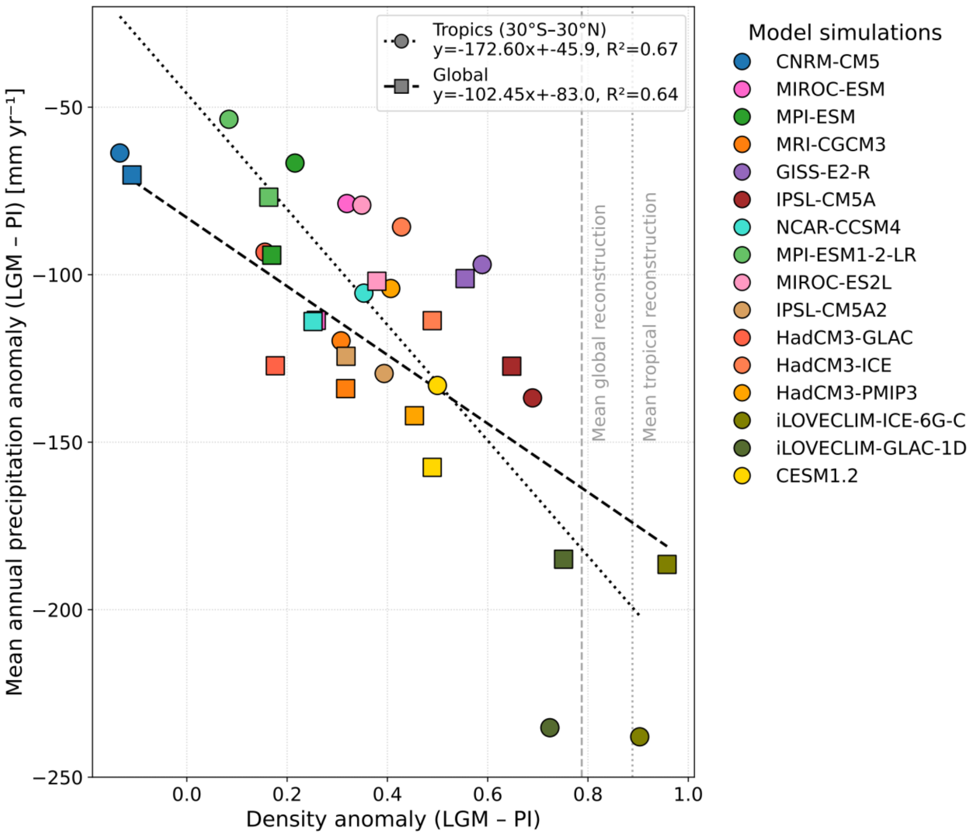


Figure 5: Mean annual precipitation anomalies (LGM – PI, mm.yr^{-1}, over land and ocean) as a function of seawater density anomalies corrected for mean ocean density changes (kg/m3) relative to sea level-induced salinity increase at LGM. Scatter points show the relationship for the tropics (30°S–30°N, circle) and the global ocean (square). Dashed lines indicate linear regressions for each region, with the corresponding slope, intercept, and R^2. Precipitation anomalies are from Kageyama et al. (2021). Grey dashed lines indicate the mean tropical and global reconstructions. No linear relationship is found between SST anomalies and mean annual precipitation anomalies (not shown), indicating that the link between density anomalies and precipitation is primarily driven by salinity changes.

In summary, our results indicate that SST effects alone cannot explain the density anomaly differences between reconstructions and simulations. Instead, salinity differences account for most of the model–data anomaly density discrepancy (Fig. 4) and are directly linked to reductions in tropical precipitation. Models that simulate stronger tropical precipitation decreases reproduce the observed LGM surface density anomalies more accurately, emphasizing the importance of representing low-latitude hydrological feedbacks to capture the full magnitude of glacial ocean density changes.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

3.2. Comparison of global distribution of surface density anomalies

Before analysing regional contrasts, we first evaluate model performances considering the global distribution of absolute surface density anomalies (LGM-PI) by aggregating data from all selected ocean basins (Fig. 6 & Fig. 7, Fig. C1).

This approach allows us to assess whether models reproduce both the magnitude and variability of the reconstructed anomalies at the global scale.

As shown in Section 3.1, reconstruction-based anomalies have a median value of approximately 1.5, whereas the values obtained from model simulations vary widely, ranging from 0.6 to 1.7 (Fig. 6). Interquartile ranges (IQRs) support this divergence: the proxy data exhibit a broader spread (IQR ~0.8), while models show smaller variability (IQR between 0.31 and 0.77), indicating that most models underestimate both the mean anomaly and its variability (Fig. 7). For example, CNRM-CM5 strongly underestimates the median (0.60 vs. 1.57), while iLOVECLIM-ICE-6G-C overestimates it (1.72 vs. 1.58) (Fig. 6 and Fig. 7).

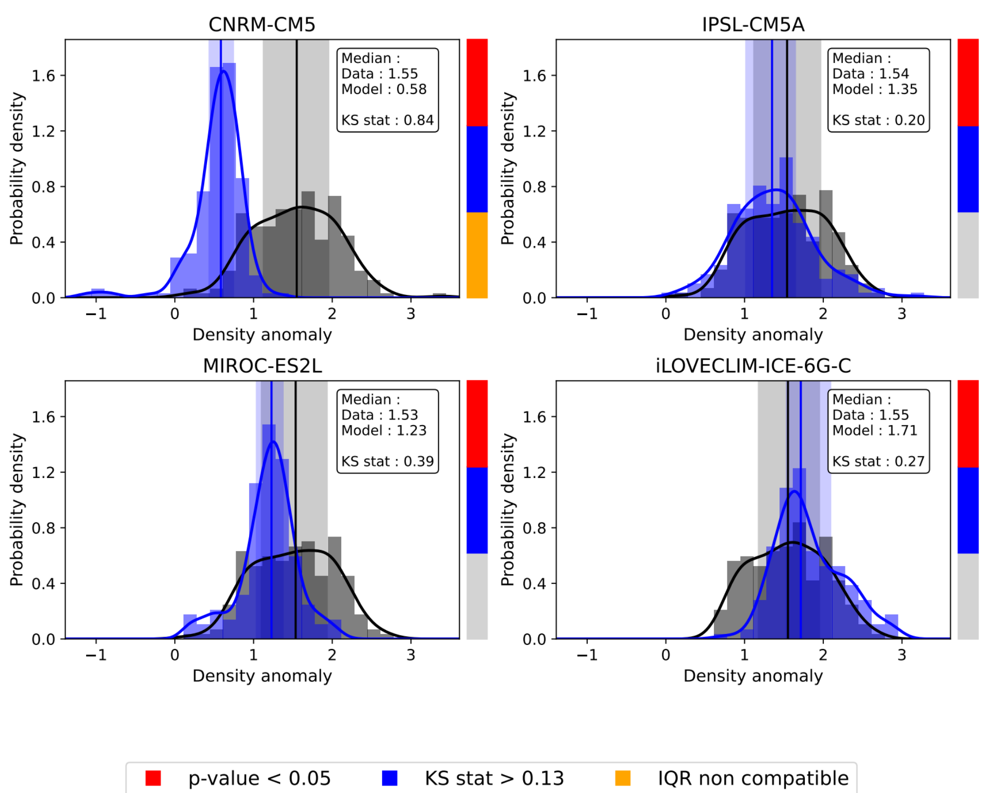


Figure 6: Distribution histograms of surface density anomalies (LGM-PI, kg/m3) for an example of four model simulations. Density reconstructions are shown in black, model simulations in blue. Kernel Density Estimates (KDEs) illustrate the central

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

tendency and overall shape of the distributions. Vertical lines indicate the median of each distribution, and shaded envelopes represent the interquartile ranges (IQRs), providing a measure of data spread that is independent of extreme values. The histograms display the frequency of values, complementing the KDEs and IQRs to give an integrated view of distribution characteristics. Note that uncertainties associated with reconstructions and model heterogeneity are not shown, as the figure focuses on distributional comparison. Colored indicators on the right side of each panel highlight whether the comparison fails to meet any of the following criteria: p-value < 0.05, KS statistic ≥ 0.13, or non-overlapping IQRs between data and model distributions. A colored flag indicates that the corresponding criterion is not satisfied.

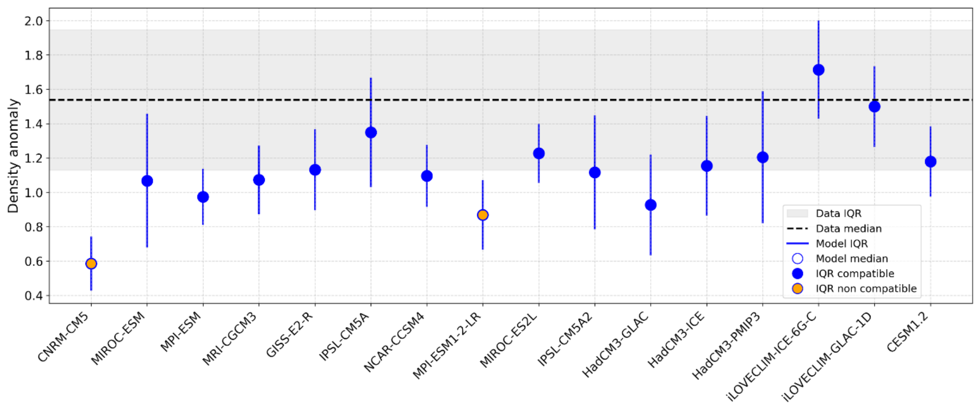


Figure 7: Comparison of observed and simulated density anomaly (kg/m3) statistics at the global scale. The shaded grey band indicates the interquartile range (IQR) of the reconstructions data, and the dashed black line marks the median of the reconstructions. For each model simulation, the blue vertical bars represent the modelled IQR, while the circular markers denote the model median. Blue markers indicate model simulations with IQR values consistent with reconstructions (IQR compatible), whereas orange markers highlight models with larger deviations (IQR non compatible).

Model-data agreement was evaluated using three complementary statistical criteria (see Sect. 2.2): a p-value > 0.05 from a two-sample Kolmogorov-Smirnov (KS) test, a KS statistic below 0.13 and overlapping interquartile ranges (IQRs) between data and model distributions, indicating consistent variability. Together, these criteria evaluate whether models reproduce not only the central value but also the overall shape of the reconstructed distribution.

As shown in Fig. 6 and Fig. 7, 14 of the 16 model simulations (87.5%) satisfy one criterion, and none meet all three simultaneously. In all cases, only the IQR criterion is fulfilled, while KS statistics and p-values remain outside thresholds, confirming that global model-data agreement remains limited.

Because density anomalies depend on both temperature and salinity, we also compared the distribution of SST anomalies between models and reconstructions using the MARGO (MARGO project, 2009) and Tierney et al. (2020) datasets (Fig. S2 and S3). Using the MARGO (MARGO project, 2009) reconstructions, 13 out of 16 simulations (81%) meet the IQR criterion, indicating that model and proxy interquartile ranges overlap substantially despite systematic offsets. None of the simulations meet the KS or p-value criteria, suggesting that although the overall spread of modelled and reconstructed SST anomalies is comparable, their detailed distribution shapes remain statistically different.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

In contrast, for the Tierney et al. (2020) dataset, 87.5% (14/16) satisfy the IQR criterion, and 2 of them also meet the p-value criterion, indicating that a few model simulations (IPSL-CM5A and IPSL-CM5A2) reproduce both amplitude and overall distribution of reconstructed SST anomalies.

Finally, part of the mismatch may arise from the uneven spatial distribution of reconstructions, which are mostly located near coastal areas (Fig. 2). These coastal regions are particularly complex to simulate due to influences such as continental runoff and oceanic upwelling, which are often poorly captured by global climate models. Consequently, kernel density estimates derived from reconstructions appear flatter, suggesting greater variability not captured by the models. Additionally, reconstructions in some key upwelling zones remain problematic, as highlighted by Caley et al. (2025), further complicating the comparison and evaluation of model outputs against data.

3.3. Global evaluation of surface density: models vs. reconstructions (PI & LGM)

One limitation of working with anomalies is that any observed difference between data and models in absolute surface density cannot be directly linked to either the LGM or the PI baseline. To address this, we analysed the two periods separately, the LGM and the PI, as shown in Fig. 8 and Fig 9. For each period, model surface density values were extracted at the grid points corresponding to the reconstruction sites, allowing a direct model–data comparison. In these figures, we compare the surface density values in the simulations to the surface density values in the reconstructions, for both LGM and PI, at the global scale (all basins combined). If the data–model agreement were perfect, a 1:1 linear relationship would be expected (see Sect. 2.2 for statistical details).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

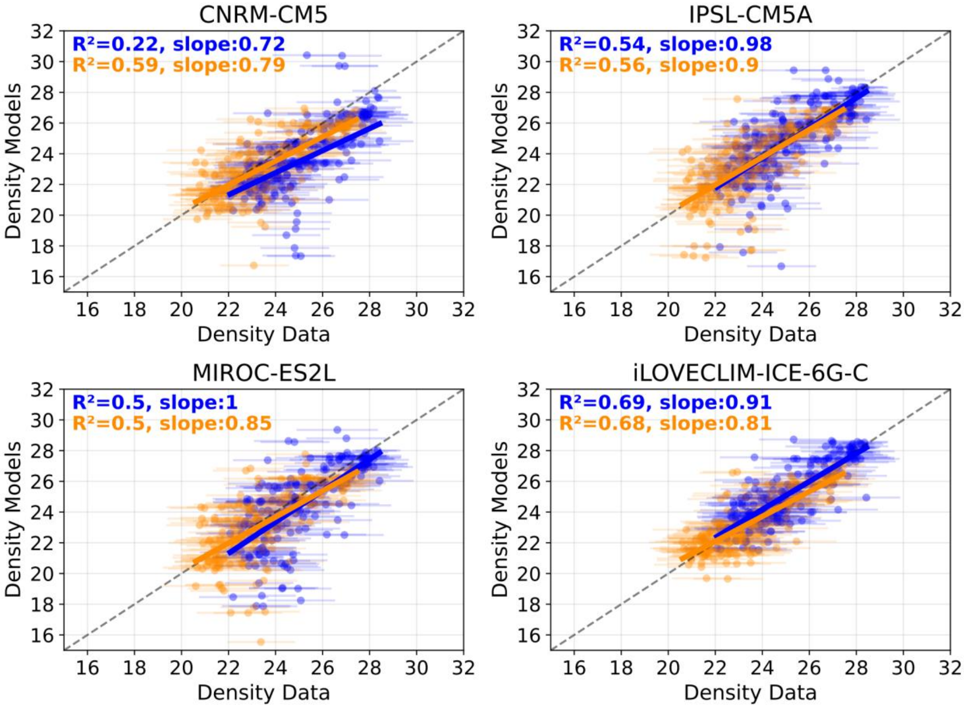


Figure 8: Linear regressions between absolute surface density (kg/m3) from proxy-based reconstructions (x-axis) and model simulations (y-axis), aggregated at the global scale (across all selected basins) for an example of four simulations. Results are shown for the LGM period (blue) and the piControl period (orange). Error bars on the x-axis represent the 95% confidence intervals of the reconstructed values. The slope and R^2 values correspond to standard linear regressions, without accounting for uncertainties on the x-axis (the Monte Carlo method was not applied here). All regressions shown are statistically significant ($p < 0.05$).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

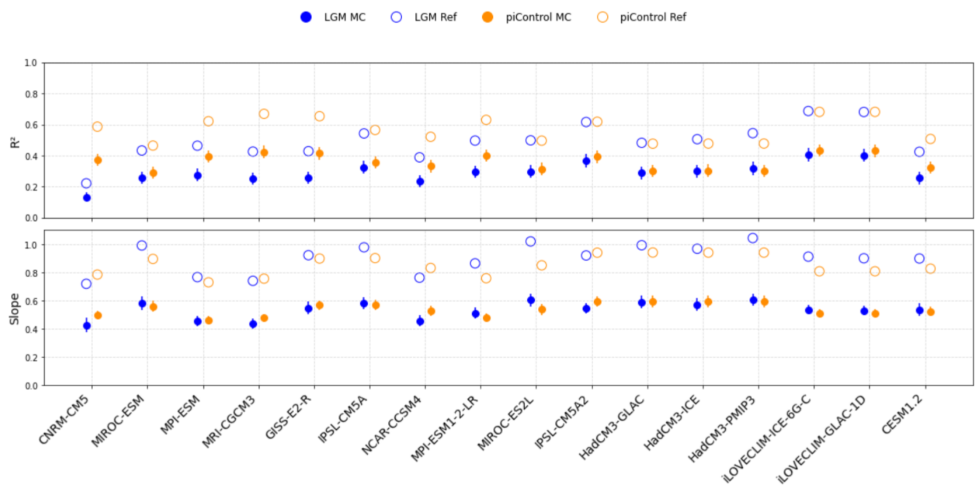


Figure 9: Coefficients of determination (R^2) and regression at the global scale, for each model simulations, evaluated separately for the LGM (blue) and PI (orange) periods. Values are shown for both standard (Ref, empty circles) least-squares regression and uncertainty-aware estimates using Monte Carlo (MC) simulations (n = 10,000 iterations), which propagate reconstruction uncertainties. The error bars for MC estimates represent variability across iterations (±1σ). All regressions shown are statistically significant ($p < 0.05$).

Across all simulations, the regressions are statistically significant (p-value < 0.05), both for standard regressions and for those including uncertainties through Monte Carlo (MC) simulations. This confirms the existence of a strong and robust relationship between reconstructed and simulated surface densities at the grid points where reconstructions are available, for PI and LGM periods.

Among these significant regressions, the R^2 value and slope are then used to identify the model simulations that best match the reconstructions. During the LGM (blue empty circles in Fig. 9), 8 out of 16 simulations (50 %) have R^2 values above ~0.5, and 3 simulations (~19 %) exceed 0.6. In terms of the slope, 75 % of the LGM simulations (12/16) fall between 0.8 and 1.2, indicating good reproduction of the amplitude of the reconstructed density changes. We also note in Fig. 8 and Fig. D1 a shift at the LGM for certain simulations (e.g., intercept not at 0, closer to −1), indicating that some models do not have sufficient density compared to the data. This is particularly evident in CNRM-CM5, MIROC-ES2L and HadCM3-ICE simulations. For the piControl period (orange empty circles on Fig. 9), the proportions are ~75 % for $R^2 > 0.5$ and ~75 % for slopes within the 0.8–1.2 range.

When examining SST, model–data agreement is generally good with the MARGO dataset (Fig. S4 (a)). However, comparison with the Tierney et al. (2020) dataset (Fig. S4 (b)) suggests that some models are slightly too warm. SST biases may therefore explain part of the shift, but the dominant contribution likely comes from salinity, through biases in the representation of the hydrological cycle.

However, in areas with lower density (between 21 and 25 in the data), some model simulations (e.g. MIROC-ES2L, Fig. 8) tend to underestimate density, showing values between 16 and 20 in most cases. This underestimation is also observed during the PI period. In these

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

areas, not all models exhibit this behaviour, and show highly divergent results, leading to a large inter-model spread.

The LGM global comparison shows a statistical agreement between model simulations and proxy-based reconstructions. While all regressions are significant, R^2 and slopes highlight the models with the best performance: 50 % of the models achieve R^2 above 0.5, and 75 % display regression slopes between 0.8 and 1.2. This indicates that most simulations capture the general variability and amplitude of surface densities reasonably well, when looking the LGM and the PI period separately. It is therefore important to examine whether this agreement persists at the regional scale.

4. Evaluate model simulations at regional scale

Despite significant global-scale correlations, the mismatch in the statistical distributions between models and reconstructions reveals that key spatial patterns are not captured uniformly. This motivates a basin-by-basin analysis to better understand the regional origins of these discrepancies.

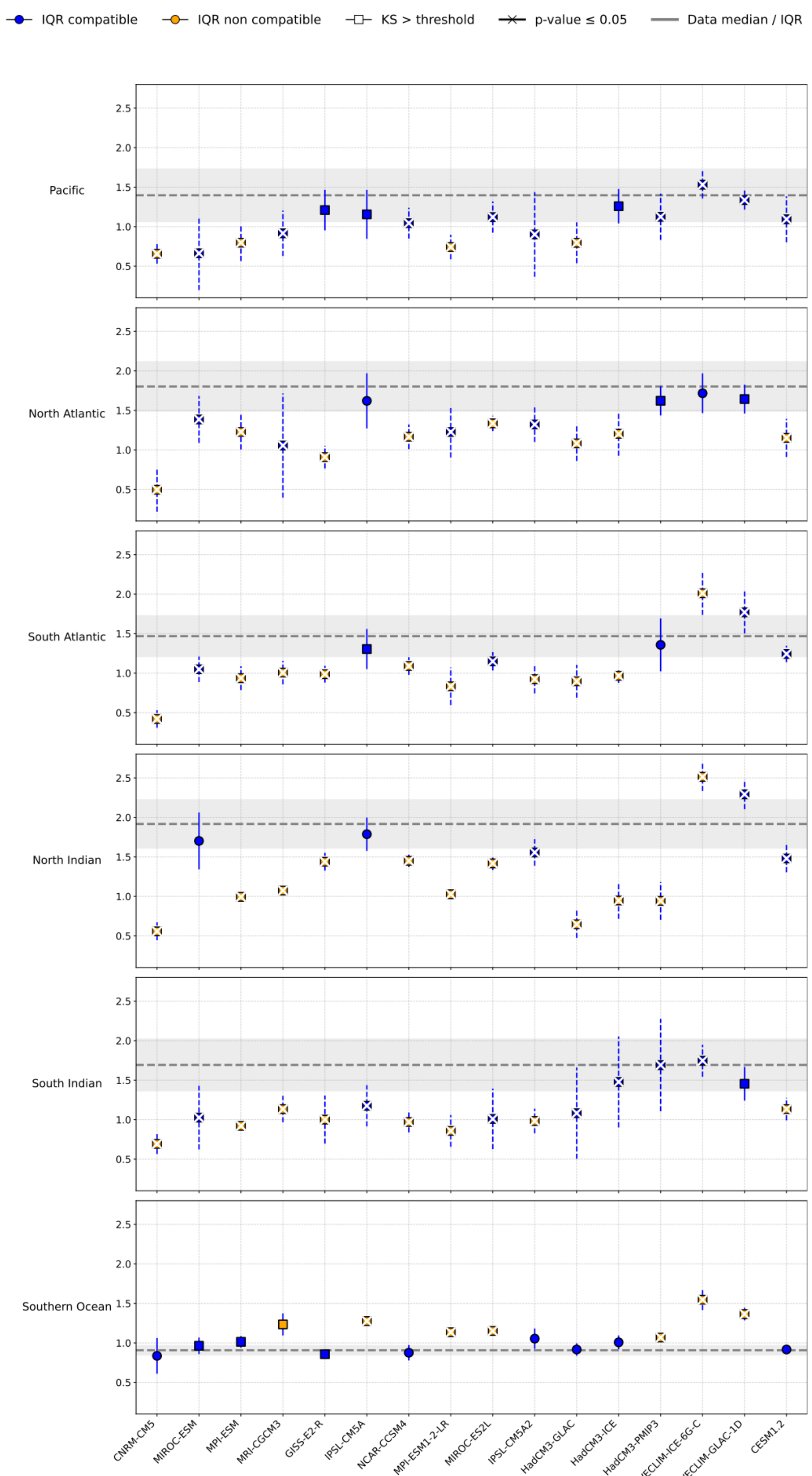
4.1. Comparison of regional distribution of surface density anomalies

To assess whether model simulations differ significantly from proxy-based reconstructions in terms of absolute surface density anomalies (LGM – PI), we performed statistical comparisons across six ocean basins, based on three complementary criteria (see Sect. 2.2).

Across all basins, a total of 96 model–data comparisons were performed (16 model simulations x 6 basins) (Fig. 10). Among them, 49 comparisons (51%) satisfy at least one of the three criteria, while only 11 comparisons (11%) meet all three criteria simultaneously, indicating that strong regional agreement remains rare. Overall, the most robust multi-basin performance is achieved by IPSL-CM5A, while other model simulations show mixed or basin-dependent skill.

At the basin scale, the Southern Ocean shows the most encouraging agreement. Multiple models (e.g., CNRM-CM5, CESM1.2) pass at least one test, and notably, 6 of the 11 cases meeting all three criteria occur in this basin, reflecting relatively high model-data consistency. However, the Southern Ocean is also characterized by sparse data coverage and a limited number of observations. By contrast, the North Indian and South Atlantic basins display systematic mismatches across nearly all models with very few passing any of the criteria (2-5 model simulations depending on the test), reflecting broader difficulties in reproducing regional distributions in these areas. The Pacific, North Atlantic and South Indian basins show intermediate performance, with some models capturing IQRs but fewer passing KS or p-value tests.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026



553

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Figure 10: Comparison of model simulations medians, of surface density anomaly (kg/m3), with observational data at the regional scale. For each ocean, the grey shaded band represents the observed median ± IQR, with the dashed line indicating the median. Model medians are shown as colored markers: blue circles indicate model simulations with IQR compatible with observations, while orange circles indicate IQR not compatible. Squares mark model simulations with a KS statistic above the basin-specific threshold, and crosses denote models with significant differences from observations (p-value ≤ 0.05). Vertical lines show the model IQR, with dashed lines highlighting cases where distributions are statistically similar (p-value > 0.05).

Overall, these results suggest that only a minority of current PMIP3 and PMIP4 model simulations reproduce proxy-based regional distributions of absolute surface density anomalies with acceptable skill, define here as simultaneously satisfying all three criteria. In our dataset, only 11 out of 96 model-basin comparisons (~11%) meet all three conditions, highlighting the difficulty of reproducing past regional ocean dynamics. However, some model–data differences may reflect spatial sampling biases, as proxy sites are mainly coastal while model outputs represent open-ocean conditions, and in several regions, simulations remain close to observational ranges despite marginal statistical mismatches.

4.2. Regional evaluation of surface density: models vs. reconstructions (PI & LGM)

To test whether the strong global agreement between model simulations and proxy-derived surface density truly holds at the regional scale, we repeated the regression analysis of Sect. 3.3 for each model simulation and for each of the six oceanic basins, producing a total of 96 model-period regressions. The complete results, including Monte Carlo (MC) uncertainty estimates, are provided in the Supplementary Material (Fig. S5 and Table S1). For clarity, only regressions with statistically significant R^2 or slope ($p \leq 0.05$ from the standard analysis) are shown, ensuring that the supplementary tables summarize only robust model-data relationships.

In the South Indian Ocean, agreement is consistently strong for both periods. During the LGM, all 16 model regressions are statistically significant, with a mean R^2 of 0.82 ± 0.05, MC R^2 of 0.71 ± 0.06 (Fig. 11), and all slopes falling within the acceptable range of 0.8–1.2 (Table S1). The region emerges as the most reliably simulated basin. However, a shift is observed in some simulations, with simulated densities being too low compared to the observations. SST regressions (not shown) do not reveal trends large enough to explain this shift, suggesting instead a bias in salinity linked to the hydrological cycle.

The Pacific Ocean also exhibits consistently good agreement. All regressions are statistically robust, with MC R^2 exceeding 0.5 in 84.4% of the cases (mainly during the PI) (Table S1). However, only 59.4% of simulations produce slopes within the 0.8-1.2 range. For the LGM, all 16 regressions are significant, with mean R^2 0.57 ± 0.21 (Fig. 11). This indicates that spatial patterns are generally well reproduced, but the magnitude of reconstructed density anomalies is not always correctly captured. The Pacific can nevertheless be considered one of the more reliably simulated regions.

The South Atlantic, shows mixed performance. All regressions are statistically robust, but low R^2 values suggest that poorly reproduced points along the African coast reduce the overall fit. For the LGM, 6 model simulations out of 16 (37.5%) achieves $R^2 > 0.5$. Regarding slope, 62.5% of simulations (10/16) fall within the acceptable 0.8–1.2 range, compared to 56.25% (9/16) during the PI (Table S1).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

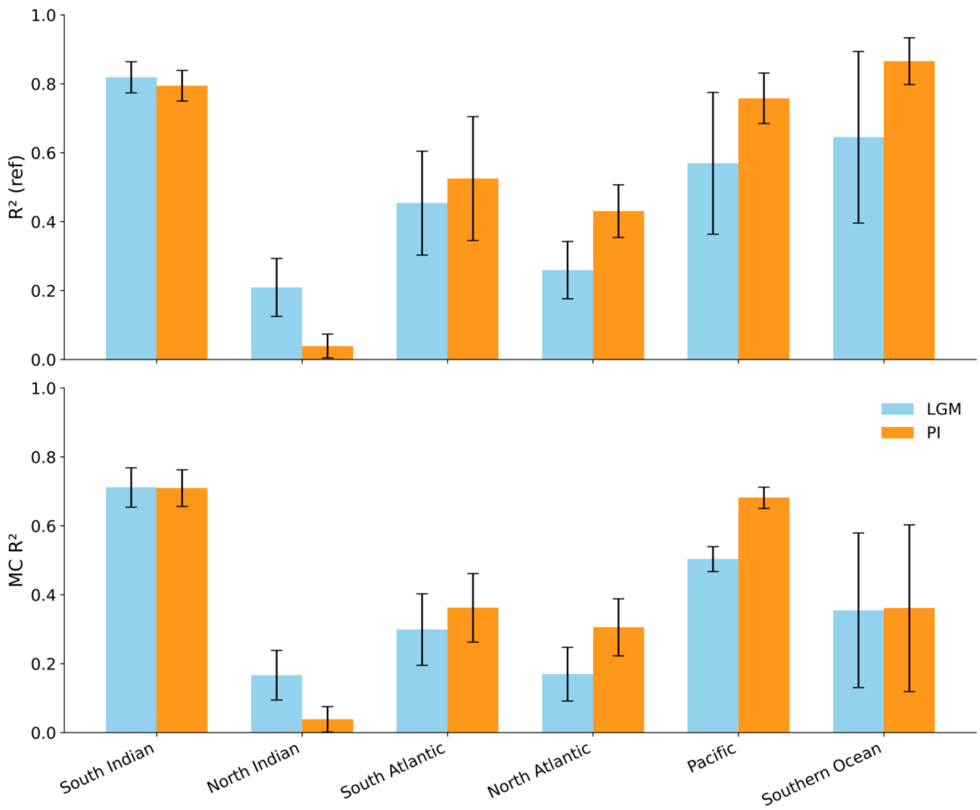
The North Atlantic shows moderate agreement. All regressions are statistically robust, yet only 2 model simulations achieve $R^2 > 0.5$ for the PI, none for the LGM. However, 53% of simulations display slopes between 0.8 and 1.2 (Table S1). During the LGM, the mean R^2 is only 0.26 ± 0.08, MC R^2 0.17 ± 0.08. This indicates that while some models reproduce the magnitude of density, spatial coherence with the reconstructions remains weak.

In contrast, the North Indian Ocean shows poor agreement between surface density model simulations and reconstructions. Fewer than half of regressions are statistically robust, most of them corresponding to the LGM (Table S1). During the LGM, 12 regressions are significant, yet the mean R^2 is only 0.21 ± 0.08, MC R^2 0.17 ± 0.07 (Fig. 11), and only 4 models show slopes within the 0.8–1.2 range. For the remaining cases, R^2 values are close to 0.1, reflecting very weak spatial coherence. Weak agreement is also observed for temperature reconstructions (not show). This basin therefore stands out as one of the least reliably simulated regions.

The Southern Ocean exhibits complex and limited agreement. Standard regressions for the LGM yield a mean R^2 of 0.64 ± 0.25 (Fig. 11), suggesting reasonable fits in some cases, but the corresponding MC-adjusted R^2 drops to 0.35 ± 0.22, highlighting the high uncertainty due to sparse observational coverage. In addition, the number of regressions that remain statistically significant (p-value ≤ 0.05) decrease sharply when considering the Monte Carlo p-values, further emphasizing the lack of robust fits. Only 3 of the LGM slopes fall within the acceptable 0.8–1.2 range. For the PI, standard R^2 values are higher (0.87 ± 0.07), yet MC R^2 remains low (0.36 ± 0.24) (Fig. 11). These results indicate that, despite apparently good fits in some regressions, the Southern Ocean's performance is not robust, and the limited observational constraints dominate the MC uncertainty rather than systematic model biases.

To summarize, across the 16 simulations and six oceanic basins, 83 % of the regressions are statistically significant when considering both LGM and PI periods. In detail, 92 % of LGM simulations (88/96) and 75 % of PI simulations (72/96) show p-values below 0.05 (Table S1), confirming a generally strong and robust relationship between reconstructed and simulated surface densities at the regional scale. When further considering model performance in terms of both $R^2 > 0.5$ and slopes within the 0.8–1.2 range, only 33.3 % of regressions during the LGM (32/96) and 33.3 % during the PI (31/96) meet these criteria, with the majority of successful regressions occurring in the South Indian Ocean (Table S1). The North Indian Ocean and Southern Ocean remain the most challenging regions to simulate, but for different reasons: in the North Indian Ocean, models struggle to reproduce spatial coherence, whereas in the Southern Ocean, the lack of robust fits is primarily due to limited data coverage. These same basins also exhibit the poorest agreement in SST regressions for both the MARGO (MARGO project, 2009) and Tierney et al. (2020) reconstructions (not shown here), suggesting that model biases in these regions are at least partly related to temperature misrepresentation.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026



<u>Figure 11:</u> Average R^2 of model–data regressions by ocean and period. Top panel shows the mean R^2 (ref) across all climate model simulations for each oceanic basin, with standard deviation as error bars. Bottom panel shows the mean Monte-Carlo R^2 (MC R^2) and its standard deviation. Blue bars correspond to the Last Glacial Maximum (LGM) period, and orange bars correspond to the pre-industrial (PI) period. Higher R^2 values indicate better agreement between model simulations and proxy data.

To further explore regional model–data agreement and better understand the sources of spatial discrepancies highlighted in the previous section, we focus on the Indian Ocean as a case study. This basin is a key component of the global climate system, in particular because of its strong connection with the monsoon hydrological cycle, the influence of large river fluxes on salinity, and the resulting impact on surface density which depends on both temperature and salinity.

4.3. Evaluation of models at regional scale: focus on the Indian Ocean

The Indian Ocean region, including part of the Indo-Pacific Warm Pool (IPWP), plays a critical role in the global climate system due to strong coupling between ocean and atmosphere, particularly through the Indo-Pacific Walker circulation. This circulation influences the zonal distribution of SST and thermocline depth, providing the background conditions for phenomena such as the Indian Ocean Dipole (IOD) (Saji et al., 1999; Abram et al., 2020). The

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Indian Ocean is also closely linked to the monsoon hydrological cycle, with precipitation patterns affecting surface salinity and, consequently, density, making it particularly sensitive to both temperature and salinity variations. However, climate models often misrepresent the mean state of the IOD, leading to biases in simulating climate variability (Weller and Cai, 2013; Cai and Cowan, 2013). Notably, Abram et al. (2020) report that many models produce an overly strong thermocline-SST feedback due to a misrepresentation of the mean state, particularly an exaggerated zonal thermocline slope, which artificially increases the strength and frequency of simulated IOD events. Although research on the IOD has expanded over the past two decades, uncertainties remain regarding the controls and long-term evolution of the Indian Ocean's mean state, especially under different climate boundary conditions. The limited timeframe of the instrumental record and persistent model biases make paleoclimate reconstructions essential for investigating past mean states and for testing model performance beyond the range of modern variability (Abram et al., 2020). To date, paleoclimate data of SST suggest that past periods, such as during the LGM, mid-Holocene or 17th century, tend to have a mean state that is more typical of a positive IOD-like, which is systematically associated with elevated IOD variability. This indicates a tight coupling between the mean state and interannual dynamics (Abram et al., 2020).

We now focus on the Indian Ocean sector to assess the degree of model-data agreement in the reconstructed west-east surface density gradient. SST and precipitation driving SSS changes create a west-east surface salinity gradient (Fig. 12) and therefore a density gradient. Recent observations show that, at interannual timescales, SSS variability is strongly linked to the IOD and ENSO, and can interact with SST anomalies through a SST–precipitation–SSS feedback. This suggests that SSS may at times amplify rather than offset SST-related changes (Zhang et al., 2016).

In this section, the 'iLOVECLIM-ICE-6G-C' and 'iLOVECLIM-GLAC-1D' simulations have been excluded due to a salinity bias in the northern Indian Ocean region caused by excess precipitation that reduces salinity in the iLOVECLIM model in comparison to reconstructions as shown by Roche and Caley (2013).

We focus on the LGM mean state of the Indian Ocean, using a West-East gradient. For this, two boxes were defined: "Indian West", which contains 6 density reconstructions and "Indian East", which contains 5. These boxes were chosen based on the definition of the IOD by Saji et al. (1999), and slightly adjusted to ensure sufficient points within each box. The "Indian West" box is defined as 50°E-70°E, 12°S-12°N, corresponding to the tropical western Indian Ocean. The "Indian East" box is defined as 90°E-110°E, 10°S-2°N, corresponding to the tropical south-eastern Indian Ocean. For model–data comparisons, model values are extracted at the exact grid cells where reconstructions are available. The SST points used by Tierney et al. (2020) in the "Indian East" box are geographically close to the density reconstructions, which favours a tighter spatial match between SST and density data. In contrast, reconstructions from the MARGO (MARGO project, 2009) database in the "Indian West" box are located farther offshore, showing less spatial overlap with the available density reconstructions.

Before analyzing the LGM–piControl difference, we verified that models reproduce the West–East gradient observed during the pre-industrial period. To do this we compared the West–East gradient of SST, salinity, and surface density from each piControl simulation with EN4 observational data (1900–1999) extracted at the same grid cells as LGM reconstructions (Fig. E1). We find that all model simulations agree with the observations in terms of the gradient's sign: a negative West–East gradient for temperature, and positive gradients for salinity and density.

Temperature shows greater inter-model spread, while for salinity and density, the model median almost matches the gradient value from EN4 observations (Fig. E1). The proxy-based reconstructions West–East density gradient closely matches the EN4 observations, confirming consistency between modern observations and Late Holocene reconstructions. Regarding SST, we find a significant difference of around 1°C between the EN4 SST and the Tierney et al. (2020) SST reconstructions in the same locations (Fig. E1). Thus, using the West–East surface density gradient as a temporal difference (LGM–piControl) allows model evaluation to focus on LGM performance, as piControl simulations are consistent with observations.

We examine the surface density difference between the LGM and PI periods across the two boxes (Fig. 12 (a)). The proxy-based reconstruction shows a negative west–east density anomaly, with a value close to -1 (Fig. 12 (a)). About half of the model simulations reproduce the correct sign, with CNRM-CM5, MRI-CGCM3, GISS-E2-R, HadCM3-GLAC1D, HadCM3-ICE-6G_C, HadCM3-PMIP3 and CESM1.2 falling within the 68% uncertainty range of the reconstruction. HadCM3-ICE-6G_C and CESM1.2 best capture both magnitude and sign.

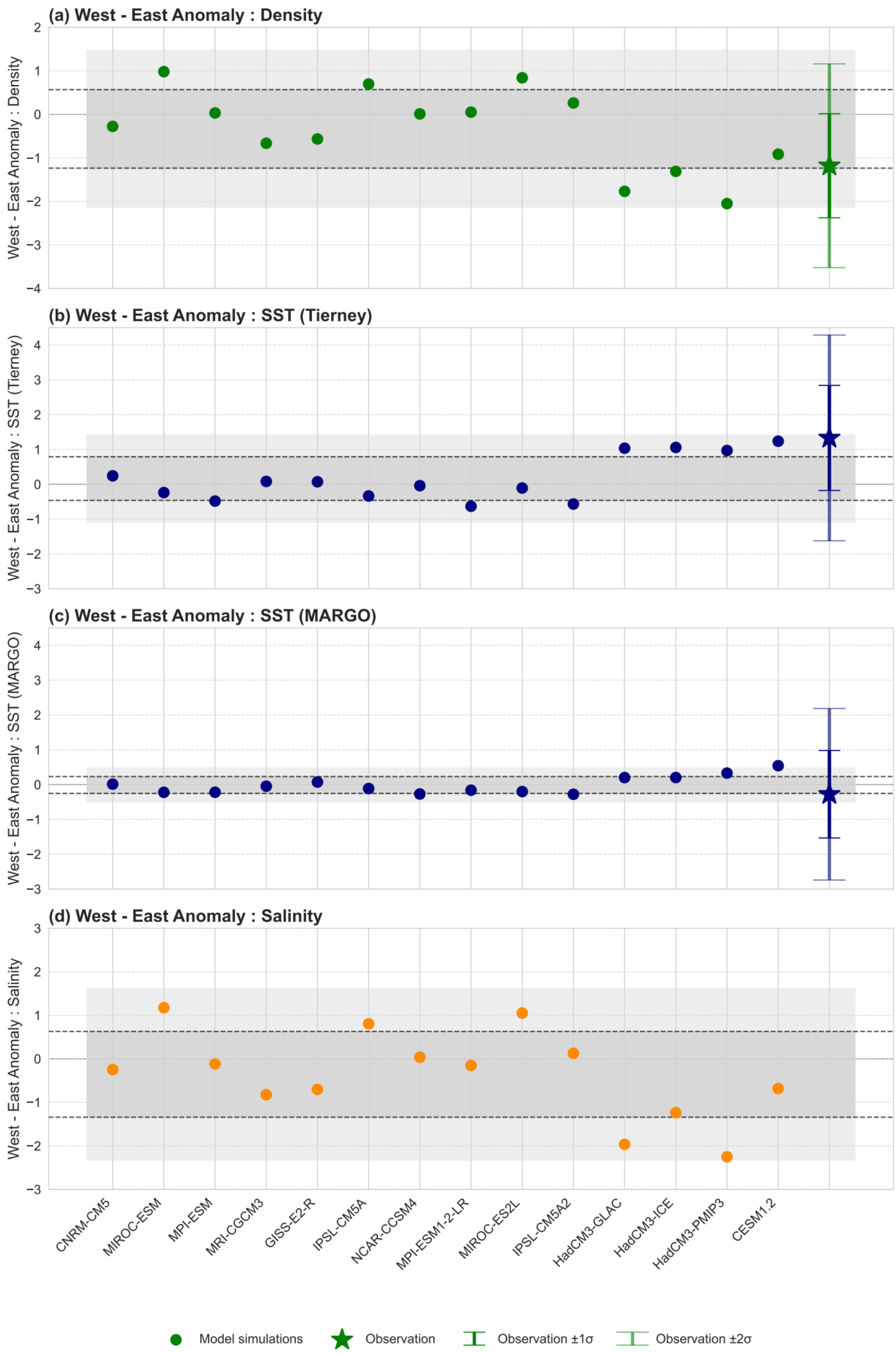
These model–data mismatches are consistent with known CMIP-class biases: many models fail to reproduce the LGM west–east density anomaly due to errors in Walker circulation, monsoon dynamics, and regional precipitation (Feng et al., 2023; McKenna et al., 2024). Such atmospheric circulation biases propagate into the ocean (thermocline slope, zonal SST gradients) and thereby affect both temperature and salinity-driven contributions to surface density. In this light, the spread of model outcomes at the LGM can be partly attributed to differing model responses to LGM boundary conditions and to systematic CMIP biases in tropical atmospheric circulation.

To investigate whether temperature and/or salinity biases could be responsible for the model–data mismatch, we first examine the west–east gradient anomaly in temperature. On the temperature side (Fig. 12 (b) and (c)), the two datasets show differing patterns: the MARGO (MARGO project, 2009) database reports a small West–East gradient (–0.28), whereas the Tierney et al. (2020) dataset shows a positive anomaly (1.33). The SST reconstructions in MARGO (MARGO project, 2009) in this zone are mainly based on foraminiferal assemblages, whereas the reconstructions in the Tierney et al. (2020) database are mainly based on Mg/Ca proxies. As the Tierney reconstructions are geographically closer to the density records, we use them (Fig. 12 (b)) for interpreting the SST anomaly. The positive West–East gradient in Tierney et al. (2020) does not imply the western Indian Ocean was warmer than the east during the LGM; rather, it indicates a relative reduction in the zonal SST gradient compared to

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

preindustrial conditions. This is consistent with the observed west–east density anomaly, where relatively weaker cooling in the west would contribute to lower density than in the east. SST anomalies exhibit smaller inter-model spread than density anomalies. The simulations that most closely reproduce the observed SST anomaly pattern between LGM and piControl are HadCM3-GLAC-1D, HadCM3-ICE-6G_C, HadCM3-PMIP3, and CESM1.2.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026



743

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Figure 12: Model–data comparison of West–East anomalies (LGM – piControl) in the Indian Ocean. Stars denote reconstruction-based anomalies, while shaded areas indicate model uncertainties (±1σ, light shading ±2σ). All model values are extracted at the same grid cells as the observational sites to ensure a spatially consistent comparison. Regional uncertainties are obtained by combining uncertainties from all sediment cores from each box (West and East). For each core, the uncertainty is represented by a Gaussian standard deviation derived from its reported confidence interval. We construct inverse-variance reliability weights using two components: the within-core variance (from the CI-based σ) and the between-core spatial variance (the variance of core-specific means). The regional mean is the weighted average of the core means, and its uncertainty follows the law of total variance, ensuring that both local reconstruction uncertainty and spatial variability are preserved. To obtain smooth uncertainty bands, we then generate Monte-Carlo draws from this mixture. Uncertainty on the West–East anomaly is then obtained by standard error propagation, assuming independent regional means (square root of the sum of squared standard deviations). **(a)** Surface density anomaly (kg/m^3). Density reconstructions are derived from δ^{18}Oc measurements of planktonic foraminifera and converted into density estimates using the Bayesian calibration method of Caley et al. (2025). Observational uncertainty is shown by the 68% (dark green) and 95% (light green) confidence intervals. **(b)** Sea surface temperature (SST) anomaly (°C) based on Tierney et al. (2020). Observational uncertainties are shown as 68% (dark blue) and 95% (blue). **(c)** Same as (b), but using the MARGO database (MARGO Project, 2009). **(d)** Sea surface salinity (SSS) anomaly (g/kg).

DiNezio et al. (2018) investigated the LGM–piControl SST anomaly in this part of the Indian Ocean using the CESM1 model, revealing a West–East SST gradient in this region. Their study showed that two main mechanisms drive the observed glacial-interglacial climate changes: first, the exposure of the Sahul shelf enhances ocean-atmosphere feedbacks that alter rainfall and temperature gradients across the Indian Ocean; second, Northern Hemisphere cooling weakens monsoonal systems by reducing moisture supply, especially over the Arabian Sea. This sensitivity is also dependent on how the newly exposed continental shelf is represented in the model. Factors such as the prescribed surface roughness, vegetation type, or albedo in the now-exposed Indonesian region can significantly impact the simulated atmospheric circulation (DiNezio and Tierney, 2013; Dinezio et al., 2018). In particular, the ability of the convection scheme to respond differently to land versus ocean surfaces plays a critical role in shaping regional precipitation patterns (Chemel et al., 2014). These model design choices likely contribute to the spread of results across PMIP3 and PMIP4 models and their varying skill in reproducing observed SST and surface density gradients.

Salinity biases were also assessed using LGM–PI anomalies and West–East gradients (Fig. 12 (d)). Without direct LGM salinity reconstructions, this relies on model outputs. Inter-model spread in salinity is comparable to density and larger than SST, emphasizing the role of freshwater fluxes and hydrological processes in shaping density changes. Models with the most negative salinity anomalies (MRI-CGCM3, GISS-E2-R, HadCM3-GLAC-1D, HadCM3-ICE-6G_C, HadCM3-PMIP3 and CESM1.2) best reproduce the observed west–east density gradient. Models failing to produce sufficiently negative salinity anomalies often have biases in monsoon precipitation location and intensity, directly affecting surface salinity. Dinezio and Tierney (2013) showed that accurate West–East salinity gradients require correct precipitation patterns. This underscores the importance of correctly simulating the Indian Ocean hydrological cycle (including major rivers) and atmospheric circulation to reproduce salinity-driven density gradients and associated climate impacts.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

All HadCM3 simulations follow the PMIP4 protocol for implementing LGM boundary conditions. This protocol ensures that the differences between ice-sheet reconstructions are consistently applied—not only in terms of ice-sheet mask and elevation, but also land–sea distribution, bathymetry, and far-field topography. Among them, the HadCM3-ICE-6G_C simulation provides the best agreement with the observed west–east surface density gradient. This improved performance is consistent with previous results (Izumi et al., 2023), who showed that the ICE-6G_C ice-sheet configuration induces distinct atmospheric circulation responses compared to other reconstructions. These include shifts in jet structure and stationary waves, as well as differences in surface albedo forcing and sea-ice expansion. Such large-scale circulation adjustments feedback onto the Indo-Pacific climate, which likely explains the slightly more realistic simulation of the west–east density gradient in HadCM3-ICE-6G_C.

Marine sediment core reconstructions provide robust support for a distinct pattern of SST and density changes in the Indian Ocean during the LGM. Our results confirm a strong cooling in the eastern Indian Ocean contrasted with milder cooling in the western basin, leading to a reduction of the zonal SST gradient compared to preindustrial conditions (DiNezio et al., 2018).

We also demonstrate that this zonal SST changes is associated with zonal salinity and density changes. This pattern supports the interpretation that the LGM Indian Ocean mean state was marked by a weakened zonal SST gradient, primarily due to intensified cooling in the eastern basin and a shallower thermocline. Such mean state changes, as shown by Abram et al. (2020), likely enhanced interannual variability across the basin. This underscores the value of palaeoclimate for improving our understanding Indian Ocean mean state under varying boundary conditions.

Of the 14 model simulations, only 7 (50%) successfully reproduce the observed West–East tropical gradient both in SST and density during the LGM, when accounting for the 68% uncertainty range of the reconstructions. This highlights current climate model limitations, indicating that key processes are insufficiently represented or that boundary condition choices, especially ice-sheet reconstructions, could influence model–data agreement.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

5. Conclusion:

The quantitative density reconstruction method based on δ^{18}Oc, developed by Caley et al. (2025), provides a valuable dataset for evaluating climate model simulations from PMIP3 and PMIP4 in terms of LGM density changes. Instead of comparing ensemble means, our analysis evaluates each model simulation individually, revealing a wide range of behaviours in both the amplitude and spatial structure of simulated density anomalies. Overall, models tend to slightly underestimate the magnitude of density anomalies compared to proxy-based reconstructions, which exhibit a mean anomaly of ~1.5 kg/m³, while model simulations average around 1.0 kg/m³. Despite this, at the global scale, 100% of simulations show statistically significant correlations with reconstructions ($p < 0.05$) for LGM and PI periods, and ~50% reach $R^2 > 0.5$, with ~75% reproducing regression slopes within 0.8–1.2, indicating that large-scale patterns are generally captured.

The comparison between model outputs and paleodata suggests that models reproduce the sign of density changes and broad latitudinal patterns, but notable regional discrepancies remain. Across six ocean basins, only ~33% of model–basin combinations achieve robust agreement simultaneously, highlighting limitations at finer scales. For instance, in the Indian Ocean, while mean density anomalies are reasonably reproduced, only 7 out of 14 models (50%) capture the weakened West–East density gradient. This misrepresentation of the mean state of the IOD could lead to biases in climate variability. Discrepancies are linked to biases in SST and salinity changes, and to insufficient representation of low-latitude precipitation reductions, which are critical for reproducing density anomalies in tropical regions.

These results emphasize that evaluating individual simulations rather than ensemble means is essential, as ensemble averaging can mask inter-model differences and region-specific biases. They also highlight the need to expand LGM density reconstructions, particularly in poorly sampled open-ocean regions, to provide stronger observational constraints on models. In the future, integrating these datasets within statistical observational-constraint frameworks and data assimilation approaches could help identify models that most accurately reproduce past climate states and could ultimately improve confidence in future climate projections.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Appendices:

A. *Study basins:*

We first considered the major oceanic basins: The Atlantic, Indian, and Pacific Oceans, the circumpolar Southern Ocean. To better capture regional processes and analyze these regions more precisely, some of these large basins were subdivided into northern and southern parts, resulting in a total of six ocean basins. The Atlantic and Indian Oceans were split with updated latitudinal boundaries. Specifically, the North Atlantic, following the density reconstruction method of Caley et al. (2025), extends from 0° to 40°N, while the South Atlantic is restricted to 0°N. Similarly, the North Indian Basin extends from 0°S northward, and the South Indian Basin is restricted to 0°S.

The Mediterranean Sea was excluded from the analysis, as this region is particularly difficult for climate models to simulate and the limited number of proxy data points resulted in statistically insignificant regressions.

The Pacific and the Southern Ocean retain their original boundaries. The study basins were defined using the NOAA WOA23 mask file (1°×1° grid: https://www.ncei.noaa.gov/data/oceans/woa/WOA23/MASKS/basinmask_01.msk), with modifications applied to create these six final regions (Fig. A1).

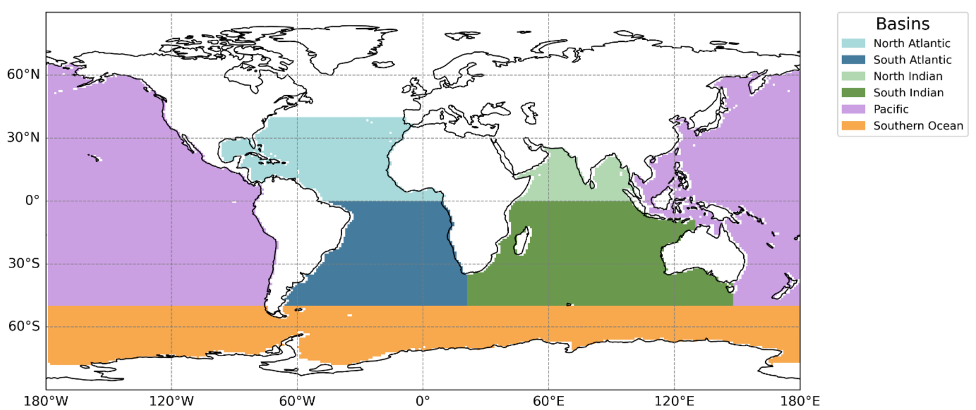


Figure A1: Map of selected basins in this study. The study basins were created from the NOAA WOA23 mask file (which has a 1°x1° grid) (website https://www.ncei.noaa.gov/data/oceans/woa/WOA23/MASKS/basinmask_01.msk) then some basins were modified, in particular to divide them into North and South. 6 basins were defined and studied: North Indian (latitudinal limit at 0° S), South Indian (latitudinal limit at 0° S), North Atlantic (between 0° and 40° N), South Atlantic (latitudinal limit at 0° N), Pacific, and Southern Ocean. North Atlantic is cropped at 40°N following the Caley et al. (2025) density-reconstruction method.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

B. Latitudinal profiles of surface density, SSS and SST anomalies (LGM-PI):

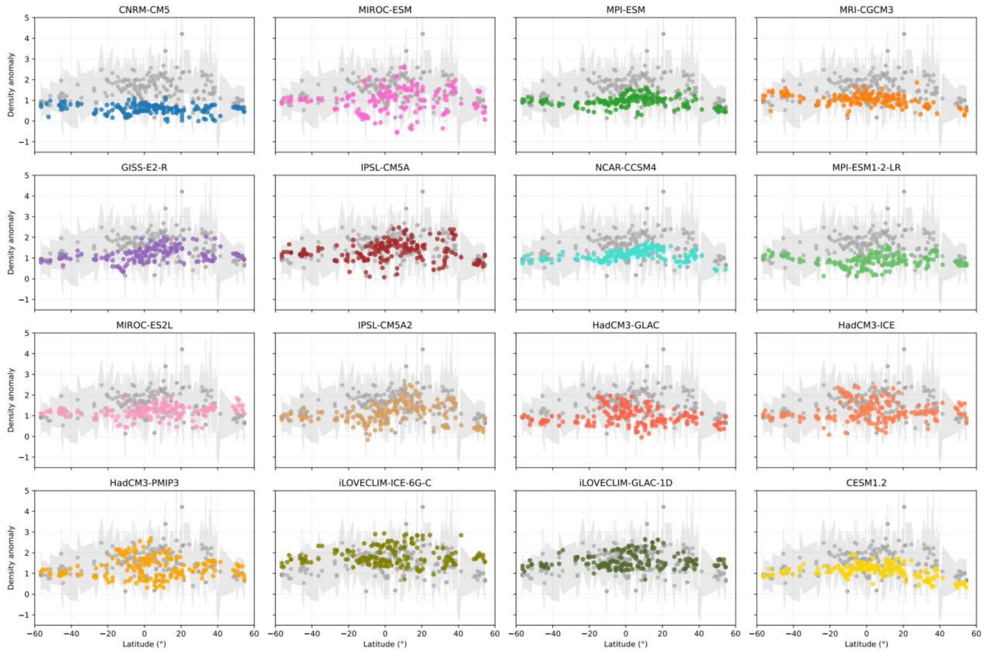


Figure B1: Density anomaly (kg/m³) as a function of latitude for each model simulation (colored dots), compared with observational data (grey dots) and the 68% confidence interval (grey shading). Model outliers, identified using the interquartile range (IQR) method (values outside 1.5×IQR), were excluded to minimize the influence of extreme values and highlight robust spatial patters. This filtering enhances the main latitudinal structure and modelled density anomalies while preserving all available latitude sampling points.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

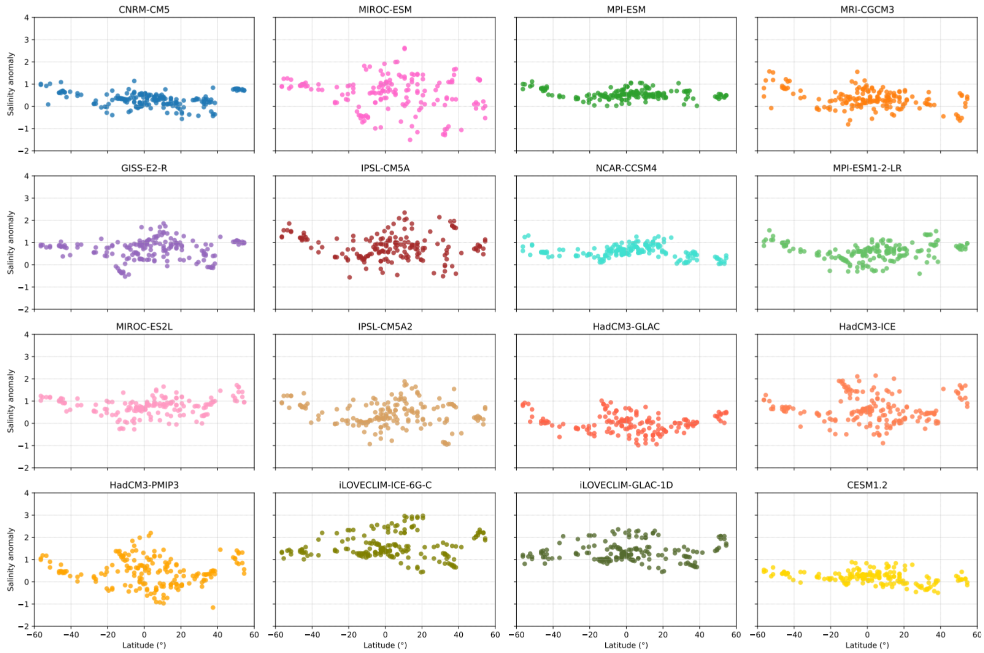


Figure B2: Zonal distribution of salinity anomalies (LGM–PI) simulated by each model (colored dots). Model outliers, identified using the interquartile range (IQR) method (values outside 1.5×IQR), were excluded to minimize the influence of extreme values and highlight consistent spatial patterns. This filtering emphasizes the latitudinal structure of simulated salinity anomalies across models while preserving all available sampling points.

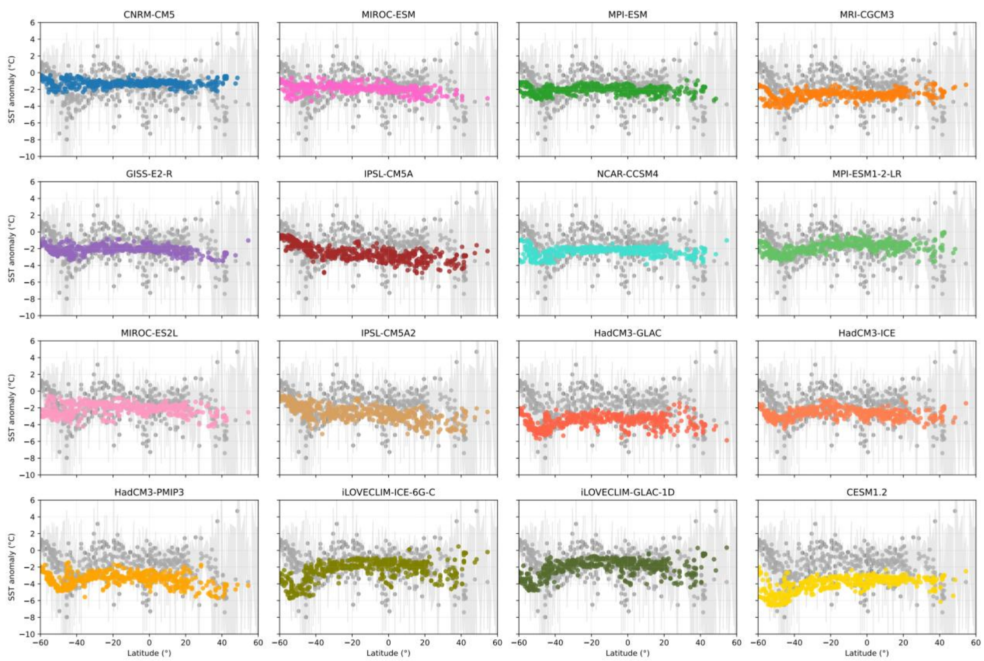
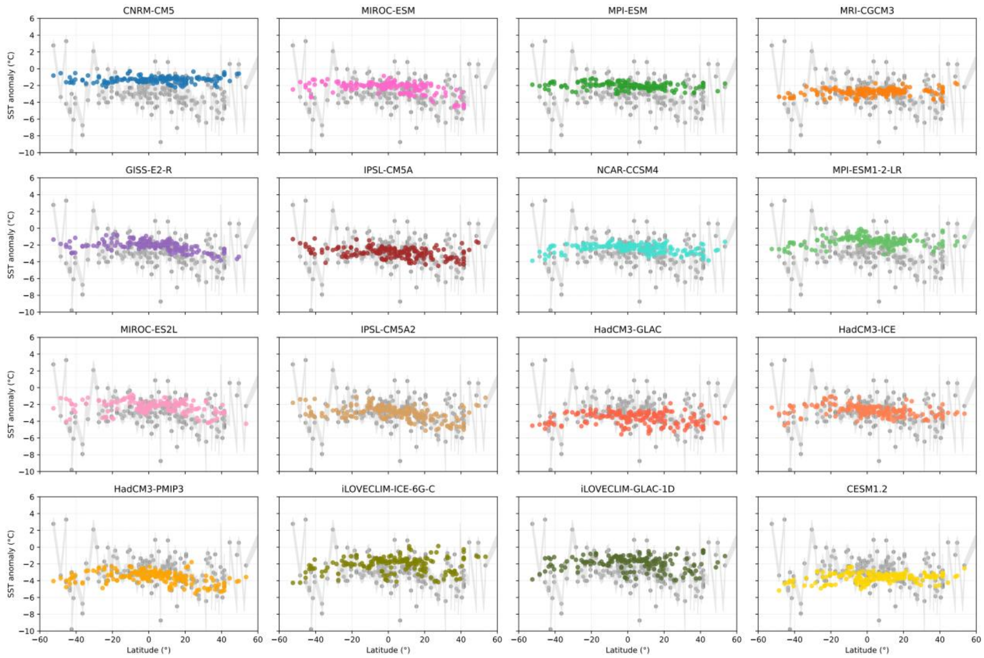


Figure B3: Zonal distribution of SST anomalies (LGM–PI) from the MARGO (MARGO project, 2009) reconstruction (grey dots) and ± 1σ uncertainties (grey shading), compared with model simulations (colored dots). Model outliers, identified using the

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

interquartile range (IQR) method (values outside 1.5×IQR), were excluded to minimize the impact of extreme values and highlight robust latitudinal patterns. This filtering enhances the readability of the large-scale SST anomaly structure while preserving all available latitude points.



<u>Figure B4:</u> Zonal distribution of SST anomalies (LGM–PI) from Tierney et al. (2020) reconstruction (grey dots) and ± 1σ uncertainties (grey shading), compared with model simulations (colored dots). Model outliers, identified using the interquartile range (IQR) method (values outside 1.5×IQR), were excluded to minimize the impact of extreme values and highlight consistent spatial patterns. This filtering clarifies the latitudinal structure of simulated SST anomalies while preserving all available observation–model comparison points.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

C. Comparison of global distribution of surface density anomalies:

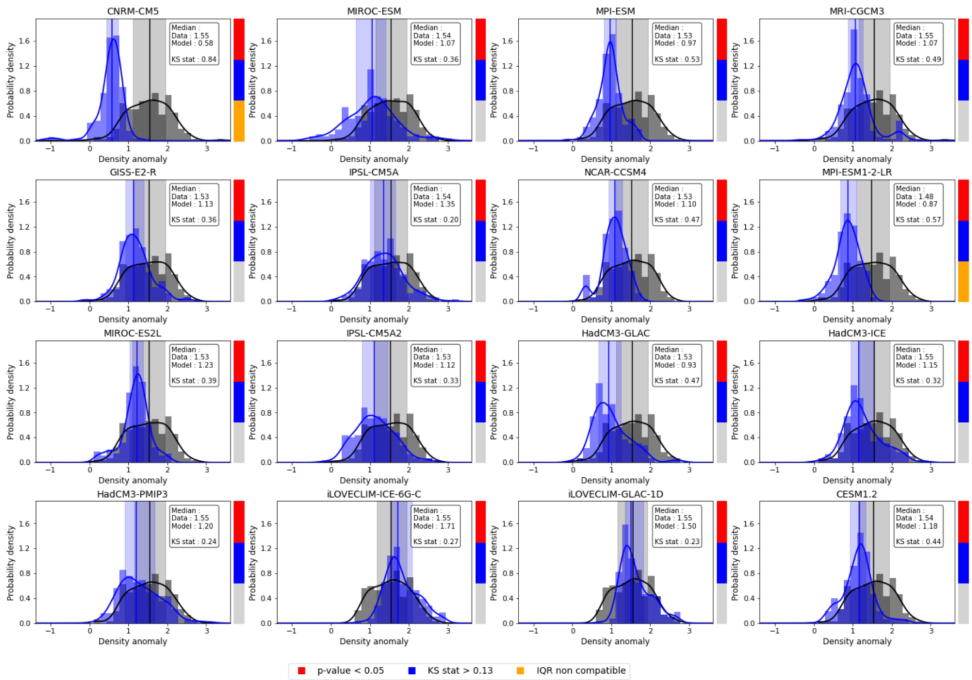


Figure C1: Distribution histograms of surface density anomalies (LGM-PI, kg/m3) at the global scale. Density reconstructions are shown in black, model simulations in blue. Kernel Density Estimates (KDEs) illustrate the central tendency and overall shape of the distributions. Vertical lines indicate the median of each distribution, and shaded envelopes represent the interquartile ranges (IQRs), providing a measure of data spread that is independent of extreme values. The histograms display the frequency of values, complementing the KDEs and IQRs to give an integrated view of distribution characteristics. Note that uncertainties associated with reconstructions and model heterogeneity are not shown, as the figure focuses on distributional comparison. Colored indicators on the right side of each panel highlight whether the comparison fails to meet any of the following criteria: p-value < 0.05, KS statistic ≥ 0.13, or non-overlapping IQRs between data and model distributions. A colored flag indicates that the corresponding criterion is not satisfied.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

D. Global Evaluation of Surface Density: Models vs. Reconstructions (PI & LGM):

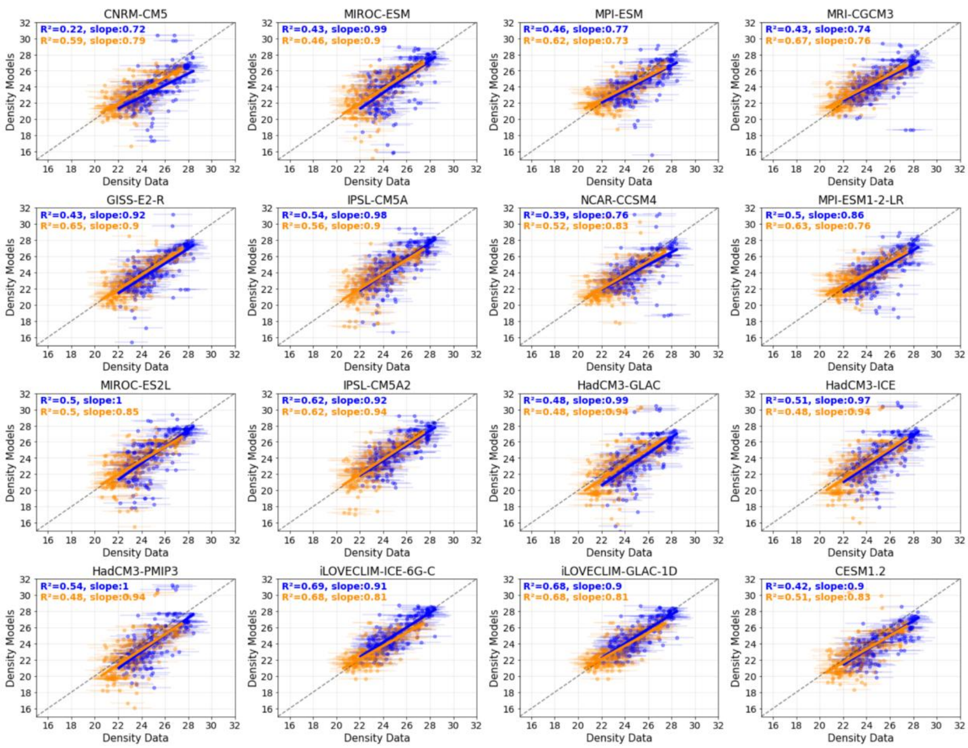


Figure D1: Linear regressions between absolute surface density (kg/m³) from proxy-based reconstructions (x-axis) and model simulations (y-axis), aggregated at the global scale (across all selected basins). Results are shown for the LGM period (blue) and the piControl period (orange). Error bars on the x-axis represent the 95% confidence intervals of the reconstructed values. The slope and R^2 values correspond to standard linear regressions, without accounting for uncertainties on the x-axis (the Monte Carlo method was not applied here). All regressions shown are statistically significant ($p < 0.05$).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

E. Focus on the Indian Ocean:

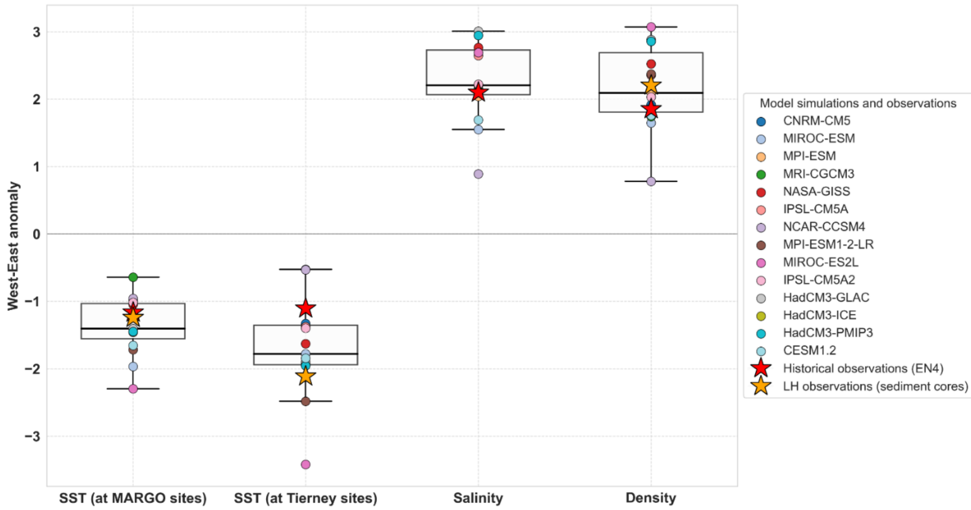


Figure E1: Boxplots showing West–East anomalies (Indian_East – Indian_West) in Sea Surface Temperature (°C), (from the MARGO and Tierney datasets, separately), Sea Surface Salinity (g/kg), and sea surface density (kg/m³) during the pre-industrial period. The boxplots represent the distribution of anomalies from climate model simulations only. Each colored dot corresponds to the anomaly from an individual model simulation. Red stars indicate observed values from the EN4 dataset (Good et al., 2013), averaged over the period 1900–1999 to provide a historical reference. Orange stars indicate reconstructions from databases (PI from WOA18 for MARGO (MARGO project, 2009), LH for Tierney et al. (2020) and LH for density (Caley et al.2025)).

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Data availability:

Most of the CMIP5/PMIP3 and CMIP6/PMIP4 climate model simulations used in this study are publicly available through the Earth System Grid Federation (ESGF). The simulations IPSL-CM5A2, HadCM3-GLAC-1D, HadCM3-ICE-6G_C, HadCM3-PMIP3, iLOVECLIM-GLAC-1D, iLOVECLIM-ICE-6G_C, and CESM1.2 are not available via ESGF. CESM1.2 simulation outputs are openly available from Zenodo (https://zenodo.org/records/14957995). The HadCM3 model simulations can be accessed at https://www.paleo.bristol.ac.uk/ummodel/scripts/papers/. iLOVECLIM simulations outputs are available upon request from Nathalie Bouttes, and IPSL-CM5A2 data can be obtained by contacting Masa Kageyama. References for all simulations are provided in Table 1.

The δ^{18}Oc database and the Python code to compute surface ocean density are related to Caley et al. 2025 and can be found at https://github.com/nicrie/density_uncertainty/tree/main/data

Author contribution:

TC and HB designed the study. HB and TC designed the analyses, and HB performed them. MK, PB and DS assisted in retrieving model simulations data. HB and TC analysed the results with contribution and discussion of all co-authors. HB produced the figures and wrote the article with help from TC and input from all co-authors.

Competing interests:

The authors declare that they have no competing interest.

Acknowledgements:

We thank Valentin Portmann for his support and expertise in statistics and Python-based data processing, and for providing the regridded EN4 dataset used in this study. We also thank Didier Roche for carefully reviewing the manuscript and providing constructive feedback. We are grateful to the paleoclimate modelling community for providing access to model outputs used in this study, and we warmly thank Masa Kageyama, Jiang Zhu, Nathaelle Bouttes, and Ruza Ivanovic for their help in retrieving PMIP simulations.

The authors would also like to thank all the modelling groups who provided the PMIP3 and PMIP4 outputs used in this analysis, as well as the CMIP panel and ESGF infrastructures for making the data available, and WCRP and CLIVAR for supporting the PMIP project. Héloïse Barathieu acknowledges the use of the IPSL (ESPRI – Ensemble de Services Pour la Recherche l'IPSL – computing and data centre (https://mesocentre.ipsl.fr/, last access: July 2025).

Financial support:

This research was supported by the ANR HYDRATE project (grant no. ANR-21-CE01-0001) of the French Agence Nationale de la Recherche.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

References:

Abram, N. J., Hargreaves, J. A., Wright, N. M., Thirumalai, K., Ummenhofer, C. C., and England, M. H.: Palaeoclimate perspectives on the Indian Ocean dipole, Quat. Sci. Rev., 237, 106302, https://doi.org/10.1016/j.quascirev.2020.106302, 2020.

Adloff, M., Reick, C. H., and Claussen, M.: Earth system model simulations show different feedback strengths of the terrestrial carbon cycle under glacial and interglacial conditions, Earth Syst. Dynam., 9, 413–425, https://doi.org/10.5194/esd-9-413-2018, 2018.

Annan, J. D., Hargreaves, J. C., and Mauritsen, T.: A new global surface temperature reconstruction for the Last Glacial Maximum, *Clim. Past*, 18, 1883–1898, https://doi.org/10.5194/cp-18-1883-2022, 2022.

Ayache, M., Swingedouw, D., Mary, Y., Eynaud, F., and Colin, C.: Multi-centennial variability of the AMOC over the Holocene: A new reconstruction based on multiple proxy-derived SST records, Glob. Planet. Change, 170, 172–189, https://doi.org/10.1016/j.gloplacha.2018.08.016, 2018.

Bouttes, N., Lhardy, F., Quiquet, A., Paillard, D., Goosse, H., and Roche, D. M.: Deglacial climate changes as forced by different ice sheet reconstructions, Clim. Past, 19, 1027–1042, https://doi.org/10.5194/cp-19-1027-2023, 2023.

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

Brady, E. C., Otto-Bliesner, B. L., Kay, J. E., and Rosenbloom, N.: Sensitivity to glacial forcing in the CCSM4, J. Clim., 26, 1901–1925, https://doi.org/10.1175/JCLI-D-11-00416.1, 2013.

Brierley, C., Thirumalai, K., Grindrod, E., and Barnsley, J.: Indian Ocean variability changes in the Paleoclimate Modelling Intercomparison Project, Clim. Past, 19, 681–701, https://doi.org/10.5194/cp-19-681-2023, 2023.

Cai, W., and Cowan, T.: Why is the amplitude of the Indian Ocean Dipole overly large in CMIP3 and CMIP5 climate models?, Geophys. Res. Lett., 40, 1200–1205, https://doi.org/10.1002/grl.50208, 2013.

Caley, T., Rieger, N., Werner, M., Barathieu, H., Happé, T., Roche, D. M., ...: Past Ocean surface density from planktonic foraminifera calcite δ18O, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-2459, 10 June 2025.

Chemel, C., Russo, M. R., Hosking, J. S., Telford, P. J., and Pyle, J. A.: Sensitivity of tropical deep convection in global models: effects of horizontal resolution, surface constraints, and 3D atmospheric nudging, Atmos. Sci. Lett., 16, 148–154, https://doi.org/10.1002/asl2.540, 2014.

DiNezio, P. N., and Tierney, J. E.: The effect of sea level on glacial Indo-Pacific climate, Nat. Geosci., 6, 485–491, https://doi.org/10.1038/ngeo1823, 2013.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

DiNezio, P. N., Tierney, J. E., Otto-Bliesner, B. L., Timmermann, A., Bhattacharya, T., Rosenbloom, N., and Brady, E.: Glacial changes in tropical climate amplified by the Indian Ocean, Sci. Adv., 4, eaat9658, https://doi.org/10.1126/sciadv.aat9658, 2018.

Douville, H., Raghavan, K., Renwick, J., Allan, R. P., Arias, P. A., Barlow, M., Cerezo-Mota, R., Cherchi, A., Gan, T. Y., Gergis, J., Jiang, D., Khan, A., Pokam Mba, W., Rosenfeld, D., Tierney, J., and Zolina, O.: Water Cycle Changes, in: Climate Change 2021 – The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1055–1210, https://doi.org/10.1017/9781009157896.010, 2023.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Clim. Dyn., 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 423–552, https://doi.org/10.1017/9781009157896.005, 2021.

Feng, J., Lian, T., and Chen, D.: Tropical Indian Ocean mixed layer bias in CMIP6 CGCMs primarily attributed to the AGCM surface wind bias, J. Climate, 36, 4169–4188, https://doi.org/10.1175/JCLI-D-22-0546.1, 2023.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of climate models, in: Climate Change 2013: The Physical

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 741–866, https://doi.org/10.1017/CBO9781107415324.020, 2014.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P., Yang, Z., and Zhang, M.: The community climate system model version 4, J. Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans, 118, 6704–6716, https://doi-org.docelec.u-bordeaux.fr/10.1002/2013JC009067, 2013.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Kato, M., Kawamiya, M., and Takemura, T.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geosci. Model Dev., 13, 2197–2244, https://doi.org/10.5194/gmd-13-2197-2020, 2020.

Harrison, S. P., Bartlein, P. J., Brewer, S., Prentice, I. C., Boyd, M., Hessler, I., Otto-Bliesner, B., Brady, E., Foley, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Clim. Dyn., 43, 671–688, https://doi.org/10.1007/s00382-013-1922-6, 2014.

Izumi, K., Valdes, P., Ivanovic, R., and Gregoire, L.: Impacts of the PMIP4 ice sheets on Northern Hemisphere climate during the last glacial period, Clim. Dyn., 60, 2481–2499, https://doi.org/10.1007/s00382-022-06456-1, 2023.

Joussaume, S., and Taylor, K. E.: Status of the Paleoclimate Modeling Intercomparison Project (PMIP), in: Proceedings of the first international AMIP scientific conference, WRCP Report, 425–430, 1995.

Kageyama, M., Albani, S., Braconnot, P., Harrison, S. P., Hopcroft, P. O., Ivanovic, R. F., Joussaume, S., Lora, J. M., Mikolajewicz, U., Roche, D. M., Volodin, E., and Zheng, W.: The PMIP4 contribution to CMIP6–Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments, Geosci. Model Dev., 10, 4035–4055, https://doi.org/10.5194/gmd-10-4035-2017, 2017.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Rachmayani, R., Zheng, W., and Zhou, T.: The PMIP4 contribution to CMIP6–Part 1: Overview and over-arching analysis plan, Geosci. Model Dev., 11, 1033–1057, https://doi.org/10.5194/gmd-11-1033-2018, 2018.

Kageyama, M., Harrison, S. P., Kapsch, M.-L., Lofverstrom, M., Lora, J. M., Mikolajewicz, U., Sherriff-Tadano, S., Vadsaria, T., Abe-Ouchi, A., Bouttes, N., Chandan, D., Gregoire, L. J., Ivanovic, R. F., Izumi, K., LeGrande, A. N., Lhardy, F., Lohmann, G., Morozova, P. A., Ohgaito, R., Paul, A., Peltier, W. R., Poulsen, C. J., Quiquet, A., Roche, D. M., Shi, X., Tierney, J. E., Valdes,

P. J., Volodin, E., and Zhu, J.: The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations, Clim. Past, 17, 1065–1089, https://doi.org/10.5194/cp-17-1065-2021, 2021

Kageyama, M., Braconnot, P., Chiessi, C. M., Rehfeld, K., Ait Brahim, Y., Dütsch, M., Gwinneth, B., Hou, A., Loutre, M.-F., Hendrizan, M., Meissner, K., Mongwe, P., Otto-Bliesner, B., Pezzi, L. P., Rovere, A., Seltzer, A., Sime, L., and Zhu, J.: Lessons from paleoclimates for recent and future climate change: opportunities and insights, *Front. Clim.*, 6, 1511997, https://doi.org/10.3389/fclim.2024.1511997, 2024

Lhardy, F., Bouttes, N., Roche, D. M., Crosta, X., Waelbroeck, C., and Paillard, D.: Impact of Southern Ocean surface conditions on deep ocean circulation during the LGM: a model analysis, Clim. Past, 17, 1139–1159, https://doi.org/10.5194/cp-17-1139-2021, 2021.

MARGO Project Members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, Nat. Geosci., 2, 127–132, https://doi.org/10.1038/NGEO411, 2009.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Crueger, T., Esch, M., Fast, I., Fiedler, S., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Kornblueh, L., Lohmann, K., Mäkelä, J., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., Storch, J. S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2, J. Adv. Model. Earth Syst., 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

McKenna, S., Santoso, A., Sen Gupta, A., and Taschetto, A. S.: Understanding biases in Indian Ocean seasonal SST in CMIP6 models, *J. Geophys. Res. Oceans*, 129, e2023JC020330, https://doi.org/10.1029/2023JC020330, 2024.

Mix, A. C., Bard, E., and Schneider, R.: Environmental processes of the ice age: land, oceans, glaciers (EPILOG), *Quat. Sci. Rev.*, 20, 627–657, https://doi.org/10.1016/S0277-3791(00)00145-1, 2001.

NOAA: WOA23 basin mask file (1°×1° grid). National Centers for Environmental Information, NOAA, USA. https://www.ncei.noaa.gov/data/oceans/woa/WOA23/MASKS/basinmask_01.msk, accessed 2025.

NODC: World Ocean Atlas 1998 (WOA98). National Oceanographic Data Center, Silver Spring, MD, USA. https://psl.noaa.gov/data/gridded/data.nodc.woa98.html, 1998.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Roche, D. M., and Caley, T.: δ[18]O water isotope in the iLOVECLIM model (version 1.0) – Part 2: Evaluation of model results against observed δ[18]O in water samples, *Geosci. Model Dev.*, 6, 1493–1504, https://doi.org/10.5194/gmd-6-1493-2013, 2013.

Roquet, F., Madec, G., McDougall, T. J., and Barker, P. M.: Accurate polynomial expressions for the density and specific volume of seawater using the TEOS-10 standard, *Ocean Model.*, 90, 29–43, https://doi.org/10.1016/j.ocemod.2015.04.002, 2015.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T.: A dipole mode in the tropical Indian Ocean, *Nature*, 401, 360–363, https://doi.org/10.1038/43854, 1999.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Del Genio, A., Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, 6, 141–184, https://doi.org/10.1002/2013MS000265, 2014.

Sepulchre, P., Caubel, A., Ladant, J.-B., Bopp, L., Boucher, O., Braconnot, P., Brockmann, P., Cozic, A., Dufresne, J.-L., Estella-Perez, V., Foujols, M.-A., Ghattas, J., Hauglustaine, D., Hourdin, F., Kageyama, M., Khodri, M., Krinner, G., Levavasseur, G., Marchand, M., Ottlé, C., Peylin, P., Servonnat, J., Swingedouw, D., Szopa, S., Tardif, D., Terray, P., Vuichard, N., and Guilyardi, E.: IPSL-CM5A2 – an Earth system model designed for multi-millennial climate simulations, *Geosci. Model Dev.*, 13, 3011–3053, https://doi.org/10.5194/gmd-13-3011-2020, 2020.

Sueyoshi, T., Ohgaito, R., Yamamoto, A., Chikamoto, M. O., Hajima, T., Okajima, H., Yoshimori, M., Abe, M., O'Ishi, R., Watanabe, S., and Abe-Ouchi, A.: Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM, *Geosci. Model Dev.*, 6, 819–836, https://doi.org/10.5194/gmd-6-819-2013, 2013.

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J.: Glacial cooling and climate sensitivity revisited, Nature, 584, 569–573, https://doi.org/10.1038/s41586-020-2617-x, 2020.

Ullman, D. J., LeGrande, A. N., Carlson, A. E., Anslow, F. S., and Licciardi, J. M.: Assessing the impact of Laurentide Ice Sheet topography on glacial climate, Clim. Past, 10, 487–507, https://doi.org/10.5194/cp-10-487-2014, 2014.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, *Clim. Dyn.*, 40, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y, 2013.

https://doi.org/10.5194/egusphere-2026-254
Preprint. Discussion started: 2 February 2026

Weller, E., and Cai, W.: Realism of the Indian Ocean Dipole in CMIP5 models: The implications for climate projections, *J. Clim.*, 26, 6649–6659, https://doi.org/10.1175/JCLI-D-12-00807.1, 2013.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., Mizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T., and Kitoh, A.: A new global climate model of the Meteorological Research Institute: MRI-CGCM3 – Model description and basic performance, *J. Meteorol. Soc. Jpn.*, 90, 23–64, https://doi.org/10.2151/jmsj.2012-A02, 2012.

Zhang, Y., Du, Y., and Qu, T.: A sea surface salinity dipole mode in the tropical Indian Ocean, Clim. Dyn., 47, 2573–2585, https://doi.org/10.1007/s00382-016-2984-z, 2016.