ARTICLE

https://doi.org/10.1038/s41586-019-0901-4

Revisiting Antarctic ice loss due to marine ice-cliff instability

Tamsin L. Edwards[1]*, Mark A. Brandon[2], Gael Durand[3], Neil R. Edwards[2], Nicholas R. Golledge[4,5], Philip B. Holden[2], Isabel J. Nias[6], Antony J. Payne[7], Catherine Ritz[3] & Andreas Wernecke[2]

Predictions for sea-level rise this century due to melt from Antarctica range from zero to more than one metre. The highest predictions are driven by the controversial marine ice-cliff instability (MICI) hypothesis, which assumes that coastal ice cliffs can rapidly collapse after ice shelves disintegrate, as a result of surface and sub-shelf melting caused by global warming. But MICI has not been observed in the modern era and it remains unclear whether it is required to reproduce sea-level variations in the geological past. Here we quantify ice-sheet modelling uncertainties for the original MICI study and show that the probability distributions are skewed towards lower values (under very high greenhouse gas concentrations, the most likely value is 45 centimetres). However, MICI is not required to reproduce sea-level changes due to Antarctic ice loss in the mid-Pliocene epoch, the last interglacial period or 1992–2017; without it we find that the projections agree with previous studies (all 95th percentiles are less than 43 centimetres). We conclude that previous interpretations of these MICI projections over-estimate sea-level rise this century; because the MICI hypothesis is not well constrained, confidence in projections with MICI would require a greater range of observationally constrained models of ice-shelf vulnerability and ice-cliff collapse.

Projections of the Antarctic contribution to global-mean sea-level rise this century from process-based models vary widely[1–6]. In particular, DeConto and Pollard[6] (DP16) hypothesized a MICI process[7] that results in mean values exceeding 1 m by 2100 under some methodological choices. However, the DP16 results are sensitive to these methodological choices (Table 1; Extended Data Fig. 1a, b) and the shapes of the probability distributions are very poorly known (Extended Data Fig. 2), leading to extremely wide probability intervals (Table 1). This considerable uncertainty poses challenges for robust and cost-effective management of coastal flood risk.

The Antarctic contribution to global-mean sea level has two parts: increasing snowfall, which is expected to reduce global-mean sea level by a few centimetres this century, and ice discharge into the ocean, which is very uncertain[1]. The latter is determined by outflow of ice across the grounding line (the boundary between floating and grounded ice), which can increase as a result of faster ice flow or inland retreat of the grounding line. Ice discharge can increase if buttressing by ice shelves is reduced by (1) ice-shelf thinning, caused by enhanced oceanic melting due to circulation changes[8] or direct warming, or (2) partial or total ice-shelf collapse, caused by widening of surface crevasses by meltwater due to atmospheric warming[9,10].

Marine parts of the ice sheet, which lie on bedrock below sea level, are potentially vulnerable to two hypothesized positive feedback effects that may have led to past collapse of the West Antarctic Ice Sheet[11]. Both are based on physical mechanisms with theoretical foundations, but it is not yet clear whether they are indeed positive feedback effects leading to widespread, rapid and sustained ice losses. The first, marine ice-sheet instability (MISI)[12], is a self-sustaining retreat of the grounding line in regions where the bedrock slopes downward inland, and is triggered by ice-shelf thinning or collapse. Ice thickness at the grounding line increases (owing to the bedrock slope), which leads to faster ice flow and thus further retreat. Satellite and modelling evidence suggests that MISI is underway in West Antarctica[13–15], driven by warm circumpolar deep water breaching the continental shelf—although it is unclear to what degree this has been affected by human activities[1,16,17]. The second instability, MICI[6,7], is a self-sustaining retreat of the ice front in regions where the ice is 100 m or more above the ocean surface[18], and is triggered by ice-shelf collapse. These tall ice cliffs are structurally unstable, and their collapse could leave behind further tall cliffs, resulting in sustained ice losses. Observational evidence for MICI is indirect: an absence of ice cliffs taller than 100 m and rapid retreat of the front of the Jakobshavn (Greenland) and Crane (Antarctic) glaciers (see 'Knowledge gaps and future directions').

DP16 uses a model of the Antarctic Ice Sheet with a parameterization of MICI[7], generating a 64-member ensemble by varying three parameters, which control the relationship between ocean temperature and basal melting, ice-shelf disintegration and the maximum rate of ice-cliff collapse. Projections to 2500 are made under three Representative Concentration Pathways (RCPs): RCP2.6, RCP4.5 and RCP8.5, for very low, low-to-medium and very high greenhouse gas concentrations, respectively[1]. The projections are calibrated by accepting only ensemble members that reproduce reconstructed Antarctic sea-level contributions in the mid-Pliocene (about 3 million years ago) and last interglacial (roughly 130,000 to 115,000 years ago), and present results for two methodological choices. The first is for the Pliocene calibration, using a sea-level interval of 5–15 m ('low Pliocene') or 10–20 m ('high Pliocene'). The latter increases sea-level contributions by up to 40 cm by 2100 and 2.5 m by 2500 under RCP8.5. The second methodological choice is whether to use an ocean temperature correction of +3 °C in West Antarctica to improve simulations of the present-day ice sheet. The 'bias-corrected' option has up to 15 cm greater sea-level contribution this century than the 'bias-uncorrected' one, but it makes little difference by 2500. Results for RCP8.5 at 2100 are given in Table 1; the corresponding distributions are shown in Extended Data Fig. 2.

[1]Department of Geography, King's College London, London, UK. [2]School of Environment, Earth and Ecosystem Sciences, Faculty of Science, Technology, Engineering and Mathematics, Open University, Milton Keynes, UK. [3]Université Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France. [4]Antarctic Research Centre, Victoria University of Wellington, Wellington, New Zealand. [5]GNS Science, Avalon, Lower Hutt, New Zealand. [6]Earth System Science Interdisciplinary Center, College Park, MD, USA. [7]Centre for Polar Observation and Modelling, School of Geographical Sciences, University of Bristol, Bristol, UK. *e-mail: tamsin.edwards@kcl.ac.uk

Table 1 | Antarctic predictions at 2100 under RCP8.5 from DP16

	Low Pliocene		High Pliocene	
	Bias-uncorrected	Bias-corrected	Bias-uncorrected	Bias-corrected
Mean ± 1 s.d.	64 ± 49	79 ± 46	105 ± 30	114 ± 36
≥68% probability interval	[−22, 150]	[−2, 160]	[51, 158]	[51, 177]
≥90% probability interval	[−90, 217]	[−65, 223]	[9, 200]	[1, 227]

Means, standard deviations and implied probability intervals are given in units of centimetres sea-level equivalent for the DP16 ensemble at 2100 under RCP8.5, for their four methodological choices (see text). We derive the probability intervals using minimal assumptions about the shape of the distribution, assuming only finite mean and variance.

We use statistical techniques of uncertainty quantification for computationally expensive computer models to re-examine and estimate probability distributions for the DP16 projections. We calibrate with the Pliocene, last interglacial and satellite (1992–2017) eras and make probabilistic projections with and without MICI, and compare with other probabilistic model projections and with a Gaussian interpretation of DP16. Finally, we outline knowledge gaps and suggest future directions.

Projections for Antarctica

We estimate probabilistic projections for the Antarctic contribution to sea-level rise by emulating the DP16 ice-sheet model (Methods). Emulation is a technique used to quantify how the outputs of a computer model vary as a function of its input parameters, with the purpose of predicting outputs for any choice of parameter values, enabling us to generate a far larger ensemble than with the original model and to present results with and without MICI. We assume that all parameter values are equally likely within the original ranges, on the basis of discussions with the authors of DP16 (R. DeConto, personal communication). Estimating probability distributions enables meaningful comparisons with other studies and decision-making using sea-level exceedance probabilities under MICI and no-MICI scenarios. Our method has two further additions: calibration with both palaeodata and satellite data and accounting for error in the ice-sheet model.

Reconstructions of past climate change provide important tests of models, particularly when the changes were large and/or warmer than today, but their uncertainties are typically large and often poorly defined[19]. Recent observations have smaller signals but much smaller uncertainties. The two provide complementary information, so we use both. We use the low Pliocene interval (which is equivalent to a combined range of 5–20 m because the highest simulation is 12.4 m) for two reasons: first, because of the large reconstruction uncertainty (values lower than 10 m cannot be ruled out; for example, a more recent estimate[20] has a maximum of 13 m); and second, because the DP16 projections are very sensitive to the lower bound of the high Pliocene (Extended Data Fig. 1a, b). The relationships between the RCP8.5 sea-level contribution at 2100 and sea-level change for the three past eras are shown in Extended Data Fig. 3.

To estimate probability distributions we use history matching, in which implausible model versions are excluded, rather than the more commonly used Bayesian calibration, in which model versions are weighted by their agreement with observations using a likelihood function (a metric of model success), for several reasons. First, the concept of history matching is the same as in DP16, which allows us to make a simpler and more transparent comparison. This method effectively estimates what would have been found in DP16 with substantially greater computing resources, calibration with satellite data and accounting for model error. Second, history matching is more cautious than Bayesian model calibration: if no model versions match the data, all are excluded. Finally, we do not know the shape of the crucial Bayesian likelihood function for the Pliocene and last interglacial; this would require estimates of the mean and error distribution of the palaeodata, rather than assuming that all values within the interval are equally likely. Guessing these properties might shift (wrong mean) or narrow (wrong distribution) the final probability distributions.

Accounting for model error (often known as model discrepancy[21]) widens the calibration intervals of acceptance (Extended Data Figs. 3, 4, from grey shaded boxes to dashed lines) and is necessary to avoid over-confidence[22,23]: the aim is to account for the structural error of the model and other uncertainties that are not sampled in the ensemble. These discrepancy terms are tolerances that reflect how well we expect the ice-sheet model to reproduce reality. We specify them using expert judgement, including the judgement that model errors are greater than reconstruction and observation errors[4,24] (that is, we judge that our confidence in simulating reality with the ice-sheet model is lower than that in observing or reconstructing it from measurements). Reconstruction errors are not defined in DP16, so we conservatively use half the palaeodata range (Pliocene, 5 m; last interglacial, 2 m) to avoid under-estimating the uncertainty. For the satellite period, the sea-level change is 0.756 ± 0.386 cm for 1992–2017[25]; we conservatively specify the model error as 0.5 cm.

We present projections at 2100 in Fig. 1 and Table 2. The distributions are skewed—the modes are consistently lower than the medians and means. The results are not strongly dependent on the lower bound of the Pliocene calibration, unlike the DP16 ensemble, owing to the much larger ensemble size (Extended Data Fig. 1c, RCP8.5 at 2100 with MICI). Emulated projections without MICI are much lower than those with MICI, and are consistent with previous projections[4] (Fig. 1b; see 'Multi-model comparisons'). The results are robust to changes in calibration era and discrepancy (Extended Data Fig. 5).

Crucially, our results show that ice-cliff instability is not required to reproduce sea-level changes in these three very different eras: 55% of the MICI and 51% of the no-MICI emulator ensemble members simultaneously pass calibration with the Pliocene, last interglacial and satellite eras (Extended Data Fig. 4, larger emulator blue circles within the dashed box). MICI increases the ensemble range to encompass more of the data intervals, but the emulator can identify many more areas of the parameter space of the model that are successful—including many without MICI. MICI is therefore not necessary for realistic simulations of these periods, so this positive feedback hypothesis cannot be supported or ruled out with this data and calibration method. In fact, the Pliocene does not rule out any ensemble members, because accounting for model error widens the calibration interval to accept them all (Extended Data Fig. 3a).

The original DP16 projections have substantial probabilities of net sea-level fall this century, with the RCP8.5 low-Pliocene mean ± 1 s.d. envelope including negative values until the 2070s. The emulated projections reflect this (Fig. 2), although with lower probability (the 5th percentile is negative until the 2070s). Calibration selects mostly positive sea-level contributions during the satellite era (Extended Data Fig. 3c), then surface accumulation increases with warming (particularly for RCP8.5) and dominates over ice discharge in many ensemble members during this period.

We estimate when the hypothesized MICI feedback would accelerate sea-level rise. Projections with MICI quickly start to diverge from those without MICI for all RCPs in the 2020s (95th percentiles; Fig. 2), as a result of contributions from the Antarctic Peninsula (Fig. 4c of ref. [6]). The MICI projections depend on the RCP by mid-century, whereas the emergence of a clear, RCP-dependent signal without MICI begins in the 2060s–2070s.

We apply the same emulation and calibration methods to the full DP16 time series to 2500 (Fig. 3a). The RCP8.5 distribution remains very skewed, with the mode at the low end of the range; the same is true for RCP4.5 until the 2340s, when the mode jumps to the high end of the distribution (from 1.7 m to 4.6 m) and remains there (as seen for 2500). Almost all of the long-term uncertainty arises from MICI. The no-MICI projections remain narrow over multiple centuries—particularly for RCP8.5, which becomes more narrow—because the sea-level contribution in the DP16 ensemble depends on the other two parameters (that control ice-shelf vulnerability and basal melting)

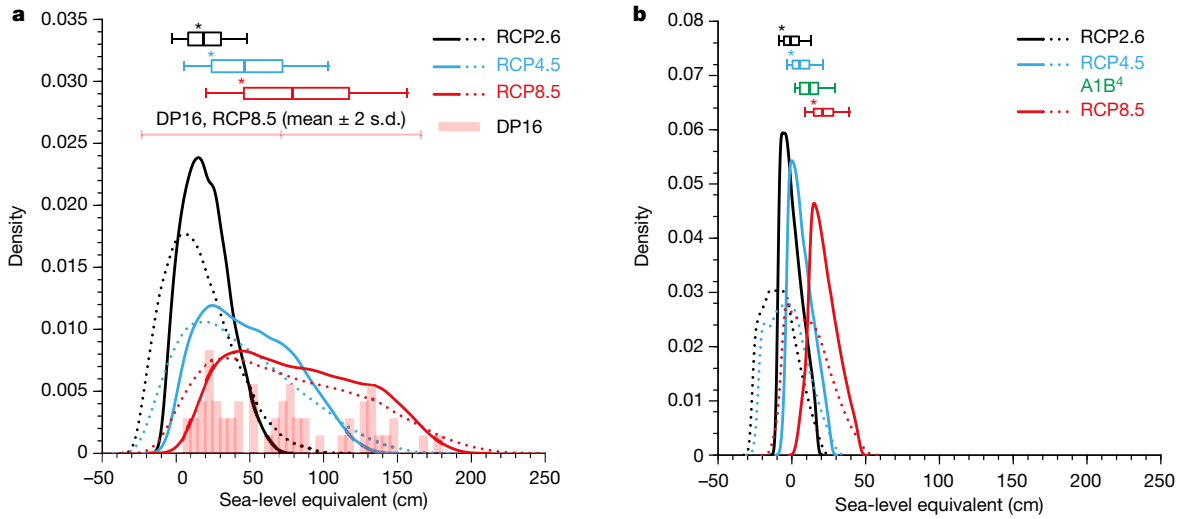


Fig. 1 | Probabilistic projections of the Antarctic contribution to sea level at 2100. a, **b**, Projections estimated under three RCPs with (**a**) and without (**b**) MICI parameterization, obtained from our emulation of the DP16 ice-sheet model ensemble. Dotted lines are density estimates from the uncalibrated emulator ensemble ($n = 10{,}000$); solid lines are calibrated with reconstructions for the last interglacial and Pliocene and satellite data[25] from 1992 to 2017 ($n = 5{,}532$ in **a**; $n = 5{,}074$ in **b**) . Boxes and whiskers show the 5th, 25th, 50th, 75th and 95th percentiles; asterisks indicate the mode. The DP16 ensemble members for RCP8.5 (low-Pliocene calibration, bias-corrected and bias-uncorrected combined) are shown as a histogram and mean ± 2 s.d. interval in **a**, scaled to the same height as the calibrated projection (ref. [6] and additional simulations by R. DeConto, personal communication). A projection[4] for the Antarctic contribution due to ice discharge under the medium–high climate scenario A1B (see text) is shown in **b**.

less over the long-term than during this century. This suggests that the DP16 ensemble either substantially under-samples model uncertainties relevant to long-term change or that the model is structurally deficient because the sensitivity to important parameters diminishes under warming. We therefore consider the post-2100 projections to be less reliable.

The projected probabilities of a sea-level contribution exceeding 1 m over time are shown in Fig. 3b. For high probabilities, the difference in exceedance time between projections under RCP8.5 and RCP4.5 is generally much greater than that between projections with and without MICI under RCP8.5. In addition, RCP2.6—which represents strong mitigation of greenhouse gas concentrations broadly consistent with the 2015 Paris Agreement—is the only RCP to ensure a low probability of high sea-level rise.

Multi-model comparisons

In Fig. 4, we compare the emulated projections at 2100 under RCP8.5 and RCP2.6 with other studies. We compare with only probabilistic projections[2–5], because these have a clear interpretation, and studies that incorporate at least some process-based modelling (rather than only expert elicitation or extrapolation), because we are interested in the uncertainties of physical modelling and we expect Antarctica to be governed by different processes in the past and future (which is not accounted for by extrapolation).

We find the emulated no-MICI results agree well with other studies: 95th percentiles are around 30–40 cm under high scenarios and 10–20 cm under low scenarios, despite the use of very different models and approaches (and some differences in scenario and contribution definitions; see Methods). A recent projection[26] that incorporates ice–ocean–atmosphere feedback effects is also consistent (14 cm under RCP8.5, similar to our mode of 15 cm, and the emergence of a signal from mid-century). Our no-MICI projections for RCP4.5 (median, 5 cm; 66% probability interval, [−1, 15] cm) are very similar to those of the 2013 Intergovernmental Panel on Climate Change (IPCC) assessment for 2100 relative to 1986–2005[1] (median, 5 cm; ≥66% probability interval, [−5, 15] cm). Because the IPCC estimates for Antarctic ice discharge do not depend on the greenhouse gas scenario, our projections for RCP2.6 (median, −1 cm; 66% probability interval, [−7, 7] cm) are lower than those of the IPCC (median, 6 cm; ≥66% probability interval, [−4, 16] cm) and those for RCP8.5 (median, 21 cm; 66% probability interval, [13, 31] cm) are higher than the IPCC estimates (median, 4 cm; ≥66% probability interval, [−8, 14] cm).

Table 2 | Projections for the Antarctic contribution to sea level in 2100

	RCP2.6		RCP4.5		RCP8.5	
	No MICI	MICI	No MICI	MICI	No MICI	MICI
Sea-level contributions (cm)						
Mode	−6	15	0	24	15	45
Median	−1	19	5	46	21	79
Mean	0	20	7	49	22	83
68% interval	[−7, 8]	[4, 36]	[−1, 15]	[16, 83]	[13, 32]	[35, 133]
90% interval	[−9, 13]	[−3, 48]	[−3, 21]	[5, 103]	[9, 39]	[20, 157]
Exceedance probabilities						
≥30 cm	–	26%	–	68%	20%	88%
≥50 cm	–	4%	–	46%	–	71%
≥1 m	–	–	–	6%	–	36%

Contributions to sea level (in units of centimetres sea-level equivalent) with and without DP16 MICI parameterization, calibrated using data from the Pliocene, last interglacial and satellite (1997–2017) eras.

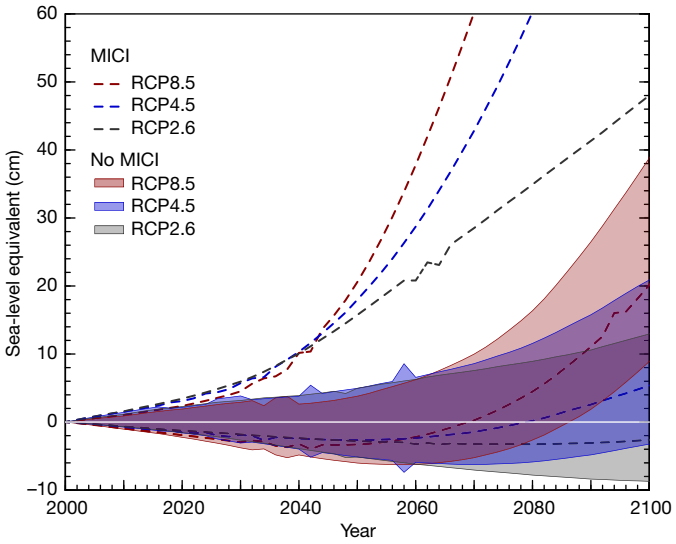


Fig. 2 | Emergence of ice-cliff instability. The projected 5%–95% probability intervals for Antarctic sea-level contributions this century are shown with and without the DP16 MICI parameterization.

Probabilistic interpretations[27] of DP16 have been used to estimate high-end total global-mean sea level by taking the high-Pliocene bias-corrected mean and standard deviation and assuming that the distribution is Gaussian, yielding probabilities of the Antarctic contribution to sea level exceeding 0.5 m and 1 m by 2100 under RCP8.5 of 96% and 65%, respectively. We argue that this interpretation is not justifiable, because the DP16 distributions are skewed (Fig. 1a, Extended Data Fig. 2) and the high-Pliocene constraint is not robust (discussed above). However, using minimal assumptions about the shape of the distribution instead would mean that the probability intervals were very poorly constrained (Table 1). Our estimates of the distribution shape give lower exceedance probabilities: 71% and 36%, respectively (Table 2). We conclude that, although considerable sea-level rise is possible under the probability distributions estimated from DP16, the previous interpretation[27] systematically over-estimates the probability of high sea-level contributions from Antarctica this century.

Only one probabilistic projection has been made[4] beyond 2100. At 2200, our emulated estimates of DP16 projections without MICI under RCP8.5 (median, 4.0 m; 90% probability interval, [3.7, 4.2] m) are an order of magnitude higher than these probabilistic projections[4] under the medium–high Special Report on Emissions Scenario (SRES) scenario A1B[1] (median, 0.41 m; 90% probability interval, [0.04, 0.72] m; Fig. 3a) and more than double previous projections[28] for RCP8.5 (0.88 m and 1.52 m). Beyond 2200, the emulated projections under RCP8.5 become increasingly inconsistent with the latter[28] (Fig. 3a): the no-MICI 2.5th percentile at 2500 is higher even under a doubling of RCP8.5 temperature changes. This is surprising because DP16 greenhouse gas concentrations are capped from the year 2175. However, the no-MICI 90% probability intervals for RCP4.5 and RCP2.6 are consistent with the previous study[28].

These findings suggest that the DP16 model may be over-sensitive to very large atmospheric temperature changes, even without MICI. The response is not self-limiting, owing to widespread ice-shelf sensitivity to warming and/or a lack of local factors mitigating MISI (such as bedrock topography, basal traction and sliding, theoretical constraints on ice stresses at the grounding line and predicted climatic triggers), in contrast to findings from other ice-sheet and ice-shelf models[4,9,14,15,28,29].

Knowledge gaps and future directions

Our analysis has two aims: to estimate the probability distributions implied by DP16 and to evaluate ways the original study could be built on to improve confidence in Antarctic projections. Altering the DP16 climate or ice-sheet models, or extending the parameter ranges of the ensemble, are beyond the scope of this study. For example, we could test the effect of reducing the range of the parameter for ice-cliff collapse (VCLIF; Extended Data Fig. 5), but not of increasing it. These estimates therefore incorporate many of the limitations of DP16, and should be seen as a first step towards a full assessment of Antarctic sea-level uncertainty.

We made pragmatic, simple choices, such as using the same palaeodata intervals as in DP16 and uniform distributions for the parameters. Future work should explore alternatives, such as new sampling of the climate forcing and ice-sheet model parameters, calibrating with palaeodata reconstructions with well-defined uncertainty estimates and spatio-temporal patterns from satellite data and potentially Bayesian calibration methods. We are confident that the tails of the sea-level distributions (which are essential to decision-making) have not been truncated too much by the calibration, because we use a 99.7% probability interval for the satellite data (Methods) and the palaeodata have very little influence (Extended Data Fig. 5). Nevertheless, we present projections only to the 95th percentile, to reflect our judgement about the precision of these estimates. Most importantly, the presence or absence of MICI is by far the largest uncertainty in sea-level rise this century that could be quantified in this study.

Although the maximum height of ice cliffs is founded in theory and supported indirectly by observations and geological evidence[18,30], very little is known about whether initial cliff collapse would lead to a positive feedback effect (that is, MICI), how such a feedback effect would vary in different locations, the consequent rate of ice wastage and how long it would last. MICI might be mitigated by cool, fresh meltwater entering the ocean, buttressing by ice mélange or changes in relative sea-level from gravitational and solid-Earth effects. Greenland's Helheim and Jakobshavn glaciers have high rates of ice wastage, but this is dominated by their fast flow, not grounding-line retreat. Reducing the maximum ice wastage value by 20% to 4 km yr^{-1} reduces the projected median under RCP8.5 by 14% and the 95th percentile by 17% (Extended Data Fig. 5); higher maximum values (which it is not possible to explore in this study) would probably have the opposite effect. The parameterization of ice loss by MICI in DP16 is very simple, and the low resolution of the model might also over-estimate the occurrence of tall cliffs. A range of models and parameterizations is therefore needed.

Triggers are also poorly understood. DP16 predicts early and widespread ice-shelf surface melting (see extended data figure 4 of ref. [6]) and collapse, due to high atmospheric warming, high sensitivity of melting and collapse to warming, or both. This prediction is in contrast to studies that use process-based ice-shelf models, which predict up to 5–6 times less surface melting around the Antarctic Peninsula and 3–8 times less on the West Antarctic Abbot ice shelf by 2100 under RCP8.5[10], and predict that only shelves along the Antarctic Peninsula are vulnerable this century under SRES A1B[9] and RCP8.5[10]. Observational evidence of ice-shelf melting has highlighted amplifying and mitigating processes[31–33], and atmosphere and ocean models have limitations such as present-day biases and missing processes, so further process studies and monitoring are required. The DP16 model shows low sensitivity to ocean melting (figure 6 of DP16) and an apparently unconstrained response to atmospheric warming (Fig. 3a), in contrast to other models[4,9,14,15,28,29,34]. Again, a greater diversity of models is needed, along with standardized extension of greenhouse gas concentration scenarios, to estimate ice-sheet stability on multi-centennial timescales. For the Pliocene, 2 °C ocean warming is used in DP16 whereas a more recent study[35] estimates it to be 3 °C, so the contribution to sea-level rise may be under-estimated.

Using palaeo-reconstructions to calibrate models requires robust quantification of reconstruction uncertainties. History matching typically uses an interval of mean ± 3 s.d., which for continuous and unimodal distributions corresponds to 95% or greater probability[36] for calibration with one observation. The interval for the mid-Pliocene Antarctic contribution to sea-level used here (5–15 m), which provides no

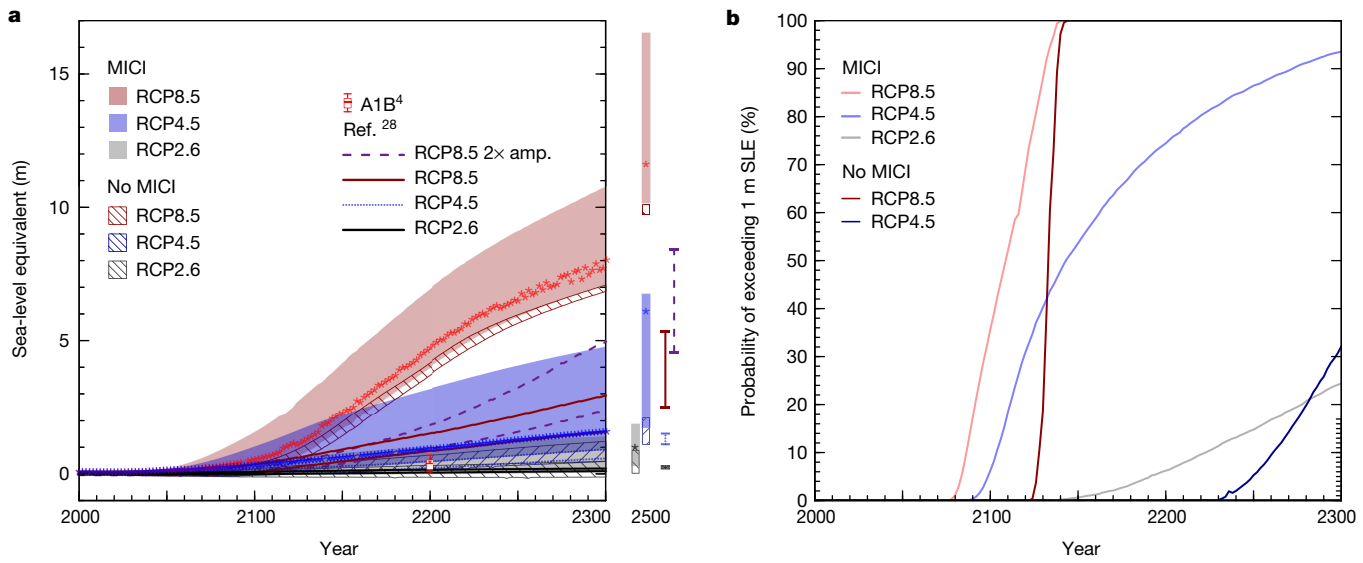


Fig. 3 | Long-term projections of Antarctic sea-level contribution. **a**, Projected 5%–95% probability intervals for Antarctic sea-level contributions to 2300 (time series within figure) and at 2500 (vertical bars on the right), with (shaded) and without (hatched) MICI parameterization, under three RCPs. The asterisks indicate the modes of the distributions with MICI. Previous results[28] under RCP8.5 (solid red line), under RCP8.5 with doubled atmosphere and ocean temperature changes ('2× amp.'; dashed purple line), under RCP4.5 (dotted blue line) and under RCP2.6 (solid black line) are also plotted. The box and whisker plot at 2200 illustrates the 5th, 25th, 50th, 75th and 95th percentiles for a previous projection[4] under the medium–high climate scenario A1B; the asterisk indicates the mode. **b**, Projected probability of the Antarctic sea-level contribution exceeding 1 m sea-level equivalent (SLE) over the same period.

constraint on the DP16 ensemble, is narrower than previous estimates—approximately 4–24 m (95% range) from a reconstruction of total global-mean sea-level change[37], −1 m to 13 m (with less confidence in the lower bound)[20] and 3–14.2 m (95% range)[35] for the early Pliocene. In addition, it has been argued[38] that global-mean sea level during the Pliocene is effectively unknown. For the last interglacial, we assumed that the

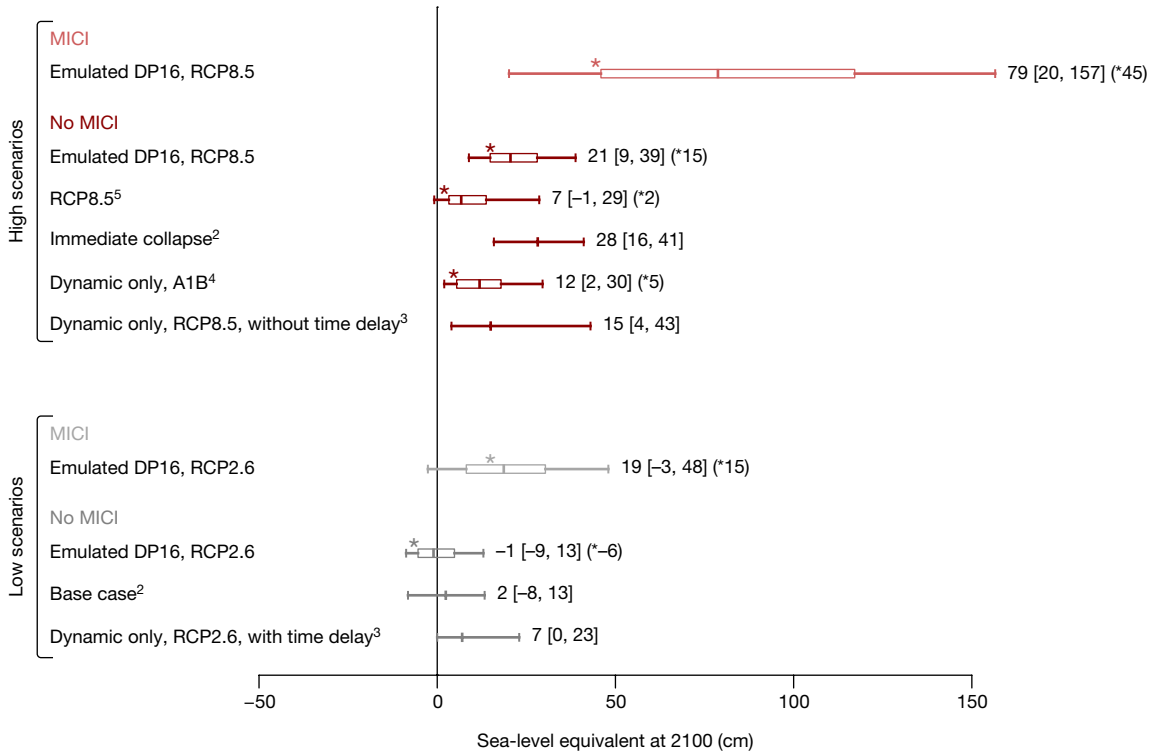


Fig. 4 | Multi-model comparison. Projections from this study at 2100, based on the emulation of DP16, with and without MICI, are shown along with results from other probabilistic modelling studies[2–5]. Boxes and whiskers show the 5th, 25th, 50th, 75th and 95th percentiles; asterisks indicate the mode. Numbers alongside each plot indicate the median, the 5%–95% probability interval and the mode (when available; in parentheses and asterisked; mode for ref. [5] supplied by K. L. Ruckert). 'High' scenarios (pink and red) are for high-end[3,5] (RCP8.5) or medium–high[4] (SRES A1B[1]) greenhouse gas emissions or concentrations, or immediate collapse of part of West Antarctica[2] (in this case the 5th percentile and median are estimated from digitization). 'Low' scenarios (grey and black) are for low greenhouse gas concentrations (RCP2.6)[3] or other baseline cases[2]. The 'dynamic only' data[3,4] include contributions from ice discharge only; high-scenario (low-scenario) data from ref. [3] are from models with ice shelves, without (with) time delay.

DP16 range (3.5–7.4 m) is sufficiently broad; however, one estimate[39] of global-mean sea level suggests a 90% interval for Antarctica of around 1.6–7.5 m and another study that suggests an 80% probability interval of 1.3–13.3 m[40] would nearly eliminate the last interglacial as a constraint. Long-term deformations in Earth's surface could potentially increase estimates of total global-mean sea level at the last interglacial by up to several metres[41]. Emulated projections calibrated with only the satellite era are almost identical to those calibrated with all three eras (Extended Data Fig. 5), indicating that the evaluations with palaeodata have little effect. Using Bayesian calibration (weighting ensemble members by their difference from the data) might yield a stronger constraint, but would require estimates of mean values and error distributions for the reconstructions.

The DP16 ensemble design is not optimal: it includes large gaps and effectively duplicated simulations, and under-samples model uncertainties. Not incorporating model error in the calibration also means that the projections are probably too narrow and over-confident—a problem amplified by sensitivity to the lower bound for the Pliocene. Ensemble designs should be space-filling[4,42] and test which uncertainties are most important to sample (for example, using pre-calibration[43,44]); emulation allows efficient ensemble design and sensitivity analysis. Statistically meaningful calibrations (such as history matching and Bayesian updating, with model discrepancy) provide and demand more information about the data constraints and improve the interpretation and robustness of the resulting projections.

Currently there are few probabilistic Antarctic model projections, and those that exist assess different uncertainties in different ways. We propose for the future a 'grand ensemble', designed across multiple, diverse ice-sheet models, that simultaneously and systematically samples parameters, structures, boundary conditions and initial conditions[34]. Co-ordinated design would enable multi-model emulation—a statistically rigorous method for interpreting and combining different model projections—to estimate probability distributions that account for structural uncertainties across multiple models. The Ice Sheet Model Intercomparison Project (ISMIP6)[45] has been created to make Greenland and Antarctic ice-sheet projections for the next IPCC assessment, and presents an ideal opportunity to design such a framework.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0901-4.

Received: 31 October 2017; Accepted: 4 January 2019;
Published online 6 February 2019.

1. IPCC. *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the IPCC Fifth Assessment Report* (Cambridge Univ. Press, Cambridge, 2013).
2. Little, C. M., Oppenheimer, M. & Urban, N. M. Upper bounds on twenty-first-century Antarctic ice loss assessed using a probabilistic framework. *Nat. Clim. Chang.* **3**, 654–659 (2013).
3. Levermann, A. et al. Projecting Antarctic ice discharge using response functions from SeaRISE ice-sheet models. *Earth Syst. Dynam.* **5**, 271–293 (2014).
4. Ritz, C. et al. Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature* **528**, 115–118 (2015).
5. Ruckert, K. L. et al. Assessing the impact of retreat mechanisms in a simple Antarctic Ice Sheet model using Bayesian calibration. *PLoS One* **12**, e0170052 (2017).
6. DeConto, R. M. & Pollard, D. Contribution of Antarctica to past and future sea-level rise. *Nature* **531**, 591–597 (2016).
7. Pollard, D., DeConto, R. M. & Alley, R. B. Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth Planet. Sci. Lett.* **412**, 112–121 (2015).
8. Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J. & Rae, J. Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. *Nature* **485**, 225–228 (2012).
9. Kuipers Munneke, P., Ligtenberg, S. R. M., van den Broeke, M. R. & Vaughan, D. G. Firn air depletion as a precursor of Antarctic ice-shelf collapse. *J. Glaciol.* **60**, 205–214 (2014).
10. Trusel, L. D. et al. Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios. *Nat. Geosci.* **8**, 927–932 (2015).
11. Vaughan, D. G. West Antarctic Ice Sheet collapse – the fall and rise of a paradigm. *Clim. Change* **91**, 65–79 (2008).
12. Schoof, C. Ice sheet grounding line dynamics: steady states, stability, and hysteresis. *J. Geophys. Res.* **112**, F03S28 (2007).
13. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. *Geophys. Res. Lett.* **41**, 3502–3509 (2014).
14. Favier, L. et al. Retreat of Pine Island Glacier controlled by marine ice-sheet instability. *Nat. Clim. Chang.* **4**, 117–121 (2014).
15. Joughin, I., Smith, B. E. & Medley, B. Marine ice sheet collapse potentially under way for the Thwaites Glacier basin, West Antarctica. *Science* **344**, 735–738 (2014).
16. Waugh, D. W., Primeau, F., DeVries, T. & Holzer, M. Recent changes in the ventilation of the southern oceans. *Science* **339**, 568–570 (2013).
17. Previdi, M. & Polvani, L. M. Climate system response to stratospheric ozone depletion and recovery. *Q. J. R. Meteorol. Soc.* **140**, 2401–2419 (2014).
18. Bassis, J. N. & Walker, C. C. Upper and lower limits on the stability of calving glaciers from the yield strength envelope of ice. *Proc. R. Soc. Lond. A.* **468**, 913–931 (2012).
19. Sweeney, J., Salter-Townshend, M., Edwards, T., Buck, C. E. & Parnell, A. C. Statistical challenges in estimating past climate changes. Wiley Interdiscip. *Rev. Comput. Stat.* **10**, e1437 (2018).
20. Gasson, E., DeConto, R. M. & Pollard, D. Modeling the oxygen isotope composition of the Antarctic ice sheet and its significance to Pliocene sea-level. *Geology* **44**, 827–830 (2016).
21. Williamson, D., Blaker, A., Hampton, C. & Salter, J. Identifying and removing structural biases in climate models with history matching. *Clim. Dyn.* **45**, 1299–1324 (2015).
22. McNeall, D. et al. The impact of structural error on parameter constraint in a climate model. *Earth Syst. Dynam.* **7**, 917–935 (2016).
23. Williamson, D., Blaker, A. T. & Sinha, B. Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model. *Geosci. Model Dev.* **10**, 1789–1816 (2017).
24. Gladstone, R. M. et al. Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth Planet. Sci. Lett.* **333–334**, 191–199 (2012).
25. The IMBIE team. Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature* **558**, 219–222 (2018).
26. Golledge, N. R. et al. Global environmental consequences of twenty-first-century ice-sheet melt. *Nature* **566**, https://doi.org/10.1038/s41586-019-0889-9 (2018).
27. Le Bars, D., Drijfhout, S. & de Vries, H. A high-end sea-level rise probabilistic projection including rapid Antarctic ice sheet mass loss. *Environ. Res. Lett.* **12**, 044013 (2017).
28. Golledge, N. R. et al. The multi-millennial Antarctic commitment to future sea-level rise. *Nature* **526**, 421–425 (2015).
29. Cornford, S. L. et al. Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate. *Cryosphere* **9**, 1579–1600 (2015).
30. Wise, M. G., Dowdeswell, J. A., Jakobsson, M. & Larter, R. D. Evidence of marine ice-cliff instability in Pinelsland Bay from iceberg-keel plough marks. *Nature* **550**, 506–510 (2017).
31. Bell, R. E. et al. Antarctic ice shelf potentially stabilized by export of meltwater in surface river. *Nature* **544**, 344–348 (2017).
32. Kingslake, J., Ely, J. C., Das, I. & Bell, R. E. Widespread movement of meltwater onto and across Antarctic ice shelves. *Nature* **544**, 349–352 (2017).
33. Gourmelen, N. et al. Channelized melting drives thinning under a rapidly melting Antarctic ice shelf. *Geophys. Res. Lett.* **44**, 9796–9804 (2017).
34. Pattyn, F., Favier, L., Sun, S. & Durand, G. Progress in numerical modeling of Antarctic ice-sheet dynamics. *Curr. Clim. Change Rep.* **3**, 174–184 (2017).
35. Golledge, N. et al. Antarctic climate and ice-sheet configuration during the early Pliocene interglacial at 4.23 Ma. *Clim. Past* **13**, 959–975 (2017).
36. Pukelsheim, F. The three sigma rule. *Am. Stat.* **48**, 88–91 (1994).
37. Miller, K. G. et al. High tide of the warm Pliocene: implications of global sea-level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012).
38. Raymo, M. E. et al. The accuracy of mid-Pliocene $\delta^{18}O$-based ice volume and sea level reconstructions. *Earth Sci. Rev.* **177**, 291–302 (2018).
39. Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M. A probabilistic assessment of sea-level variations within the last interglacial stage. *Geophys. J. Int.* **193**, 711–716 (2013).
40. Düsterhus, A., Tamisiea, M. E. & Jevrejeva, S. Estimating the sea-level highstand during the last interglacial: a probabilistic massive ensemble approach. *Geophys. J. Int.* **206**, 900–920 (2016).
41. Austermann, J., Mitrovica, J. X., Huybers, P. & Rovere, A. Detection of a dynamic topography signal in last interglacial sea-level records. *Sci. Adv.* **3**, e1700457 (2017).
42. Nias, I., Cornford, S. L. & Payne, A. J. Contrasting model sensitivity of the Amundsen Sea embayment ice streams. *J. Glaciol.* **62**, 552–562 (2016).
43. Holden, P. B., Edwards, N. R., Oliver, K. I. C., Lenton, T. M. & Wilkinson, R. D. A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1. *Clim. Dyn.* **35**, 785–806 (2010).
44. Edwards, N. R., Cameron, D. & Rougier, J. Precalibrating an intermediate complexity climate model. *Clim. Dyn.* **37**, 1469–1482 (2011).
45. Nowicki, S. M. J. et al. Ice Sheet Model Intercomparison Project (ISMIP6) contribution to CMIP6. *Geosci. Model Dev.* **9**, 4521–4545 (2016).

Acknowledgements T.L.E., N.R.E. and P.B.H. were supported by EPSRC Research on Changes of Variability and Environmental Risk (ReCoVER:

EP/M008495/1) under the Quantifying Uncertainty in Antarctic Ice Sheet Instability (QUAntIS) project (RFFLP 006). T.L.E. was also supported by the EPSRC-funded Past Earth Network (EP/M008363/1) and the Université Joseph Fourier–Grenoble International visitor fund 'Campagne INVITES'. N.R.G. is supported by contract VUW1501 from the Royal Society Te Aparangi. I.J.N. was supported by the NERC iSTAR-C project Dynamical control on the response of Pine Island Glacier (NE/J005738/1) and now by the NASA Sea Level Change programme. A.W. is supported by the Open University Faculty of Science, Technology, Engineering and Mathematics. We thank R. DeConto for running additional simulations necessary for the analysis, suggestions and discussions, and K. Ruckert and A. Levermann for providing data. We thank the statisticians who attended the Past Earth Network events 'Assessing Palaeoclimate Uncertainty' (held jointly with the Environmental Statistics Section of the Royal Statistics Society, Cambridge, August 2016), 'Emulators workshop' (Leeds, June 2017) and writing retreat (Callow Hall, August 2017) for advice, particularly I. Vernon, P. Challenor and J. Rougier. We thank D. Le Bars, D. McNeall, D. Demeritt and the King's Geography Hazards, Risk and Regulation reading group (G. Adamson, F. Liu, A. Razli, L. Ball and A. Heilbron) for comments on the manuscript. We also thank K.-K. Shiu and D. Campbell for supporting T.L.E. in this work.

Author contributions T.L.E. conceived the idea, carried out the analysis, produced the figures and wrote the manuscript. A.W. and P.B.H. performed preliminary analyses. A.J.P., A.W., C.R., G.D., I.J.N., M.A.B. and N.R.G. contributed ideas on glaciological and oceanic aspects; A.W., N.R.E. and P.B.H. contributed ideas on statistical aspects. All authors contributed to writing the manuscript.

Conflict of interest The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0901-4.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.L.E.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Simulator ensemble design. In DP16, three continuous parameters are perturbed, sampling four levels for each in a factorial design to generate $4^3 = 64$ ensemble members. First, the ocean melt factor OCFAC $\in \{0.1, 1, 3, 10\} \times 0.224$ m yr^{-1} °C^{-2} (note that DP16 quotes incorrect units of m yr^{-2} °C^{-2} in two places) controls sub-ice-shelf direct melting, and is defined as the factor by which the default value is multiplied. Second, the crevasse liquid depth CREVLIQ $\in \{0, 50, 100, 150\}$ m (m yr^{-1})$^{-2}$ controls ice-shelf collapse by hydrofracturing due to surface liquid, and is defined as the additional crevasse depth due to surface melt plus the rainfall rate. Third, the maximum net ice wastage rate VCLIF $\in \{0, 1, 3, 5\}$ km yr^{-1} controls cliff failure after ice shelf collapse.

For present-day and future projections, this ensemble is duplicated with the Southern Ocean bias correction (BIAS) applied. When emulating the ice-sheet model (see below) we combine these 128 ensemble members and treat the bias correction as a continuous uncertain parameter, defined as a scalar that ranges from BIAS = 0 (no bias correction, +0 °C) to BIAS = 1 (full bias correction, +3 °C).

We use time-series data for the ensemble provided by R. DeConto. When emulating the model, we found a sign error in the supplementary information of DP16: the value for the last interglacial for simulation row 6 (OCFAC = 0.1, CREVLIQ = 50, VCLIF = 1) should be +2.63 m, not −2.63 m.

Building the emulators. We use Gaussian process regression ('kriging' when used for spatial interpolation) because it is flexible, non-parametric and provides uncertainty estimates[46]. As usual for emulation of computer models, we set the 'nugget' to zero because the ice-sheet model is deterministic. We refer to a single emulator in the main text for simplicity, but this comprises separate emulators for each scalar output: sea-level change for the Pliocene and last interglacial, present-day (1992–2017) change in the RCP4.5 simulation, and the change from 2000 to every even year up to 2500 for the three RCPs. We construct, validate, calibrate and make predictions using the R software packages DiceKriging and a modified version of DiceEvaluation.

Let the function $f(\boldsymbol{x})$ be the ice-sheet model, which simulates sea-level change in a particular era (for example, the Pliocene) as a function of its input parameters $\boldsymbol{x}$. We consider only one output at a time to avoid the need for a further index. An emulator $f_{\mathrm{em}}(\boldsymbol{x})$ for a particular output of $f(\boldsymbol{x})$ can be written as

$$f_{\mathrm{em}}(\boldsymbol{x}) = \sum_j \beta_j g_j(\boldsymbol{x}) + u(\boldsymbol{x})$$

where $\boldsymbol{g}(\boldsymbol{x})$ are known functions of $\boldsymbol{x}$, $\boldsymbol{\beta}$ are regression coefficients and $u(\boldsymbol{x})$ is a stochastic process with a specified covariance function. We wish to select the subset of $\boldsymbol{x}$ that has the most influence on $f_{\mathrm{em}}(\boldsymbol{x})$.

Design and validation of the emulators consists of two parts: a step-wise model-selection procedure, to choose the mean function (that is, which simulator parameters, and interactions between these, to use as regressors), and a leave-one-out (LOO) cross-validation procedure, to evaluate which is the most suitable covariance function and whether each emulator is sufficiently accurate for our purposes. We perform these procedures for six outputs—the two palaeo-eras, the present day and the three RCP projections at 2100—to choose the overall emulator structure. The final fitting of the emulators with the full ensemble data, and their use for prediction, are discussed later.

Mean functions. There are important interactions between parameters; for example, increasing the bias correction (BIAS) increases the effect of the maximum ice wastage rate (VCLIF) on projections. However, we also wish to avoid over-fitting by including too many interaction terms. We use the function stepAIC from the R MASS package to select model terms, testing up to second-order (three-way) interactions between parameters; we use the Bayesian information criterion because it is generally more parsimonious than the Akaike information criterion. The resulting mean functions for the six outputs are

$$g_{\mathrm{palaeo}}(x) \sim (\mathrm{OCFAC}, \mathrm{CREVLIQ}, \mathrm{VCLIF}, \mathrm{CREVLIQ} * \mathrm{VCLIF})$$

for the Pliocene and last interglacial,

$$g_{\mathrm{low}}(x) \sim (\mathrm{OCFAC}, \mathrm{CREVLIQ}, \mathrm{VCLIF}, \mathrm{BIAS}, \mathrm{OCFAC} * \mathrm{VCLIF}, \mathrm{OCFAC} * \mathrm{BIAS}, \mathrm{CREVLIQ} * \mathrm{VCLIF}, \mathrm{VCLIF} * \mathrm{BIAS}, \mathrm{OCFAC} * \mathrm{VCLIF} * \mathrm{BIAS})$$

for the present day and RCP2.6 at 2100, and

$$g_{\mathrm{high}}(x) \sim (\mathrm{OCFAC}, \mathrm{CREVLIQ}, \mathrm{VCLIF}, \mathrm{BIAS}, \mathrm{OCFAC} * \mathrm{VCLIF}, \mathrm{OCFAC} * \mathrm{BIAS}, \mathrm{CREVLIQ} * \mathrm{VCLIF})$$

for RCP4.5 and RCP8.5 at 2100, where $g \sim (a, \ldots)$ means that g is a linear function of $a, \ldots$, and $a * b$ indicates an interaction term.

Covariance functions. The covariance controls the smoothness between data points, with a trade-off between accuracy and over-fitting. We compare the success of different covariance functions—Matern(5/2), Matern(3/2), exponential and power exponential (exponential family, where the exponent can vary between 0 and 2)—using the mean function selected above, and choose the one with the smallest normalized Euclidean distance in a LOO procedure. The LOO procedure involves fitting the emulator to all ensemble members except one (63 of 64 for Pliocene and last interglacial; 127 of 128 for present-day and future), and then predicting the final member to compare with the simulation. This is repeated for all combinations ($N_{\mathrm{ens}} = 64$ or 128) to provide a summary statistic. The normalized Euclidean distance is

$$d = \sqrt{\sum_{i=1}^{N_{\mathrm{ens}}} \frac{[f_{\mathrm{em}}(x_i) - f(x_i)]^2}{\sigma^2_{\mathrm{em},x_i}}}$$

where i identifies the ensemble member and σ_{em} is the emulator error for that prediction. We choose this metric because it uses the uncertainty estimate inherent in a Gaussian process emulator to standardize the residuals, so that an emulator with some large errors is not overly penalized if it has sufficiently large uncertainty estimates to generally encompass the true value. This also guards against over-fitting, by penalizing too-confident emulators. The distance metric therefore balances the two aims of emulator accuracy and appropriate confidence. The resulting covariance functions from this procedure are power exponential for the last interglacial, Matern(3/2) for 1992–2017, and exponential for the Pliocene and future outputs.

Validating and fitting the emulators. We use various validation outputs to assess emulator adequacy: root-mean-square error; Kendall's τ, a non-parametric measure of correlation, for the emulator predictions versus the simulations; and the fraction of predictions for which the simulation lies within the 95% credibility interval of the emulator, for which values lower than about 90% would indicate an over-confident emulator (that is, too-small uncertainty estimates). The root-mean-square error and Kendall's τ coefficients between the emulator predictions and simulations are: 12 cm (1.4% of the data range) and 0.958, respectively, for the Pliocene; 26 cm (2.7%) and 0.923 for the last interglacial; 0.1 cm (0.6%) and 0.972 for the present day; and 0.9–1.2 cm (0.4%–0.8%) and 0.973–0.976 for the three future projection emulators. These values indicate sufficient accuracy. The fraction of predictions within the 95% interval of the emulator is 100% for the Pliocene, 89% for the last interglacial and 91%–98% for the present day and future, indicating sufficiently large uncertainty estimates. The predictive accuracy and uncertainty estimates of the six emulators can be inspected visually by plotting the emulator predictions against the simulations and the standardized residuals (Extended Data Fig. 6).

Having judged these six emulators to be adequate, we fit each emulator with the full ensemble for that output. We use the emulator structures for the year 2100 for all time slices for that RCP.

Emulator ensemble design. We predict 10,000 points in parameter space using a maximin Latin hypercube (that is, efficiently space-filling) design. The MICI design samples from uniform distributions for all four parameters, on the basis of discussions with one of the authors of DP16 (R. DeConto, personal communication); the no-MICI design has VCLIF = 0. The effect of VCLIF, CREVLIQ and OCFAC on sea-level contributions at 2100 under RCP8.5 in the MICI case is shown in Extended Data Fig. 7, which demonstrates the strong dependence on VCLIF. The reason for some apparent gaps in emulator coverage is that the ensemble design is space-filling but does not necessarily sample points in each corner of parameter space, as the original ensemble members do.

Pliocene calibration. The low-Pliocene and high-Pliocene projections of DP16 are presented (and have been interpreted by others[27,47]) as equally plausible, but here we make the case that the high-Pliocene calibration is not robust. This is important because the RCP8.5 projections are uniquely sensitive to the particular minimum value chosen for the high-Pliocene constraint (10 m). In Extended Data Fig. 1a, b we show that a lower bound exceeding 9.6 m results in much higher means and much smaller standard deviations, because fewer than a quarter of the ensemble members pass. The sensitivity is caused by a combination of the small ensemble size and the strong correlation in the model between Pliocene sea-level and RCP8.5 projections (large circles in Extended Data Fig. 3a).

This sensitivity to the Pliocene lower bound is exacerbated by the choice of calibration method, which involves a simple accept or reject. We re-express this method in a history-matching framework[22,48] below. This binary filtering means that we should choose a sufficiently wide range of tolerance, because every rejected ensemble member is treated as completely implausible (by being removed, rather than down-weighted as in Bayesian calibration). Treating two ranges as equally plausible is not coherent, because it implies that values in the ranges 5–10 m and 15–20 m are simultaneously both plausible and implausible. The data range

chosen should be both broad and unique, to obtain a calibration that is robust and meaningful.

The Antarctic contribution to mid-Pliocene sea level has been estimated[20] to have a maximum of 13 m, which would rule out most of the high-Pliocene range. This suggests that the interval 10–20 m is not well supported. (A range of 10–13 m would be inconsistent with the large degree of uncertainty in Pliocene reconstructions[37,49].)

Increasing the upper bound from 13 m would have no effect because the maximum Pliocene change in the ensemble is 12.4 m. Decreasing the lower bound below 5 m would also make little difference because the original (no discrepancy) DP16 calibration for the last interglacial (3.5–7.4 m) rejects these ensemble members. None of the ensemble members that pass the last interglacial constraint have Pliocene sea-level changes of less than 5 m (Extended Data Fig. 4; there are no large circles directly below the shaded box). The crucial judgement is therefore whether the high-Pliocene lower bound of 10 m can be justified.

We conclude that the Pliocene Antarctic sea-level contribution is currently too uncertain to use the high-Pliocene constraint, particularly for this model and for a history matching approach, and that the low-Pliocene calibration is far more robust.

Model discrepancy. Model discrepancy, or structural error, is defined as the smallest possible difference between a model simulation and the true values, that is, how well the model could reproduce reality at its best possible (tuned) parameter values[4,21–23]. Discrepancy is an essential part of model calibration; not incorporating it implies that a model could be tuned to perfectly match reality. Using a value less than the observational error would imply that we could simulate reality better than we could measure it. Model discrepancy can, in some cases, be estimated approximately by comparing simulations with multiple observations. But if there are insufficient observations to do this, as is the case here, then discrepancy can be viewed as a tolerance to model error[48] estimated by expert judgement[4,24] (see below).

Calibrating projections. We re-express the DP16 calibration within a history matching framework, extending it to account for emulator error and model discrepancy. We adapt a previous notation[50] and model the relationship between a palaeodata reconstruction or an observation of sea-level change (Pliocene, last interglacial or 1992–2017 trend) z and the true value y as

$$z = y + \varepsilon_{\text{obs}}$$

where ε_{obs} has variance σ^2_{obs}, the square of the observational or palaeodata reconstruction error. The relationship between the true value and the simulation of this sea-level change is

$$y = f(\boldsymbol{x}^*) + \varepsilon_{\text{md}}$$

where $\boldsymbol{x}^*$ are the best values of the parameters and ε_{md} is the model discrepancy with variance σ^2_{md}. We emulate

$$f(\boldsymbol{x}) = f_{\text{em}}(\boldsymbol{x}) + \varepsilon_{\text{em},\boldsymbol{x}}$$

where $f_{\text{em}}(\boldsymbol{x})$ is the mean emulator prediction for $f(\boldsymbol{x})$ and $\varepsilon_{\text{em},\boldsymbol{x}}$ is the emulator error as before; it varies with $\boldsymbol{x}$ and is automatically estimated in Gaussian process emulation. For a given emulated output (Pliocene, last interglacial or 1992–2017 trend), we can use the standardized distance, which is also known as the implausibility I,

$$I^2(\boldsymbol{x}) = \frac{[f_{\text{em}}(\boldsymbol{x}) - z]^2}{\sigma^2_{\text{obs}} + \sigma^2_{\text{em},\boldsymbol{x}} + \sigma^2_{\text{md}}}$$

to accept or reject a given emulated ensemble member with parameter values $\boldsymbol{x}$. We interpret the accepted ensemble members as a posterior probability distribution. This represents a judgement that this distribution represents our uncertainty about future sea-level rise (given the limitations of the ice-sheet model and palaeodata), that is, that the parameter space outside the calibration intervals has a low probability of being plausible.

We use a minimum palaeodata value for the last interglacial of 3.5 m, rather than the 3.6 m quoted in DP16, for consistency with the calibrated ensemble results therein, which include a member with sea-level change for the last interglacial of 3.53 m.

The palaeodata reconstruction errors are not defined. We conservatively treat the DP16 range as a mean ± 1 s.d. interval, so use $\sigma_{\text{obs}} = 5$ m and $\sigma_{\text{obs}} = 2$ m for the Pliocene and last interglacial, respectively. The observational constraint[25] is the cumulative mass loss from 1992 to 2017 (2,720 ± 1,390 Gt), converted to centimetres sea-level equivalent by dividing by 3,600 to give a sea-level contribution of 0.756 ± 0.386 cm over this period. Model discrepancy is set to $\sigma_{\text{md}} = 0.5$ cm for 1992–2017 sea-level change.

When calibrating with palaeodata, we accept ensemble members with $I < 1$ for the Pliocene and last interglacial, so that the simplest case without emulator and model errors matches the interval used in DP16. This Pliocene range corresponds approximately to a 95% interval in some reconstructions, but the last interglacial range may correspond to a lower probability than 95% by some estimates, and so may be too strict a constraint (see main text). Calibration with satellite data accepts ensemble members with $I < 3$, to follow the usual history matching convention for well-defined errors: for a smooth unimodal distribution, $I < 3$ with probability greater than or equal to 95%[36]; for Gaussian distributions, as we expect for the satellite data errors, the probability interval is 99.7%.

Extended Data Figure 3 shows the calibration relationships for RCP8.5 at 2100—the relationships between past and future. Grey boxes show the original palaeodata constraints; dashed lines show the broader intervals after accounting for model discrepancy. Accounting for emulator error in the implausibility means that some emulator ensemble members are accepted that lie just outside the calibration interval.

Percentiles and exceedance probabilities are estimated directly from the 10,000-member emulator ensemble, and modes from kernel density estimation using an automatic (Silverman) bandwidth. We do not include emulator uncertainties in the distributions; these are small at 2100, but increase on multi-century timescales so would broaden the distributions. To improve the clarity of Fig. 3, we exclude one, three and five data points from each of the RCP8.5, RCP4.5 and RCP2.6 MICI projections, respectively, because the estimates are not continuous in time (owing to slight differences in emulator fitting).

Multi-model comparisons. We show distributions[5] provided by K. Ruckert, and estimate the distribution from another study[2] by digitization of the original figures. We re-estimate the modes from ref. [4] using an automatic bandwidth for the kernel density estimation, rather than the broader, fixed bandwidth used in the original study. We assume that differences due to definitions of time periods are small enough to be ignored; all are 2000–2100 except for ref. [2] (1990–2099) and ref. [1] (1986–2005 to 2081–2100).

Palaeodata uncertainties. Here we consider probability intervals for palaeodata constraints. Peak total sea-level change for the last interglacial has been estimated to be 6.4–10.9 m (90% probability interval)[39] and 6.1–16.7 m (80% probability)[40]. These estimates broadly encompass recent assessments that the upper end of the widely used 6–9-m range[49] could increase by several metres[41]. Subtracting a range of estimates for the contributions from Greenland, thermal expansion and glaciers (3.4–4.8 m)[5] gives Antarctic contributions of 1.6–7.5 m and 1.3–13.3 m, respectively.

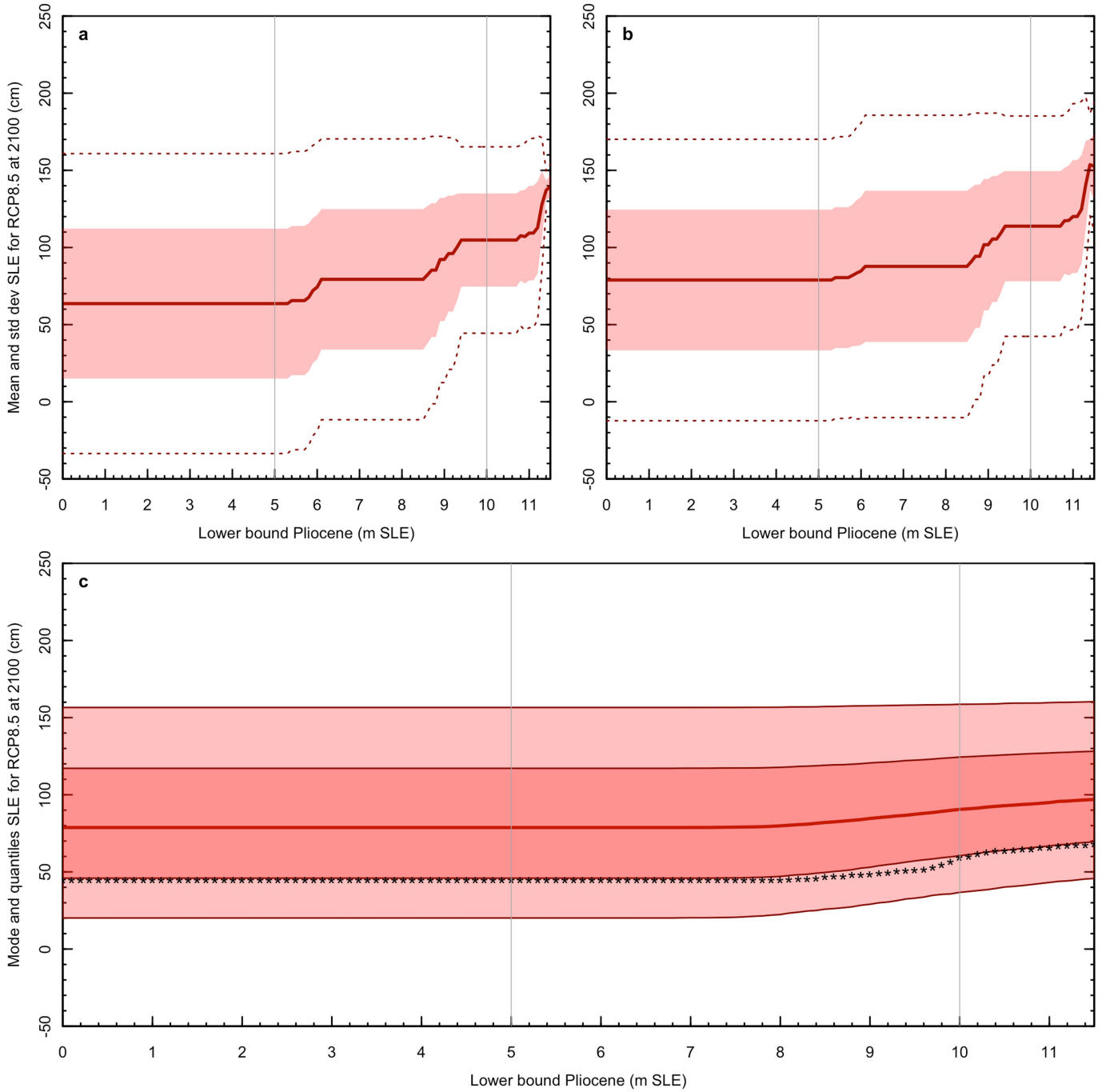
For the Pliocene, a total sea-level change of 22 ± 10 m (95% range) has been estimated[37]; subtracting 7 m for the Greenland Ice Sheet and 1 m for thermal expansion[35] would imply an Antarctic contribution of approximately 14 ± 10 m (4–24 m). There is no difference between using a combined 5–25-m range and using the low-Pliocene (5–15 m) constraint presented here, because the DP16 ensemble maximum is 12.4 m; however, for a different model or ensemble design the upper bound might have more influence. For the early Pliocene, the Antarctic contribution has been estimated[35] to be 8.6 ± 2.8 m; we use this Gaussian assumption to derive the 95% (2σ) range.

Code availability. All emulation was performed in R using the DiceKriging and DiceEvaluation packages, with minor modifications by T.L.E. The scripts and input data for the main analysis (sea-level projections at 2100) are available as a downloadable R package at https://github.com/tamsinedwards/revisitmici (v1.0.2) and can be run without installation on the cloud-based computational reproducibility platform Code Ocean at https://doi.org/10.24433/CO.4ebd8cda-35c0-4d8f-9b7c-d1b064109437.

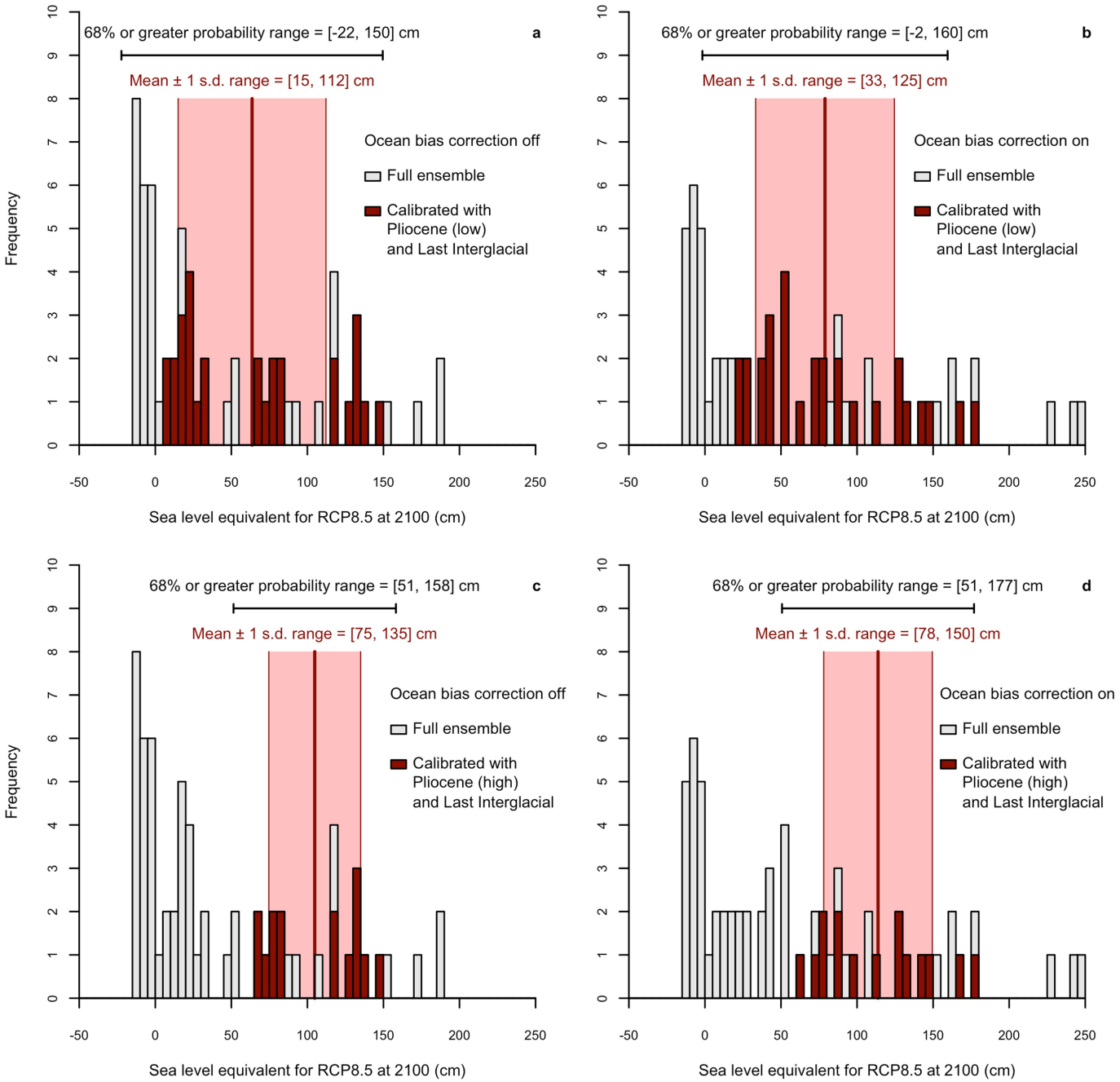
Data availability

All projections from this study are available from the corresponding author on request. Simulations of the last interglacial, Pliocene and 1992–2017 and 2000–2100 sea-level contributions for all DP16 ensemble members are available on Code Ocean (https://doi.org/10.24433/CO.4ebd8cda-35c0-4d8f-9b7c-d1b064109437). Simulations at 2500 for the subset of the DP16 ensemble that pass their calibration are available in the supplementary information of DP16.

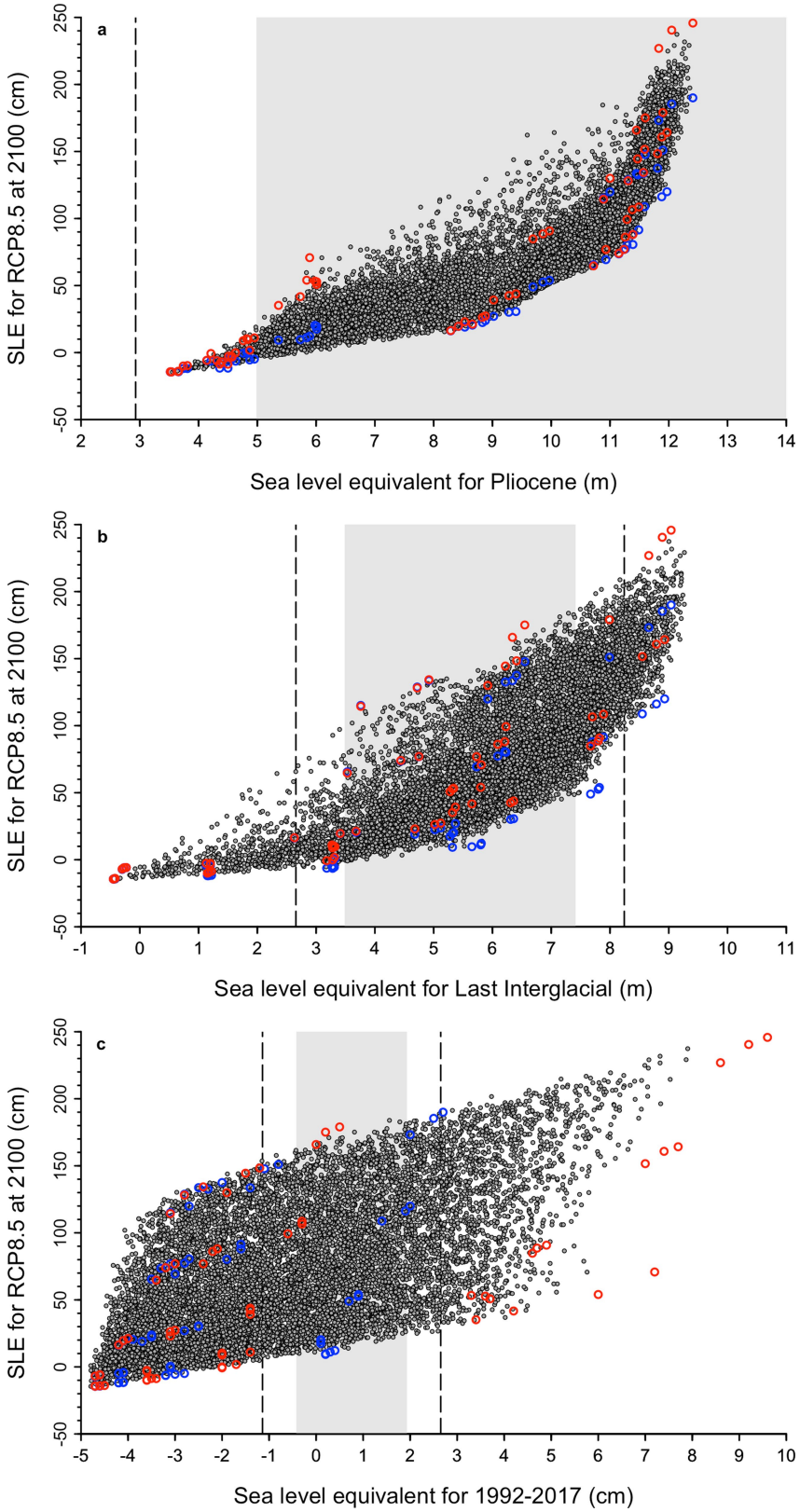
46. O'Hagan, A. Bayesian analysis of computer code outputs: a tutorial. *Reliab. Eng. Syst. Saf.* **91**, 1290–1300 (2006).
47. Sweet, W. V. et al. *Global and Regional Sea-Level Rise Scenarios for the United States*. Report No. NOS CO-OPS 083 (NOAA, 2017).
48. Williamson, D. et al. History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim. Dyn.* **41**, 1703–1729 (2013).
49. Dutton, A. et al. Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, aaa4019 (2015).
50. Vernon, I., Goldstein, M. & Bower, R. G. Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Anal.* **5**, 619–669 (2010).



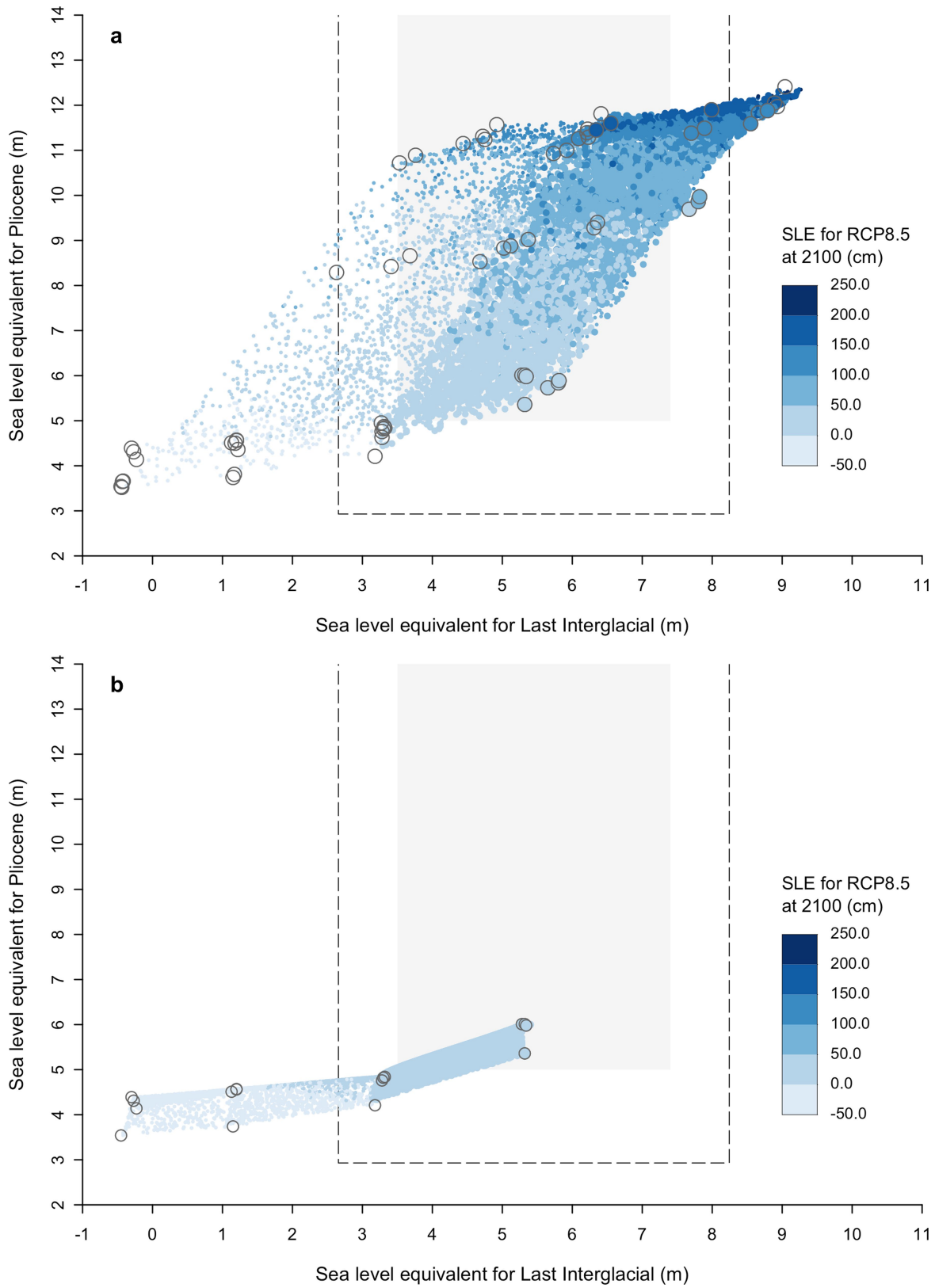
Extended Data Fig. 1 | Sensitivity of DP16 RCP8.5 projections to the lower bound of the Pliocene data. a, **b**, Bias-uncorrected (**a**) and bias-corrected (**b**) DP16 projections[6] for Antarctic sea-level contribution by 2100 under RCP8.5 as a function of the lower bound of the Pliocene data range. The solid red line and pink shading show the mean ± 1 s.d.; the dotted lines indicate ±2 s.d. **c**, Sensitivity of our emulated projections for RCP8.5 at 2100 with MICI: lines show 5th, 25th, 50th, 75th and 95th percentiles; asterisks indicate the mode.



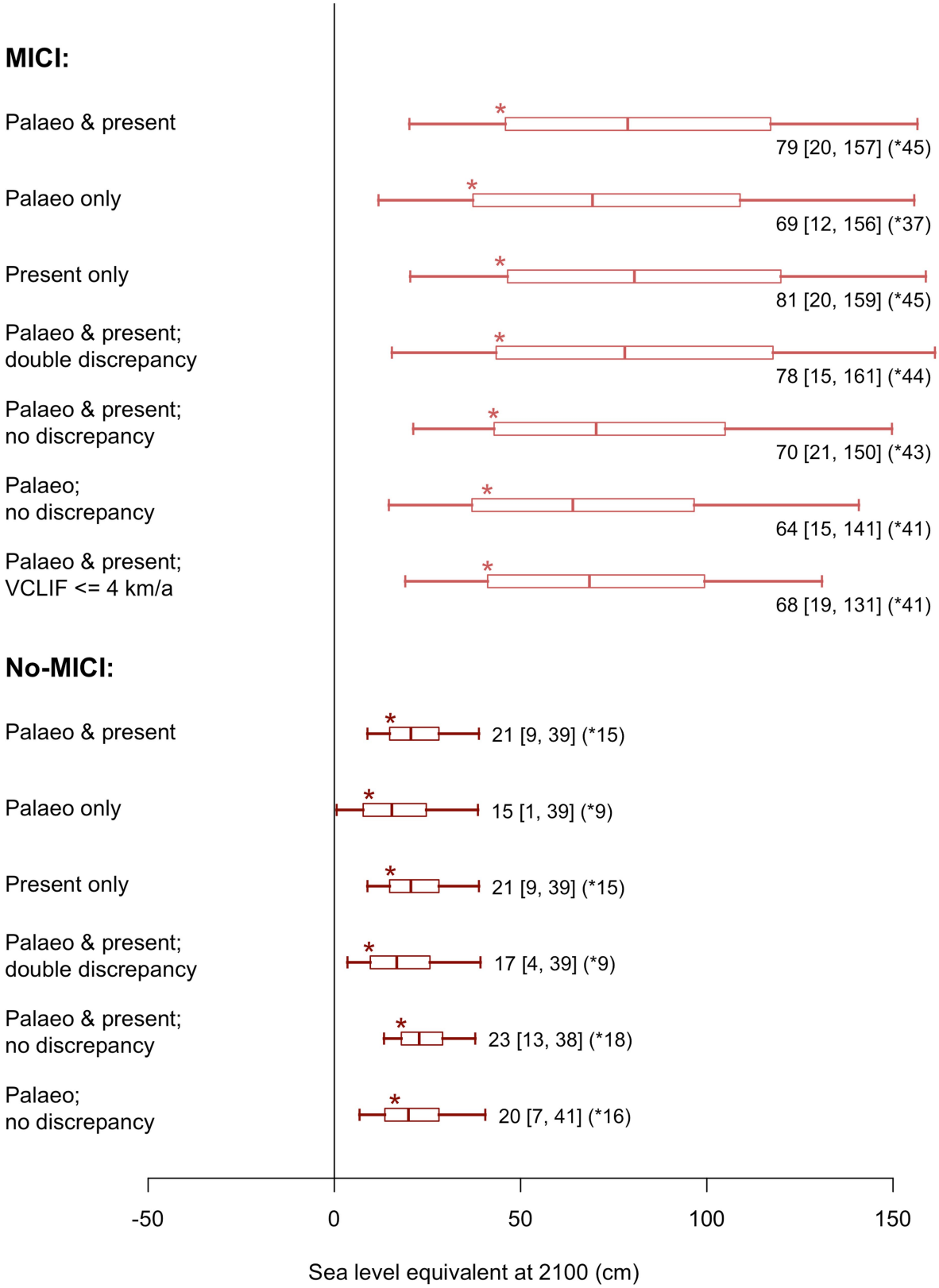
Extended Data Fig. 2 | DP16 RCP8.5 projection distributions. **a**–**d**, DP16 ensemble projections[6] for Antarctic sea-level contribution by 2100 under RCP8.5 for their four variants—low-Pliocene bias-uncorrected (**a**), low-Pliocene bias-corrected (**b**), high-Pliocene bias-uncorrected (**c**) and high-Pliocene bias-corrected (**d**)—showing the full 64-member ensemble (grey) and the subset selected by calibrating with Pliocene and last interglacial sea-level reconstructions (red). The solid red lines and pink shading show the mean ± 1 s.d.; the horizontal black lines indicate the ≥68% probability interval (see main text and Methods for more details).



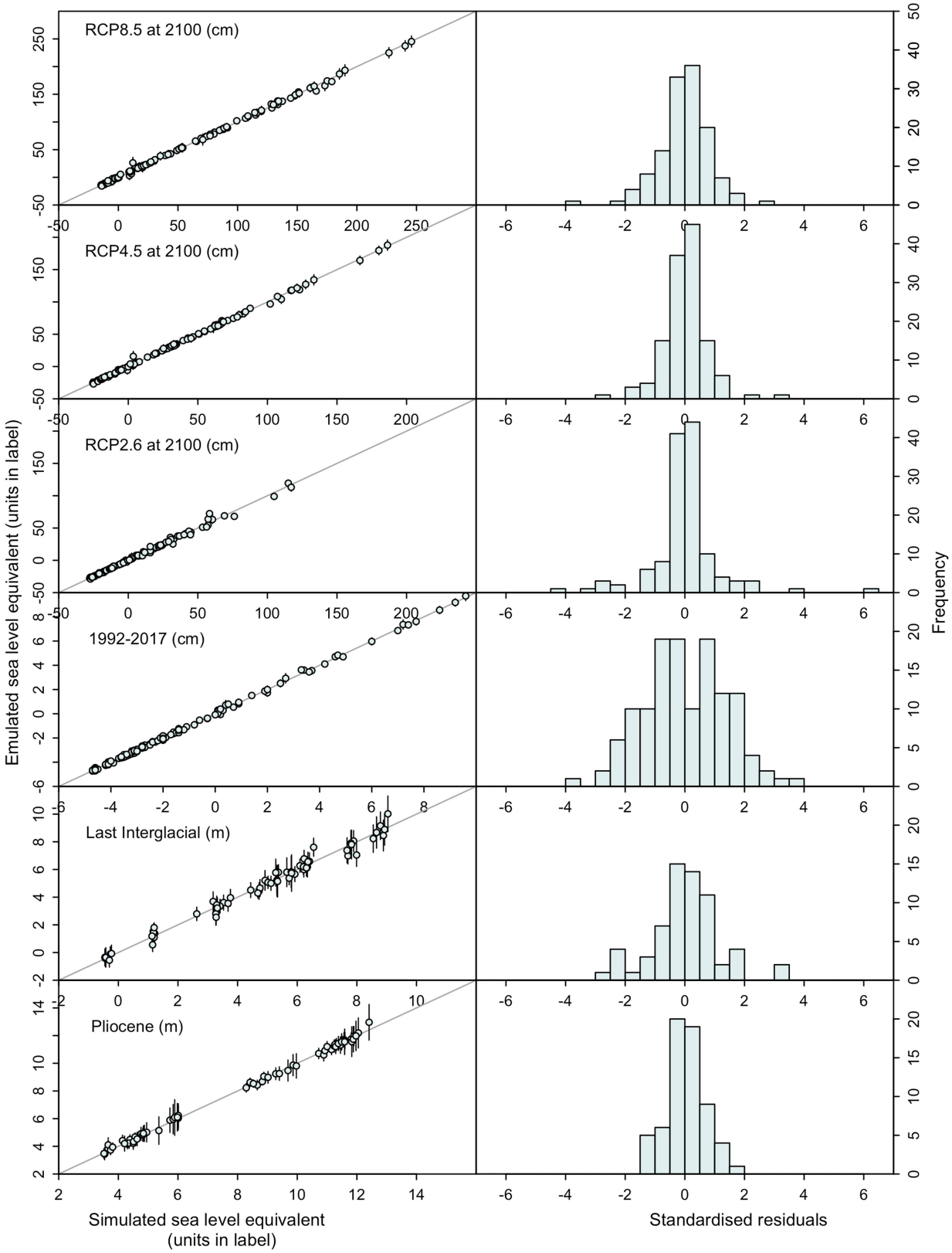
Extended Data Fig. 3 | Relationships between RCP8.5 projections at 2100 and past sea-level changes. a–c, Sea-level contribution at 2100 under RCP8.5 versus sea-level contribution during the Pliocene (**a**), last interglacial (**b**) and from 1992 to 2017 (**c**), for the emulator (small grey dots) and DP16 simulator (large open circles; ref. [6] and R. DeConto, personal communication), with ocean bias correction off (blue) and on (red). Grey shading indicates the DP16 palaeodata range (**a**, **b**) or observational mean ± 3 s.d.[25] (**c**); the dashed lines additionally include model error.



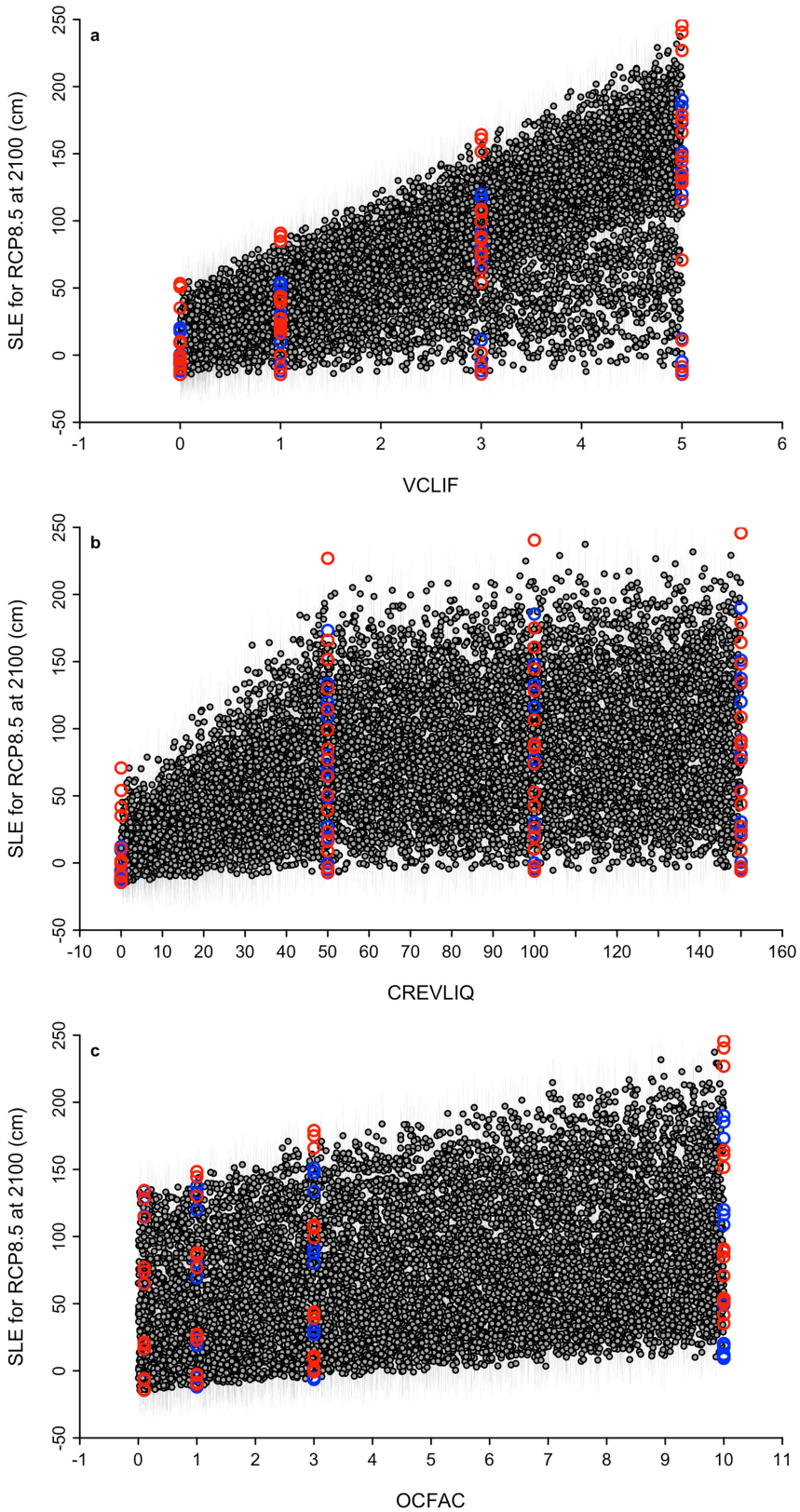
Extended Data Fig. 4 | Relationship between past and future sea-level changes with and without MICI. a, **b**, Simulator ensemble (large circles; ref. [6] and R. DeConto, personal communication) and emulated ensembles (small circles) with (**a**) and without (**b**) MICI, showing Pliocene versus last interglacial sea-level changes, with the colour scale indicating the sea-level equivalent (SLE) contribution at 2100 under RCP8.5. Large emulator points and filled simulator points are those that pass the 1992–2017 calibration[25]. The shaded rectangle indicates the bounds of the DP16 low-Pliocene and last interglacial palaeodata constraints; the dashed rectangle shows constraints in this study (which include model error).



Extended Data Fig. 5 | Sensitivity of RCP8.5 projections to MICI and calibration choices. Projections for RCP8.5 at 2100 are shown with and without MICI, for different combinations of calibration eras ('palaeo', Pliocene and last interglacial; present, 1992–2017) and model discrepancy (with, without or double). Box and whiskers show the 5th, 25th, 50th, 75th and 95th percentiles; asterisks show the mode. Numbers alongside each plot indicate the median, the 5%–95% probability interval and the mode (in parentheses and asterisked).



Extended Data Fig. 6 | Emulator validation. Left column, emulator prediction versus simulation for each ensemble member, with the emulator fitted to the other ensemble members, for each of the outputs used for building and validating emulator structure: RCP8.5, RCP4.5 and RCP2.6 sea-level contribution at 2100; 1992–2017 contribution; last interglacial; and Pliocene. Vertical error bars show 95% credibility intervals. Right column, difference between emulator predictions and simulations, standardized by emulator error, for the same six outputs. Values falling mostly between ±2 indicate that the emulator has adequate uncertainty estimates. Simulation data from ref. [6] and R. DeConto (personal communication).



Extended Data Fig. 7 | Sensitivity of RCP8.5 projections to model parameters. a–c, Sea-level contribution at 2100 under RCP8.5 versus VCLIF (**a**), CREVLIQ (**b**) and OCFAC (**c**) parameters for the emulator (small grey dots with error bars) and simulator (large open circles; blue, bias-uncorrected; red, bias-corrected). Simulation data from ref. [6] and R. DeConto (personal communication).