



King's Research Portal

DOI: 10.1038/s41586-019-0901-4

Document Version Peer reviewed version

Link to publication record in King's Research Portal

Citation for published version (APA):

Edwards, T. L., Brandon, M., Durand, G., Edwards, N., Golledge, N. R., Holden, P., Nias, I., Payne, A. J., Ritz, C., & Wernecke, A. (2019). Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*, *566*, 58-64. https://doi.org/10.1038/s41586-019-0901-4

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research. •You may not further distribute the material or use it for any profit-making activity or commercial gain •You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

1 ARTICLE

2

Revisiting Antarctic ice loss due to marine ice cliff instability

- 5
- 6 Tamsin L. Edwards* [a], Mark Brandon [b], Gael Durand [c], Neil R. Edwards [b], Nicholas
- 7 R. Golledge [d, e], Philip B. Holden [b], Isabel Nias [f], Antony J. Payne [g], Catherine Ritz
- 8 [c] and Andreas Wernecke [b].
- 9 * Corresponding author
- 10

- 12 b School of Environment Earth and Ecosystem Sciences, Faculty of Science, Technology, Engineering and
- 13 Mathematics, Open University, Walton Hall. Milton Keynes, UK
- 14 c Univ. Grenoble Alpes, CNRS, IRD, IGE, F-38000 Grenoble, France
- 15 d Antarctic Research Centre, Victoria University of Wellington, Wellington 6140, New Zealand
- 16 e GNS Science, Avalon, Lower Hutt 5011, New Zealand
- 17 f Earth System Science Interdisciplinary Center, 5825 University Research Court, Suite 4001, College Park,
- 18 Maryland 20740-3823, USA
- 19 g Centre for Polar Observation and Modelling, School of Geographical Sciences, University of Bristol,
- 20 University Road, Bristol BS8 1SS, UK
- 21
- 22

¹¹ a Department of Geography, King's College London

Predictions for sea-level rise from Antarctica this century range from zero to over one 23 metre. The highest are driven by the controversial 'marine ice cliff instability' (MICI) 24 hypothesis, where coastal ice cliffs rapidly collapse after ice shelves disintegrate from 25 surface and sub-shelf melting caused by global warming. But the MICI mechanism has 26 not been observed in the modern era, and it remains unclear whether or not it is 27 required to reproduce sea-level variations in the geological past. Here we quantify ice 28 sheet modelling uncertainties for the original MICI study and show the probability 29 distributions are skewed towards lower values (most likely value is 45 cm under very 30 high greenhouse gas concentrations). However, MICI is not required to reproduce sea-31 level changes in the mid-Pliocene, Last Interglacial or 1992-2017, and without it we find 32 the projections agree with previous studies (all 95th percentiles are less than 43 cm). We 33 therefore find previous interpretations of the MICI projections over-estimate sea-level 34 35 rise this century. The hypothesis is not well constrained: confidence in projections with MICI would require a greater diversity of observationally constrained models of ice 36 shelf vulnerability and ice cliff collapse. 37

Projections of the Antarctic contribution to global mean sea-level rise this century from 38 process-based models vary widely¹⁻⁶. In particular, DeConto and Pollard (2016)⁶ (here DP16) 39 introduced a hypothesised 'marine ice cliff instability' (MICI) process⁷ resulting in mean 40 values exceeding 1 m by 2100 under some methodological choices. However, the DP16 41 results are sensitive to these choices (Table 1: Mean \pm 1 s.d.; Extended Data Figures 1a and 42 b), and the shapes of the probability distributions are very poorly known (Extended Data 43 Figure 2), leading to extremely wide probability intervals (Table 1). This considerable 44 uncertainty poses challenges for robust and cost-effective coastal flood risk management. 45

The Antarctic contribution to global mean sea-level (GMSL) has two parts: increasing 46 snowfall, which is expected to reduce GMSL by a few centimetres this century, and ice 47 discharge into the ocean, which is very uncertain¹. The latter is determined by outflow of ice 48 across the 'grounding line' (the boundary between floating and grounded ice), which can 49 increase due to faster ice flow or inland retreat of the grounding line. Ice discharge can 50 increase if buttressing by ice shelves is reduced by (1) ice shelf thinning, caused by enhanced 51 oceanic melting due to circulation changes⁸ or direct warming, or (2) partial or total ice shelf 52 collapse, caused by widening of surface crevasses by meltwater due to atmospheric 53 warming 9,10 . 54

55 Marine parts of the ice sheet, lying on bedrock below sea-level, are potentially 56 vulnerable to two hypothesised positive feedbacks that may have led to past collapse of the 57 West Antarctic ice sheet¹¹. Both are based on physical mechanisms with theoretical

foundations, but it is not yet clear the degree to which these could lead to positive feedbacks 58 (i.e. widespread, rapid and sustained ice losses). 'Marine Ice Sheet Instability' (MISI)¹² is a 59 self-sustaining retreat of the grounding line in regions where the bedrock slopes downward 60 inland, triggered by ice shelf thinning or collapse. Ice thickness at the grounding line 61 increases (due to the bedrock slope), leading to faster ice flow, causing further retreat. 62 Satellite and modelling evidence suggests MISI is underway in West Antarctica^{13,14,15}, though 63 it is unclear the degree to which the driver of this, warm Circumpolar Deep Water breaching 64 the continental shelf, has been affected by human activities^{1,16,17}. 'Marine Ice Cliff Instability' 65 (MICI)^{6,7} is a self-sustaining retreat of the ice front in regions where the ice is 100 m or more 66 above the ocean surface¹⁸, triggered by ice shelf collapse. These tall ice cliffs are structurally 67 unstable, and their collapse could leave behind further tall cliffs, resulting in sustained ice 68 losses. Observational evidence for MICI is indirect: an absence of ice cliffs taller than 100 m, 69 and rapid retreat of the front of the Jakobshavn (Greenland) and Crane (Antarctic) glaciers 70 (see Knowledge Gaps and Future Directions). 71

DP16⁶ use an Antarctic ice sheet model with a new parameterisation of MICI⁷, 72 generating a 64-member ensemble by varying three parameters controlling the relationship 73 between ocean temperature and basal melting, ice shelf disintegration, and maximum rate of 74 ice cliff collapse. They make projections to 2500 under three Representative Concentration 75 Pathways (RCPs): RCP2.6, RCP4.5 and RCP8.5, for very low, low-to-medium and very high 76 greenhouse gas concentrations respectively¹. They calibrate these by accepting only ensemble 77 members that reproduce reconstructed Antarctic sea-level contributions in the mid-Pliocene 78 (~3 million years ago) and Last Interglacial (LIG: ~130,000-115,000 years ago) eras, and 79 present results for two methodological choices. The first is the Pliocene calibration, using an 80 interval of 5-15 m or 10-20 m; the latter increases sea-level contributions by up to 40 cm by 81 2100 and 2.5 m by 2500 under RCP8.5 (here "LowPliocene"/"HighPliocene"). The second is 82 an ocean temperature correction of +3°C in West Antarctica to improve simulations of the 83 present ice sheet ("BiasCorrected"/"BiasUncorrected"); this increases sea-level contributions 84 by up to 15 cm this century, but makes little difference by 2500. Results for RCP8.5 at 2100 85 are given in Table 1; the corresponding distributions are shown in Extended Data Figure 2. 86

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

94 New projections for Antarctica

We estimate probabilistic projections for the Antarctic contribution to sea-level rise by 95 'emulating' the DP16 ice sheet model (see Methods). This quantifies how a computer 96 model's outputs vary as a function of its input parameters, to predict outputs for any 97 parameter values, enabling us to generate a far larger ensemble than with the original model 98 and to present results both with and without MICI. For this we assume all parameter values 99 are equally likely within the original ranges, based on discussions with the DP16 authors (R. 100 DeConto, pers. comm.). Estimating probability distributions allows meaningful comparison 101 102 with other studies, and decision-making using sea-level exceedance probabilities under both MICI and No-MICI scenarios. Our method has two further additions: calibration with both 103 palaeodata and satellite data, re-expressing this in the statistical framework of 'history 104 105 matching' (see Methods), and accounting for ice sheet model error.

Reconstructions of past climate change provide important tests of models, particularly 106 when the changes were large and/or warmer than today, but their uncertainties are typically 107 large and often poorly-defined¹⁹; recent observations have smaller signals but far smaller 108 uncertainties. The two provide complementary information, so we use both. We use the 109 LowPliocene (equivalent to a combined range of 5-20 m, because the highest simulation is 110 111 12.4 m), for two reasons: the large reconstruction uncertainty (values lower than 10 m cannot be ruled out: e.g. a more recent estimate has a maximum of 13 m²⁰), and because the DP16 112 projections are very sensitive to the lower bound of the HighPliocene (Extended Data Figure 113 2a and b). The 'calibration relationships' between RCP8.5 sea-level contribution at 2100 and 114 sea-level change for the three past eras are shown in Extended Data Figure 3. 115

To estimate probability distributions we use 'history matching' (HM), where 116 implausible model versions are excluded, rather than the more commonly known Bayesian 117 calibration (BC), where model versions are weighted by their agreement with observations 118 using a likelihood function (metric of model success). This is for several reasons. The 119 concept of HM is the same as DP16, which allows us to make a simpler and more transparent 120 comparison. This method effectively estimates what DP16 would have found if they had 121 substantially greater computing resources, calibrated their ensemble with satellite data, and 122 accounted for model error. History matching is also more 'cautious' than Bayesian model 123 124 calibration: if no model versions match the data, they are all excluded, while BC retains all and upweights the 'least bad'. Finally, we do not know the shape of the crucial Bayesian 125 likelihood function for the Pliocene and LIG: this would require estimates of the palaeodata 126

mean and error distribution, rather than assuming all values within the interval are equally
likely. Guessing these might shift (wrong mean) or narrow (wrong distribution) the final
probability distributions.

Accounting for model error, 'discrepancy'²¹, widens the calibration intervals of 130 acceptance (Extended Data Figures 3 and 4: from grey shaded boxes to dashed lines) and is 131 necessary to avoid over-confidence^{22.23}: the aim is to account for model structural error and 132 other model uncertainties not sampled in the ensemble. These discrepancy terms are 133 tolerances that reflect how well we expect the ice sheet model to reproduce reality. We 134 specify them using expert judgement, including the judgement that they are greater than 135 reconstruction/observation errors^{4,24} (i.e. we judge that confidence in simulating reality with 136 the ice sheet model is lower than in observing or reconstructing it from measurements). 137 Reconstruction errors are not defined by DP16, so we conservatively use half the palaeodata 138 139 range to avoid underestimating uncertainty (Pliocene: 5 m; LIG: 2 m). For the satellite period, the sea-level change is (0.756 ± 0.386) cm for 1992-2017²⁵; we conservatively specify the 140 model error as 0.5 cm. 141

We present projections at 2100 in Figure 1 and Table 2. The distributions are skewed: modes are consistently lower than medians and means. The results are not strongly dependent on the Pliocene calibration lower bound, unlike the DP16 ensemble, due to the much larger ensemble size (Extended Data Figure 2c: RCP8.5 at 2100 with MICI). Emulated projections without MICI are much lower than those with MICI, and are consistent with previous projections by Ritz et al. (2015)⁴ (Figure 1b). The results are robust to changes in calibration era and discrepancy (Extended Data Figure 5).

Crucially, our results show ice cliff instability is not required to reproduce sea-level 149 changes in these three very different eras: 55% of the MICI and 51% of the No-MICI 150 emulator ensemble members simultaneously pass calibration with the Pliocene, LIG and 151 satellite eras (Extended Data Figure 4: larger emulator blue circles within dashed box). MICI 152 increases the ensemble range to encompass more of the data intervals, but the emulator can 153 identify many more areas of the model's parameter space that are successful: including many 154 without MICI. MICI is therefore not necessary for realistic simulations of these periods, so 155 this positive feedback hypothesis cannot be confirmed or ruled out with this data and 156 calibration method. In fact, the Pliocene does not rule out any ensemble members, because 157 158 accounting for model error widens the calibration interval to accept them all (Extended Data Figure 3a). 159

The original DP16 projections have substantial probabilities of net sea-level fall this century, with the RCP8.5 LowPliocene mean ± 1 s.d. envelope including negative values until the 2070s. The emulated projections reflect this (Figure 2), though with lower probability (5th percentile negative until the 2070s). Calibration selects mostly positive sealevel contributions during 1992-2017 (Extended Data Figure 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 hypothesised MICI feedback would accelerate sea-level rise. Contributions with MICI quickly start to diverge from those without for all RCPs: in the 2020s (95th percentiles: Figure 2), resulting from the Antarctic Peninsula (DP16: Figure 4c). Dependence of the Antarctic contribution on RCP with MICI begins mid-century, while 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 172 (Figure 3a). The RCP8.5 distribution remains very skewed, with the mode at the low end of 173 174 the range; the same is true of 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). Virtually all 175 the long-term uncertainty arises from MICI. The No-MICI projections remain narrow over 176 multiple centuries - particularly for RCP8.5, which becomes more narrow - because the sea-177 level contribution in the DP16 ensemble depends less on the parameters controlling ice shelf 178 vulnerability and basal melting over the long-term than during this century. This suggests the 179 DP16 ensemble either seriously under-samples model uncertainties relevant to long-term 180 change, or the model is structurally deficient because the sensitivity to important parameters 181 diminishes under warming. We therefore consider the post-2100 projections to be less 182 reliable. 183

The projected probabilities of exceeding 1 m sea-level contribution over time are shown in Figure 3b. These show that, for high probabilities of first exceeding 1 m Antarctic contribution to sea-level, the difference in exceedance time between RCP8.5 and RCP4.5 greenhouse gas concentration scenarios is generally much greater than between projections with or without MICI under RCP8.5. They also show that RCP2.6, i.e. strong mitigation of greenhouse gas concentrations broadly consistent with the 2015 Paris Agreement, is the only one of these scenarios to ensure a low probability of high sea-level rise.

191

192 Multi-model comparisons

Figure 4 compares the emulated projections at 2100 under RCP8.5 and RCP2.6 with other studies. We compare only with probabilistic projections²⁻⁵, 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 physical modelling
uncertainties and we expect the Antarctica to be governed by different processes in the future
than the past (which is not accounted for in extrapolation).

We find the emulated No-MICI results agree well with other studies: 95th percentiles 199 are around 30-40 cm under high scenarios and 10-20 cm under low scenarios, despite the use 200 of very different models and approaches (and some differences in scenario and contribution 201 definitions; see Methods). A recent projection by Golledge et al. (2018)²⁶ incorporating ice-202 ocean-atmosphere feedbacks is also consistent (14 cm under RCP8.5. compared with DP16-203 based mode of 15 cm; emergence of signal from mid-century). The No-MICI projections for 204 RCP4.5 are very similar to the IPCC (2013) assessment for 2100 relative to 1986–2005¹ 205 (Emulated DP16: 5 [-1, 15] cm median and 66% probability interval; IPCC: 5 [-5, 15] cm 206 median and 66% or greater probability). The IPCC (2013) estimates for Antarctic ice 207 208 discharge do not depend on greenhouse gas scenario, so the projections for RCP2.6 are slightly lower than the IPCC (Emulated DP16: -1 [-7, 7] cm, IPCC: 6 [-4, 16] cm) and 209 higher for RCP8.5 (Emulated DP16: 21 [13, 31] cm, IPCC: 4 [-8, 14] cm). 210

Le Bars et al. (2017)²⁷ make probabilistic interpretations of DP16 for assessing high-211 end total GMSL by taking the HighPliocene BiasCorrected mean and standard deviation and 212 assuming the distribution is Gaussian. This gives probabilities of exceeding 0.5 m and 1 m 213 Antarctic contribution by 2100 under RCP8.5 of 96% and 65%, respectively. We argue this 214 interpretation is not justifiable, as the original DP16 distributions are skewed (Extended Data 215 Figure 2) and the HighPliocene constraint is not robust (discussed above). Using minimal 216 assumptions about the distribution shape instead would mean probability intervals were very 217 poorly constrained (Table 1). Our estimates of the distribution shape give lower exceedance 218 probabilities: 71% and 36%, respectively (Table 2); We conclude that, although significant 219 sea-level rise is possible under the probability distributions estimated from DP16, Le Bars et 220 221 al. (2017) systematically overestimate the probability of high sea-level contribution from Antarctica this century. 222

Only Ritz et al. (2015)⁴ have made probabilistic projections beyond 2100. At 2200, the 223 emulated No-MICI projections under RCP8.5 are an order of magnitude higher than Ritz et 224 al. (2015) projections under the medium-high A1B scenario (Figure 3a; emulated median and 225 90% probability interval: 4.0 [3.7, 4.2] m; Ritz et al. (2015): 0.41 [0.04, 0.72] m) and more 226 than double the projections by Golledge et al. (2015)²⁸ for RCP8.5 (0.88 m and 1.52 m at 227 2200 for two model versions). Beyond 2200, the DP16-derived projections under RCP8.5 228 become increasingly inconsistent with Golledge et al. (2015) (Figure 3a). The 2.5th percentile 229 at 2500 without MICI is higher than the latter's projections at 2500 even under a doubling of 230

RCP8.5 temperature changes. This is particularly surprising, given DP16 greenhouse gas concentrations are capped from the year 2175. However, the RCP4.5 and RCP2.6 No-MICI projections are consistent: the Golledge et al. (2015) ranges fall within the 90% probability intervals.

This suggests the DP16 model may be over-sensitive to very large atmospheric temperature changes, even without MICI: i.e. the response is not self-limiting, due to widespread ice shelf sensitivity to warming and/or a lack of local factors mitigating MISI (e.g. bedrock topography, basal traction and sliding, theoretical constraints on ice stresses at the grounding line, and predicted climatic triggers), in contrast to findings from a diversity of other ice sheet and ice shelf models^{4,9,14,15,28,29}.

241

242 Knowledge gaps and future directions

Our analysis has two aims: to make best estimates of the probability distributions implied by 243 the DP16 study and satellite record, and to evaluate ways in which the original study could be 244 built upon to improve confidence in Antarctic projections. Altering the DP16 climate or ice 245 246 sheet models, and extending the ensemble parameter ranges, are beyond the scope of this study. For example, we could test the effect of reducing the range of the ice cliff collapse 247 parameter VCLIF (Extended Data Figure 5), but not increasing it. These estimates therefore 248 249 incorporate many of the limitations of DP16, and should be seen as a first step towards a full assessment of Antarctic sea-level uncertainty. 250

We made pragmatic, simple choices, such as using the same palaeodata intervals as 251 DP16 and uniform distributions for the parameters. Future work should explore alternatives: 252 sampling of the parameter space, palaeodata reconstructions with well-defined uncertainty 253 estimates, spatio-temporal patterns from satellite data, and Bayesian calibration. We are 254 confident that the tails of the sea-level distributions (essential to decision-making) have not 255 been truncated too much by the calibration, as we use a 99.7% probability interval for the 256 satellite data (see Methods) and the palaeodata have very little influence (Extended Data 257 Figure 5). Nevertheless we present projections only to the 95th percentile, to reflect our 258 judgement about the precision of these estimates. Most importantly, presence or absence of 259 MICI is by far the largest uncertainty in sea-level rise this century that could be quantified in 260 this study. 261

Although the maximum height of ice cliffs is founded in theory and indirectly supported by observations and geological evidence^{18,30}, very little is known about whether initial cliff collapse would lead to a positive feedback (i.e. MICI), how this would vary in

different locations, the consequent rate of ice wastage, and how long it would last. MICI 265 might be mitigated by cool, fresh meltwater entering the ocean, buttressing by ice mélange, 266 or changes in relative sea-level from gravitational and solid earth effects. Greenland's 267 Helheim and Jakobshavn glaciers have high rates of ice wastage, but this is dominated by 268 their fast flow, not grounding line retreat. Reducing the maximum ice wastage value by 20% 269 to 4 km/a reduces the RCP8.5 projected median by 14% and the 95th percentile by 17% 270 (Extended Data Figure 5), and higher maximum values (which it is not possible to explore in 271 this study) would likely have the opposite effect. The parameterisation of ice loss by MICI in 272 DP16 is very simple, and the low resolution of the model might also over-estimate the 273 occurrence of tall cliffs. A diversity of model parameterisations is therefore needed. 274

Triggers are also poorly understood. DP16 predict early and widespread surface melting 275 (DP16: Extended Data Figure 4) and ice shelf collapse, due to high atmospheric warming, 276 high sensitivity of melting/collapse to warming, or both. This is in contrast with studies using 277 process-based models, which predict up to 5-6 times less surface melting around the 278 Peninsula and 3-8 times less on the West Antarctic Abbot ice shelf by 2100 under RCP8.5¹⁰, 279 and that only shelves along the Peninsula are vulnerable this century under SRES A1B⁹ and 280 RCP8.5¹⁰. Observational evidence of ice shelf melting has highlighted both amplifying and 281 mitigating processes³¹⁻³³, and atmosphere and ocean models have limitations such as present 282 day biases and missing processes, so further process studies and monitoring are required. The 283 DP16 model shows low sensitivity to ocean melting (DP16 Figure 6) and apparently 284 unconstrained response to atmospheric warming (Figure 3a), in contrast with other 285 models^{4,9,14,15,28,29,34}. Again, a greater diversity of models is needed, along with standardised 286 extension of greenhouse gas concentration scenarios, in order to estimate ice sheet stability 287 on multi-centennial timescales. For the Pliocene, DP16 apply a 2°C ocean warming but 288 Golledge et al. (2017)³⁵ estimate this was 3°C, so the contribution to sea-level rise may be 289 under-estimated. 290

Using palaeo-reconstructions to calibrate models requires robust quantification of their 291 uncertainties. History matching calibrations typically use a mean ± 3 s.d. interval, which for 292 continuous and unimodal distributions corresponds to 95% or greater probability³⁶ for 293 calibration with one observation. For the Pliocene, total GMSL change reconstructed by 294 Miller et al. (2012)³⁷ implies an Antarctic contribution of approximately 4-24 m (95% range), 295 Gasson et al. (2016)²⁰ estimated an Antarctic contribution of -1 to 13 m (with less confidence 296 in the lower bound), Golledge et al. $(2017)^{35}$ estimated an Antarctic contribution in the early 297 Pliocene of 3-14.2 m (95% range) – all equivalent to, or wider than, the interval used here 298 (i.e. no constraint) – while Raymo et al. $(2018)^{38}$ argue that Pliocene GMSL is effectively 299

unknown. For the LIG, we have assumed the DP16 range (3.5-7.4 m) is sufficiently broad, 300 but the GMSL estimate by Kopp et al. (2013)³⁹ implies a 90% interval for Antarctica of 301 around 1.6-7.5 m, while the 80% probability interval implied by Dusterhus et al. (2016)⁴⁰ 302 (1.3-13.3 m) would virtually eliminate the LIG as a constraint. Long-term deformations in the 303 earth's surface have also recently been estimated to potentially increase estimates of total 304 GMSL at the LIG by up to several metres⁴¹. In fact, emulated projections calibrated only with 305 the satellite period are virtually identical to those calibrated with all three eras (Extended 306 Data Figure 5), indicating that these evaluations with palaeodata have little impact. Using 307 Bayesian calibration (weighting ensemble members by their difference with the data) might 308 vield a stronger constraint, but this would require estimates of mean values and error 309 distributions (e.g. Gaussian). 310

The DP16 ensemble design is not optimal: it includes large gaps and effectively 311 duplicated simulations, and under-samples model uncertainties. Failing to incorporate model 312 error in the calibration also means their projections are likely too narrow and over-confident, 313 a problem amplified by sensitivity to the Pliocene lower bound. Ensemble designs should be 314 space-filling^{4,42} and test which uncertainties are most important to sample (e.g. 'pre-315 calibration^{,43,44}); emulation allows efficient ensemble design and sensitivity analysis. 316 Statistically-meaningful calibrations (such as history matching and Bayesian updating, with 317 model discrepancy) improves interpretation of the data constraints and robustness and 318 interpretation of the resulting projections. 319

Currently there are few probabilistic Antarctic model projections, and they assess 320 different uncertainties in different ways. We propose a new vision of a 'grand ensemble' 321 designed across multiple diverse ice sheet models simultaneously, systematically sampling 322 parameters, structures, boundary conditions and initial conditions³⁴. Co-ordinated design 323 would allow multi-model emulation, a statistically rigorous method of interpreting and 324 325 combining different model projections, to estimate probability distributions that account for multiple model structural uncertainties. The Ice Sheet Model Intercomparison Project 326 (ISMIP6) is bringing together an international consortium of ice sheet modellers to make 327 projections for the Greenland and Antarctic ice sheets⁴⁵; this presents an ideal opportunity to 328 design such a framework. 329

331 **References**

IPCC. Working Group I Contribution to the IPCC Fifth Assessment Report 1. 332 Climate Change 2013: The Physical Science Basis. (Cambridge University 333 Press, 2013). 334 2. Little, C. M., Oppenheimer, M. & Urban, N. M. Upper bounds on twenty-first-335 century Antarctic ice loss assessed using a probabilistic framework. Nature 336 *Climate Change* 3, 1–6 (2013). 337 Levermann, A. et al. Projecting Antarctic ice discharge using response functions 3. 338 from SeaRISE ice-sheet models. Earth Syst. Dynam. 5, 271-293 (2014). 339 4. Ritz, C. et al. Potential sea-level rise from Antarctic ice-sheet instability 340 constrained by observations. Nature 528, 115-118 (2015). 341 5. Ruckert, K. L. et al. Assessing the Impact of Retreat Mechanisms in a Simple 342 Antarctic Ice Sheet Model Using Bayesian Calibration. PLOS ONE 12, 343 344 e0170052 (2016). DeConto, R. M. & Pollard, D. Contribution of Antarctica to past and future sea-345 6. level rise. Nature 531, 591-597 (2016). 346 Pollard, D., DeConto, R. M. & Alley, R. B. Potential Antarctic Ice Sheet retreat 347 7. driven by hydrofracturing and ice cliff failure. Earth and Planetary Science 348 Letters 412, 112–121 (2015). 349 Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J. & Rae, J. Twenty-8. 350 first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal 351 current. Nature 485, 225-228 (2012). 352 Kuipers Munneke, P., Ligtenberg, S. R. M., Van den Broeke, M. R. & Vaughan, 353 9. D. G. Firn air depletion as a precursor of Antarctic ice-shelf collapse. Journal of 354 355 Glaciology 60, 205–214 (2014). Trusel, L.D. et al. Divergent trajectories of Antarctic surface melt under two 10. 356 twenty-first-century climate scenarios. Nature Geoscience, 8(12), pp.927–932 357 (2015). 358 Vaughan, D. G. West Antarctic Ice Sheet collapse – the fall and rise of a 359 11. paradigm. Climatic Change 91, 65-79 (2008). 360 Schoof, C. Ice sheet grounding line dynamics: Steady states, stability, and 361 12. hysteresis. Journal of Geophysical Research 112, F03S28 (2007). 362 Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. 13. 363 364 Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. Geophysical Research 365 Letters 41, 3502–3509 (2014). 366 Favier, L. et al. Retreat of Pine Island Glacier controlled by marine ice-sheet 14. 367 instability. Nature Climate Change 5, 117–121 (2014). 368 Joughin, I., Smith, B. E. & Medley, B. Marine Ice Sheet Collapse Potentially 15. 369 Under Way for the Thwaites Glacier Basin, West Antarctica. Science 344, 735-370 738 (2014). 371 Waugh, D. W., Primeau, F., DeVries, T. & Holzer, M. Recent Changes in the 16. 372 Ventilation of the Southern Oceans. Science 339, 568-570 (2013). 373 Previdi, M. & Polvani, L. M. Climate system response to stratospheric ozone 374 17. 375 depletion and recovery. *Q J Roy Meteor Soc* 140, 2401–2419 (2014). Bassis, J. N. & Walker, C. C. Upper and lower limits on the stability of calving 18. 376 glaciers from the yield strength envelope of ice. Proceedings of the Royal Society 377 A: Mathematical, Physical and Engineering Sciences 468, 913–931 (2012). 378 Sweeney, J., Salter-Townshend, M., Edwards, T., Buck, C. E. & Parnell, A. C. 19. 379 Statistical Challenges in Estimating Past Climate Changes. WIREs 380 Computational Statistics, 10 (5), e1437 (2018). 381

382		https://doi.org/10.1002/wics.1437
383	20.	Gasson, E., DeConto, R. M. & Pollard, D. Modeling the oxygen isotope
384		composition of the Antarctic ice sheet and its significance to Pliocene sea-level.
385		<i>Geology</i> 44, 827–830 (2016).
386	21.	Williamson, D., Blaker, A., Hampton, C. & Salter, J. Identifying and removing
387		structural biases in climate models with history matching. <i>Climate Dynamics</i> 45,
388		1299–1324 (2014).
389	22.	McNeall, D. et al. The impact of structural error on parameter constraint in a
390		climate model. Earth Syst. Dynam. 7, 917–935 (2016).
391	23.	Williamson, D., Blaker, A. T. & Sinha, B. Tuning without over-tuning:
392		parametric uncertainty quantification for the NEMO ocean model. Geosci. Model
393		Dev. 10, 1789–1816 (2017). doi:10.5194/gmd-2016-185-AC3
394	24.	Gladstone, R. M. et al. Calibrated prediction of Pine Island Glacier retreat during
395		the 21st and 22nd centuries with a coupled flowline model. <i>Earth and Planetary</i>
396		Science Letters 333-334, 191–199 (2012).
397	25.	Shepherd, A. et al. Mass balance of the Antarctic Ice Sheet from 1992 to 2017.
398		Nature 558, 219–222 (2018).
399	26.	Golledge, N., E. D. Keller, N. Gomez, K. A. Naughten, J. Bernales, L. D. Trusel,
400		and T. L. Edwards (2018), Causes and consequences of 21st century ice sheet
401		melt. Submitted.
402	27.	Le Bars, D., Drijfhout, S. & de Vries, H. A high-end sea-level rise probabilistic
403		projection including rapid Antarctic ice sheet mass loss. Environ Res Lett 12,
404		044013 (2017).
405	28.	Golledge, N. R. et al. The multi-millennial Antarctic commitment to future sea-
406		level rise. Nature 526, 421-425 (2015).
407	29.	Cornford, S. L. et al. Century-scale simulations of the response of the West
408		Antarctic Ice Sheet to a warming climate. The Cryosphere 9, 1579–1600–
409		(2015).
410	30.	Wise, M. G., Dowdeswell, J. A., Jakobsson, M. & Larter, R. D. Evidence of
411		marine ice-cliff instability in PineIsland Bay from iceberg-keel plough marks.
412		<i>Nature</i> 550, 506–510 (2017).
413	31.	Bell, R. E. et al. Antarctic ice shelf potentially stabilized by export of meltwater
414		in surface river. Nature 544, 344–348 (2017).
415	32.	Kingslake, J., Ely, J. C., Das, I. & Bell, R. E. Widespread movement of
416		meltwater onto and across Antarctic ice shelves. Nature 544, 349-352 (2017).
417	33.	Gourmelen, N. et al. Channelized Melting Drives Thinning Under a Rapidly
418		Melting Antarctic Ice Shelf. Geophysical Research Letters 17, 173 (2017).
419	34.	Pattyn, F., Favier, L., Sun, S. & Durand, G. Progress in Numerical Modeling of
420		Antarctic Ice-Sheet Dynamics. Curr Clim Change Rep 3, 174–184 (2017).
421	35.	Golledge et al. (2017), Antarctic climate and ice-sheet configuration during the
422		early Pliocene interglacial at 4.23 Ma, <i>Climate of the Past</i> , 13(7), pp.959–975.
423	36.	Pukelsheim, F. The three sigma rule. The American Statistician 48, 88–91
424		(1994).
425	37.	Miller, K. G. et al. High tide of the warm Pliocene: Implications of global sea-
426		level for Antarctic deglaciation. Geology 40, 407–410 (2012).
427	38.	Raymo, M.E. et al., 2018. The accuracy of mid-Pliocene δ18O-based ice volume
428		and sea level reconstructions. <i>Earth-Science Reviews</i> , 177, pp.291–302.
429	39.	Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M.
430		A probabilistic assessment of sea-level variations within the last interglacial
431		stage. Geophysical Journal International 193, 711–716 (2013).
432	40.	Düsterhus, André, Tamisiea, M. E. & Jevrejeva, S. Estimating the sea-level
433		highstand during the Last Interglacial: a probabilistic massive ensemble

434		approach. Geophysical Journal International 206, 900–920 (2016).
435	41.	Austermann, J., Mitrovica, J. X., Huybers, P. & Rovere, A. Detection of a
436		dynamic topography signal in last interglacial sea-level records. Science
437		<i>Advances</i> 3, 1–9 (2017).
438	42.	Nias, I., Cornford, S. L. & Payne, A. J. Contrasting model sensitivity of the
439		Amundsen Sea embayment ice streams. Journal of Glaciology 62, 552-562
440		(2016).
441	43.	Holden, P. B., Edwards, N. R., Oliver, K. I. C., Lenton, T. M. & Wilkinson, R.
442		D. A probabilistic calibration of climate sensitivity and terrestrial carbon change
443		in GENIE-1. Climate Dynamics 35, 785–806 (2009).
444	44.	Edwards, N. R., Cameron, D. & Rougier, J. Precalibrating an intermediate
445		complexity climate model. Climate Dynamics 37, 1469–1482 (2011).
446	45.	Nowicki, S. M. J. et al. Ice Sheet Model Intercomparison Project (ISMIP6)
447		contribution to CMIP6. Geosci. Model Dev. Discuss. 1-42 (2016).
448		doi:10.5194/gmd-2016-105-SC2
449		-

450 Acknowledgements

T.L.E., N.E. and P.H. were supported by EPSRC REsearch on Changes of Variability and 451 Environmental Risk (ReCoVER: EP/M008495/1) under the Quantifying Uncertainty in 452 ANTarctic Ice Sheet instability (QUAntIS) project (RFFLP 006). T.L.E. was also supported 453 by the EPSRC-funded Past Earth Network (EP/M008363/1) and the Université Joseph 454 Fourier – Grenoble International visitor fund 'Campagne INVITES'. N.R.G is supported by 455 contract VUW1501 from the Royal Society Te Aparangi. I.N. was supported by the NERC 456 iSTAR-C project Dynamical control on the response of Pine Island Glacier (NE/J005738/1) 457 and now by the NASA Sea Level Change program. A.W. is supported by the Open 458 University Faculty of Science, Technology, Engineering and Mathematics. 459

Thanks to Rob DeConto for running additional simulations necessary for the analysis, 460 suggestions and helpful discussions, and to Kelsey Ruckert and Anders Levermann for 461 providing data. Thanks to the statisticans who attended the Past Earth Network events 462 'Assessing Palaeoclimate Uncertainty' (held jointly with the Environmental Statistics Section 463 of the Royal Statistics Society, Cambridge, August 2016), 'Emulators workshop' (Leeds, 464 June 2017), and writing retreat (Callow Hall, August 2017) for their advice, particularly Ian 465 Vernon, Peter Challenor, and Jonty Rougier. Thanks to Dewi Le Bars, Doug McNeall, David 466 Demeritt and the King's Geography Hazards, Risk and Regulation reading group (George 467 Adamson, Felicia Liu, Amar Razli, Laurence Ball and Ana Heilbron) for helpful comments 468 on the manuscript. Thanks to three anonymous reviewers for their comments and advice. 469 470 Thanks also to Dr Kai-Keen Shiu of the Gastrointestinal Oncology Unit at UCL Hospitals, and to Dallas Campbell without whom this work would not have been possible. 471

472

473 Author contributions

474	T.L.E. conceived the idea, carried out the analysis, produced the figures, and wrote the
475	manuscript. A.W. and P.B.H. performed preliminary analyses. A.J.P, A.W., C.R., G.D., I.N.,
476	M.B. and N.R.G. contributed ideas on glaciological and oceanic aspects, while A.W., N.R.E.
477	and P.B.H. contributed ideas on statistical aspects. All authors contributed to writing the
478	manuscript.
479	
480	Author information
481	The authors declare no competing financial interests. Correspondence and requests for
482	materials should be addressed to T.L.E. (tamsin.edwards@kcl.ac.uk).
483	
484	Table 1. Probabilities from DeConto and Pollard (2016) study. Means and standard deviations, and implied
485	probability intervals, for DeConto and Pollard (2016) ensemble at 2100 for RCP8.5 for their four
486	methodological choices (see text), using minimal assumptions about the distribution shape (finite mean and

487 variance: Chebyshev inequality).

RCP8.5	LOW PL	OCENE	HIGH PLIOCENE		
	Bias	Bias	Bias	Bias	
	Uncorrected	Corrected	Uncorrected	Corrected	
Antarctic contri	bution at 2100 (cm sea-level	equivalent)		
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]	

488

490 Table 2. Projections for the Antarctic contribution to sea-level in 2100. Calibrated with Pliocene, Last
491 Interglacial and satellite data (1997-2017), with and without DeConto and Pollard (2016) marine ice cliff
492 instability (MICI) parameterisation.

	RCI	RCP2.6		P4.5	RCP8.5					
	No MICI	MICI	No MICI	MICI	No MICI	MICI				
	Antarctic contribution at 2100 (cm sea-level equivalent)									
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									
\geq 30 cm		26%		68%	20%	88%				
\geq 50 cm		4%		46%		71%				
≥1 m				6%		36%				

- 496 Figure 1. Probabilistic projections of the Antarctic contribution to sea-level at 2100. Projections estimated
- under three RCPs, (a) with ice cliff instability parameterisation and (b) without, from emulation of the DeConto
- and Pollard (2016) ice sheet model ensemble. Dotted lines are the uncalibrated emulator ensemble; solid lines
- are calibrated with Last Interglacial and Pliocene reconstructions and satellite data from 1992-2017. Box and
- 500 whiskers show the [5, 25, 50, 75, 95]th percentiles; star shows the mode. The DeConto and Pollard (2016)
- 501 ensemble members for RCP8.5 (LowPliocene calibration; BiasCorrected and BiasUncorrected combined) are
- shown as a histogram and mean ± 2 s.d. interval in (a), scaled to the same height as the calibrated projection.
- 503 The projection for the Antarctic contribution due to ice discharge under the medium-high climate scenario A1B
- by Ritz et al. (2015) is also shown in (b). Data from refs. 6 and 25 and supplementary simulations by R.
- 505 DeConto (pers. comm.) (see Methods).
- 506

507 **Figure 2. Emergence of ice cliff instability.** Projected 5-95% probability intervals for Antarctic sea-level 508 contributions this century, with and without the marine ice cliff instability (MICI) parameterisation of DeConto 509 and Pollard (2016). Data from refs. 6 and 25 and supplementary simulations by R. DeConto (pers. comm.) (see 510 Methods).

- 511
- 512 Figure 3. Long-term projections of Antarctic sea-level contribution. (a) Shaded/hatched regions: projected
- 513 5-95% intervals for Antarctic sea-level contribution to 2300 and for 2500 with (shaded) and without (hatched)
- 514 ice cliff instability (MICI) parameterisation under three greenhouse gas concentration scenarios. Dots: mode of
- 515 the RCP4.5 and RCP8.5 distributions with MICI. Single lines: range of results from Golledge et al. (2015) under
- 516 RCP8.5 (solid dark red), RCP8.5 with doubled atmosphere and ocean temperature changes (dashed purple),
- 517 RCP4.5 and RCP2.6 (solid black). Box and whisker at 2200 shows Ritz et al. (2015): [5, 25, 50, 75, 95]th
- 518 percentiles and mode (*). (b). Projected probability of exceeding 1 m Antarctic sea-level contribution over the
- 519 same period. Data from refs. 6 and 25 and supplementary simulations by R. DeConto (pers. comm.) (see
- 520 Methods).
- 521

Figure 4. Multi-model comparison. Projections from this study (bold text: 'EMULATED') at 2100 based on 522 523 emulation of DeConto and Pollard (2016) (with and without ice cliff instability, 'MICI'), along with results 524 from other probabilistic modelling studies. Box and whiskers show the [5, 25, 50, 75, 95]th percentiles; star shows the mode. Numbers show the median, [5th, 95th] percentiles and, where available, the mode (*). "High 525 Scenarios" (pink/red) are for high-end (RCP8.5) or medium-high (Special Report on Emissions Scenario A1B¹) 526 527 greenhouse gas emissions or concentrations, or immediate collapse of part of West Antarctica (Little et al. (2013): 5th percentile and median are estimated from digitisation); Levermann et al. (2014) is from models with 528 529 ice shelves, without time delay. "Low Scenarios" (grey/black) are for low greenhouse gas concentrations (RCP2.6) or other baseline case (Little et al., 2013); Levermann et al. (2014) is with time delay. Levermann et 530 531 al. (2014) and Ritz et al. (2015) are for ice discharge contribution only. Data from refs. 2-6 and 25, supplementary simulations by R. DeConto (pers. comm.), and mode for ref. 5 supplied by K.L. Ruckert (pers. 532 533 comm) (see Methods).

534

536 **METHODS**

537 Simulator ensemble design

- 538 DeConto and Pollard (2016) perturb three continuous parameters, sampling four levels for 539 each in a factorial design to generate $4^3 = 64$ ensemble members:
- 540 **OCFAC**: Ocean melt factor, which controls sub-ice-shelf direct melting. Defined as a 541 factor by which the default value is multiplied. OCFAC = $\{0.1, 1, 3 \text{ and } 10\} \ge 0.224 \text{ m yr}^{-1}$ 542 °C⁻². (Note that DP16 quotes incorrect units of m yr⁻² °C⁻² in two places).
- 543 **CREVLIQ**: Crevasse liquid depth, which controls ice shelf collapse by hydrofracturing 544 due to surface liquid. Defined as the additional crevasse depth due to surface melt plus 545 rainfall rate. CREVLIQ = $\{0, 50, 100, 150\}$ m per (m yr⁻¹)⁻².
- 546 VCLIF: Maximum net ice wastage rate. Controls cliff failure after ice shelf collapse. 547 VCLIF = $\{0, 1, 3, 5\}$ km yr⁻¹.
- 548 For present day and future projections, this ensemble is duplicated with the ocean bias 549 correction applied. When emulating the ice sheet model (see below) we combine these 128 550 ensemble members and treat the bias correction as a continuous uncertain parameter:
- **BIAS:** Southern Ocean bias correction applied to present day and future simulations. Defined as a scalar ranging from 0 (no bias correction, $+0^{\circ}$ C) to 1 (full bias correction, $+3^{\circ}$ C). Active only for present day and future simulations.
- We use time series data for the ensemble provided by Rob DeConto. When emulating the model, we found a sign error in the DP16 Supplementary Information: the Last Interglacial value for simulation row 6 (OCFAC = 0.1, CREVLIQ = 50, VCLIF = 1) should be +2.63 m, not -2.63 m.
- 558

559 **Building the emulators**

We use Gaussian Process regression ('kriging' when used for spatial interpolation), because 560 it is flexible, non-parametric, and provides uncertainty estimates⁴⁶. As usual for emulation of 561 computer models, we set the 'nugget' to zero because the ice sheet model is deterministic. 562 We refer to 'the emulator' in the main text for simplicity, but this comprises separate 563 emulators for each scalar output: Pliocene and LIG sea-level change, present day (1992-2017 564 change in the RCP4.5 simulation) and the change from 2000 to every even-numbered year up 565 to 2500 for the three RCPs. We construct, validate, calibrate and make predictions using the 566 R software packages DiceKriging and a modified version of DiceEvaluation. 567

Let the function f(x) be the ice sheet model, which simulates sea-level change in a particular era (e.g. the Pliocene) as a function of the set of its input parameters, x. We 570 consider only one output at a time, to avoid the need for a further index. An emulator $f_{em}(x)$ 571 for a particular output of f(x) can be written as:

$$f_{em}(x) = \sum_j eta_j g_j(x) + u(x)$$

572

573 where g(x) are known functions of x, β are regression coefficients, and u(x) is a stochastic 574 process with a specified covariance function. We wish to select the subset of x that has the 575 most influence on $f_{em}(x)$.

Design and validation of the emulators comprises two parts: a step-wise model 576 selection procedure, to choose the mean function (i.e. which simulator parameters, and 577 interactions between these, to use as regressors), and a 'leave-one-out' (LOO) cross-578 validation procedure, to evaluate which is the most suitable covariance function and whether 579 580 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 581 choose the overall emulator structure. The final fitting of the emulators with the full ensemble 582 data, and their use for prediction, are discussed later. 583

584

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

- 592
- 593 Pliocene and Last Interglacial:

594 $g_{palaeo}(x) \sim (OCFAC, CREVLIQ, VCLIF, CREVLIQ*VCLIF)$

595

596 Present day and RCP2.6 at 2100:

597 $g_{low}(x) \sim (OCFAC, CREVLIQ, VCLIF, BIAS, OCFAC*VCLIF, OCFAC*BIAS,$

598 CREVLIQ*VCLIF, VCLIF*BIAS, OCFAC*VCLIF*BIAS)

599

600 RCP4.5 and RCP8.5 at 2100:

601 $g_{high}(x) \sim (OCFAC, CREVLIQ, VCLIF, BIAS, OCFAC*VCLIF, OCFAC*BIAS,$

602 CREVLIQ*VCLIF)

603

where $g \sim (a,b)$ means g is a linear function of a and b, etc, and a*b indicates an interaction term.

606

Covariance functions: The covariance controls the smoothness between data points, with a 607 trade-off between accuracy and over-fitting. We compare the success of different covariance 608 functions — Matern(5/2), Matern(3/2), exponential, and power-exponential (exponential 609 family, where the exponent can vary between 0 and 2) — using the mean function selected 610 above, and choose the one with the smallest normalised Euclidean distance in a LOO 611 procedure. The LOO procedure comprises fitting the emulator to all ensemble members 612 613 except one (i.e. 63 of 64 for Pliocene and LIG; 127 of 128 for present and future), and then predicting the final member to compare with the simulation itself. This is repeated for all 614 combinations ($N_{ens} = 64$ or 128) to provide a summary statistic. Normalised Euclidean 615 distance is: 616

$$d = \sqrt{\sum_{i=1}^{N_{ens}} \frac{(f_{em}(x_i) - f(x_i))^2}{\sigma_{em:x_i}^2}}$$

617

where *i* is the ensemble member and σ_{em} is the emulator error for that prediction. We choose 618 this metric because it makes use of the uncertainty estimate inherent in a Gaussian Process 619 emulator to standardise the residuals, so that an emulator with some large errors is not overly 620 penalised if it has sufficiently large uncertainty estimates to generally encompass the true 621 value. This also guards against overfitting, by penalising too-confident emulators. The 622 distance metric therefore balances the two aims of emulator accuracy and appropriate 623 confidence. The resulting covariance functions from this procedure are power-exponential for 624 the LIG, Matern(3/2) for 1992-2017, and exponential for the Pliocene and future outputs. 625

626

627 Validating and fitting the emulators

We use various validation outputs to assess emulator adequacy: RMSE; Kendall's tau, a non-628 parametric measure of correlation, for the emulator predictions versus the simulations; and 629 the fraction of predictions for which the simulation lies within the emulator 95% credibility 630 interval, for which values lower than ~90% would indicate an over-confident emulator (i.e. 631 too-small uncertainty estimates). The RMSE and Kendall rank correlation coefficients 632 between the emulator predictions and simulations are 12 cm (1.4% of the data range) and 633 0.958 respectively for the Pliocene; 26 cm (2.7%) and 0.923 for the Last Interglacial; 0.1 cm 634 (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 635

three future projection emulators, indicating sufficient accuracy. The fraction of predictions within the emulator 95% interval is 100% for the Pliocene, 89% for the LIG, and 91-98% for the present and future, indicating sufficiently large emulator uncertainty estimates. The predictive accuracy and uncertainty estimates of the six emulators can also be inspected visually by plotting the emulator predictions vs simulations and the standardised residuals (Extended Data Figure 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 timeslices
for that RCP.

645

646 Emulator ensemble design

We predict 10,000 points in the parameter space using a maximin Latin Hypercube (i.e. 647 648 efficiently space-filling) design. The MICI design samples from uniform distributions for all four parameters, based on discussion with one of the original DP16 authors (DeConto, pers. 649 comm.); the No-MICI design has VCLIF set to zero. The effect of VCLIF, CREVLIQ and 650 OCFAC on sea-level contribution at 2100 under RCP8.5 in the MICI case is shown in 651 Extended Data Figure 7, which shows the strong dependence on VCLIF. The reason for some 652 apparent gaps in emulator coverage is that the ensemble design is space-filling but does not 653 necessarily sample points in each corner of the parameter space, as the original ensemble 654 members do. 655

656

657 Pliocene calibration

The LowPliocene and HighPliocene projections of DP16 are presented (and have been 658 interpreted by others^{27,47}) as equally plausible, but here we make the case that the 659 HighPliocene calibration is not robust. This is important because the RCP8.5 projections are 660 uniquely sensitive to the particular minimum value chosen for the HighPliocene constraint 661 (10 m). Extended Data Figure 2a and b show that when the lower bound exceeds 9.6 m, this 662 results in much higher means and much smaller standard deviations, because fewer than a 663 quarter of the ensemble members pass. The sensitivity is caused by a combination of the 664 small ensemble size and the strong correlation in the model between Pliocene sea-level and 665 RCP8.5 projections (large circles in Extended Data Figure 3a). 666

667 This sensitivity to the Pliocene lower bound is exacerbated by the choice of calibration 668 method: a simple 'accept' or 'reject', which can be expressed in the 'history matching' 669 framework^{22,48} below. This binary filtering means we should choose a sufficiently wide range 670 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 values in the range 5–10 m and
15-20 m are simultaneously both plausible and implausible. The chosen data range should be
both broad and unique to obtain a calibration that is robust and meaningful.

Gasson, DeConto and Pollard $(2016)^{20}$ estimate the Antarctic contribution to mid-Pliocene sea-level has a maximum of 13 m, which would rule out most of the HighPliocene range, suggesting 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})

In fact, 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 Last Interglacial calibration (3.5-7.4 m) rejects these ensemble members: none of the ensemble members that pass the LIG constraint have Pliocene sea-level changes of less than 5 m (Extended Data Figure 4: no large circles directly below shaded box). The crucial judgement is therefore whether the 10 m HighPliocene lower bound can be justified.

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

689

690 Model discrepancy

Model 'discrepancy', or 'structural error', is defined as the smallest possible difference 691 between a model simulation and the true values: that is, how well the model could reproduce 692 reality at its best possible, 'tuned', parameter values^{4,21,22,23}. Discrepancy is an essential part 693 of model calibration: not incorporating it implies that a model could be tuned to perfectly 694 match reality. Using a value less than the observational error would imply we could simulate 695 reality better than we could measure it. Model discrepancy can, in some cases, be 696 approximately estimated by comparing simulations with multiple observations. But if there 697 are insufficient observations to do this, as is the case here, discrepancy can be viewed as a 698 tolerance to model error⁴⁸ estimated by expert judgement^{4,24} (see below). 699

700

701 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 notation by Vernon et al. $(2010)^{50}$ here. The relationship between a palaeodata reconstruction or an observation of sealevel change (Pliocene, LIG, or 1992-2017 trend), *z*, and the true value, *y*, is modelled as: 706

$$z = y + \epsilon_{obs}$$

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

0 (*)

$$y = f(x^*) + \epsilon_{md}$$

710 where x^* are the best values of the parameters, and ε_{md} is the model discrepancy with 711 variance σ_{md}^2 . We emulate f(x):

$$f(x) = f_{em}(x) + \epsilon_{em:x}$$

712713

where $f_{em}(x)$ is the mean emulator prediction for f(x), and $\varepsilon_{em:x}$ is the emulator error as before; it varies with *x*, and is automatically estimated in Gaussian Process emulation. For a given emulated output (Pliocene, LIG, 1992-2017 trend) we can use the standardised distance, also known as *implausibility*, *I*:

$$\mathcal{I}^2(x) = \frac{(f_{em}(x) - z)^2}{\sigma_{obs}^2 + \sigma_{em:x}^2 + \sigma_{md}^2}$$

718

to accept or reject a given emulated ensemble member with parameter values *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), i.e. that the parameter space outside the calibration intervals has a low probability of being plausible.

We use a minimum LIG palaeodata value of 3.5 m, rather than the 3.6 m quoted by DP16, for consistency with their calibrated ensemble results which include a member with LIG sea-level change 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 $\sigma_{Obs} = \{5, 2\}$ m for the Pliocene and LIG respectively. The observational constraint (Shepherd et al., 2018)²⁵ is the cumulative mass loss from 1992-2017, (2720 \pm 1390) Gt, converted to cm sea-level equivalent by dividing by 3600, to give (0.756 \pm 0.386) cm sea-level contribution over this period. Model discrepancy is set to $\sigma_{md} = 0.5$ cm for 1992-2017 sea-level change.

When calibrating with palaeodata, we accept ensemble members with I < 1 for the 733 Pliocene and LIG, so that the simplest case without emulator and model errors matches the 734 interval used by DP16. We note this Pliocene range approximately corresponds to a 95% 735 interval in some reconstructions, but the LIG range may correspond to a lower probability 736 than 95% by some estimates, and so may be too strict a constraint (see main text). Calibration 737 with satellite data observations accepts ensemble members with I < 3, to follow the usual 738 history matching convention for well-defined errors: for a smooth unimodal distribution, I <739 3 with probability greater than or equal to 95% (Pukelsheim, 1994)³⁶; for Gaussian 740 distributions, as we expect for the satellite data errors, the probability interval is 99.7%. 741

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 10,000member 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 these distributions. To improve the clarity of Figure 3, we exclude 1, 3 and 5 data points from each of RCP8.5, RCP4.5 and RCP2.6 MICI projections respectively, because the estimates are not continuous in time (due to slight differences in emulator fitting).

754

755 Multi-model comparisons

We show distributions from Ruckert et al. $(2016)^5$ provided by Kelsey Ruckert, and estimate the distribution at 2100 for Little et al. $(2013)^2$ by digitisation of the original figures. We reestimate the modes for Ritz et al. $(2015)^4$ distributions using an automatic bandwidth for the kernel density estimation, rather than the broader, fixed bandwidth used in the original study. We assume differences due to definitions of time period are small enough to be ignored: all are 2000-2100, except Little et al. $(2013)^2$ (1990-2099) and the IPCC¹ (1986–2005 to 2081– 2100 for Antarctic component).

763

764 Palaeodata uncertainties

765 We here consider probability intervals for palaeodata constraints. Peak total sea-level change

- for the LIG estimated by Kopp et al. $(2013)^{39}$ is 6.4-10.9 m (90% probability interval), and by
- 767 Dusterhus et al. (2016)⁴⁰ is 6.1-16.7m (80% probability). These broadly encompass recent

- assessments that the upper end of the widely used 6-9 m range⁴⁹ could increase by several metres⁴¹. Subtracting a range of estimates for the contributions from Greenland, thermal expansion and glaciers in the same way as Ruckert et al. (2016)⁵ (3.4-4.8 m) gives Antarctic contributions of 1.6-7.5 m and 1.3-13.3 m respectively.
- For the Pliocene, Miller et al. $(2012)^{37}$ estimate 22 ± 10 m (95% range) total sea-level 772 change; subtracting 7 m for the Greenland ice sheet and 1 m for thermal expansion (Golledge 773 et al., 2017)³⁵ would imply approximately 14 ± 10 m Antarctic contribution, i.e. 4-24 m. 774 There is no difference between using a combined 5-25 m range and using the LowPliocene 775 (5-15 m) constraint presented here, because the DP16 ensemble maximum is 12.4 m, though 776 for a different model or ensemble design the upper bound might have more influence. 777 Golledge et al. $(2017)^{35}$ estimate 8.6 \pm 2.8 m for the Antarctic contribution to the early 778 Pliocene, and we use their Gaussian assumption to derive the 95% (2σ) range. 779
- 780

781 Code availability

- All emulation was performed in R using the DiceKriging and DiceEvaluation packages with
- minor modifications by TLE. The scripts and input data for the main analysis (sea-level
- 784 projections at 2100) are available as a downloadable R package on GitHub
- 785 (https://github.com/tamsinedwards/revisitmici, v1.0.2) and can be run without installation on
- the cloud-based computational reproducibility platform Code Ocean at
- 787 <u>https://doi.org/10.24433/CO.4ebd8cda-35c0-4d8f-9b7c-d1b064109437</u>.
- 788

789 Data availability

- All projections from this study are available on request. Simulations of the LIG and Pliocene,
 1992-2017 mean and 2100 sea level change for all DP16 ensemble members are available in
 the Code Ocean data folder at the above link. Simulations at 2500 for the subset of the DP16
 ensemble passing their calibration are available also in the Supplementary Materials of
 DeConto and Pollard (2016).
- 795

796 Methods References

798	46.	O'Hagan, A. Bayesian analysis of computer code outputs: A tutorial. Reliability
799		Engineering and System Safety 91, 1290–1300 (2006).
800	47.	Sweet, W. V. et al. GLOBAL AND REGIONAL SEA-LEVEL RISE
801		SCENARIOS FOR THE UNITED STATES. 1–75 (2017).
802	48.	Williamson, D. et al. History matching for exploring and reducing climate model
803		parameter space using observations and a large perturbed physics ensemble.
804		<i>Climate Dynamics</i> 41, 1703–1729 (2013).

805	49.	Dutton, A. et al. Sea-level rise due to polar ice-sheet mass loss during past warm
806		periods. Science 349, aaa4019–aaa4019 (2015).
807	50.	Vernon, I., Goldstein, M. & Bower, R. G. Galaxy Formation: a Bayesian
808		Uncertainty Analysis. Bayesian Analysis 5, 619–669 (2010).

809 Extended Data

810

811 Extended Data Figure 1. Sensitivity of DeConto and Pollard RCP8.5 projections to Pliocene data lower

bound. DeConto and Pollard (2016) BiasUncorrected (a) and BiasCorrected (b) projections for Antarctic sealevel contribution by 2100 under RCP8.5 as a function of the lower bound of the Pliocene data range. Mean ± 1 s.d. range shown as central solid line with pink shading; mean ± 2 s.d. range as dotted line. (c) Sensitivity of

- 815 emulated projections for RCP8.5 at 2100 with MICI from this study: [5, 25, 50, 75, 95]th percentiles and mode
 816 (*). Data from ref. 6.
 - 817

818 Extended Data Figure 2. DeConto and Pollard RCP8.5 projection distributions. DeConto and Pollard 819 (2016) ensemble projections for Antarctic sea-level contribution by 2100 under RCP8.5 for their four variants, 820 LowPliocene BiasUncorrected (a) and BiasCorrected (b), and High Pliocene similarly (c, d), showing the full 821 64-member ensemble and the subset selected by calibrating with Pliocene and Last Interglacial sea-level 822 reconstructions. The mean ± 1 s.d. range of the ensemble is shown as a solid red line with pink shading, and the 823 68% or greater probability interval is shown as a horizontal black line (see main text and Methods for more 824 details). Data from ref. 6.

825

Extended Data Figure 3. Relationships between RCP8.5 projections at 2100 and past sea-level changes.
Sea-level contribution at 2100 under RCP8.5 versus (a) Pliocene sea-level change; (b) Last Interglacial sea-level
change; (c) sea-level change from 1992-2017, for the emulator (small grey dots) and DP16 simulator (large open
circles) with ocean bias correction off (blue) and on (red). Grey shading indicates the DP16 palaeodata range (a,
b) or observational mean ± 3 s.d. (c); the dashed line additionally includes model error. Data from refs. 6 and 25
and supplementary simulations by R. DeConto (pers. comm.) (see Methods).

832

Extended Data Figure 4. Relationship between past and future sea-level changes with and without MICI. Simulator ensemble (large circles), and emulated ensembles (small circles) with (a) cliff instability and (b) no cliff instability, showing Pliocene versus Last Interglacial sea-level changes and shaded by sea-level contribution at 2100 under RCP8.5. Large emulator points and filled simulator points pass the 1992-2017 calibration. Shaded rectangle indicates bounds of DP16 LowPliocene and Last Interglacial palaeodata constraints; dashed box shows constraints in this study, i.e. including model error. Data from refs. 6 and 25 and supplementary simulations by R. DeConto (pers. comm.) (see Methods).

840

Extended Data Figure 5. Sensitivity of RCP8.5 projections to MICI and calibration choices. Projections for RCP8.5 at 2100, with and without MICI, for different combinations of calibration eras ('palaeo': Pliocene and Last Interglacial; present: 1992-2017) and model discrepancy (with, without, and double). Box and whiskers show the [5, 25, 50, 75, 95]th percentiles; star shows the mode. Numbers show the median, [5th, 95th] percentiles and mode (*). Data from refs. 6 and 25 and supplementary simulations by R. DeConto (pers. comm.) (see Methods).

847

Extended Data Figure 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-
- 851 2017 contribution; Last Interglacial; and Pliocene. Vertical error bars show 95% credibility intervals. Right
- 852 column: Difference between emulator predictions and simulations, standardised by emulator error, for the same
- six outputs. Values falling mostly between ± 2 indicate the emulator has adequate uncertainty estimates. Data
- from ref. 6 and supplementary simulations by R. DeConto (pers. comm.) (see Methods).
- 855

857

856 Extended Data Figure 7. Sensitivity of RCP8.5 projections to model parameters. Sea-level contribution at

2100 under RCP8.5 versus VCLIF (a), CREVLIQ (b) and OCFAC (c) parameters for emulator (small grey dots

- 858 with error bars) and simulator (large open circles: BiasUncorrected blue, BiasCorrected red). Data from ref. 6
- and supplementary simulations by R. DeConto (pers. comm.) (see Methods).









Sea level equivalent at 2100 (cm)





Frequency

Frequency



Sea level equivalent for RCP8.5 at 2100 (cm)

Sea level equivalent for RCP8.5 at 2100 (cm)



Sea level equivalent for Pliocene (m)



SLE for RCP8.5



Sea level equivalent for 1992-2017 (cm)



Sea



Sea level equivalent for Last Interglacial (m)



N -	+			1					1					
	-1		0	1	2	3	4	5	6	7	8	9	10	11
Sea level equivalent for Last Interalacial (m)														









Palaeo & present; no discrepancy

Palaeo; no discrepancy

Palaeo & present; VCLIF $\leq 4 \text{ km/a}$

No-MICI:

Palaeo & present

Palaeo only

Present only

Palaeo & present; double discrepancy

Palaeo & present; no discrepancy

Palaeo; no discrepancy









Standardised residuals





VCLIF

6



