Supporting Information for "Could the Last Interglacial Constrain Projections of Future Antarctic Ice Mass Loss and Sea-level Rise?"

Daniel M. Gilford[1,2], Erica L. Ashe[2], Robert M. DeConto[3], Robert E. Kopp[1,2], David Pollard[4], Alessio Rovere[5]

[1]Institute of Earth, Ocean, and Atmospheric Sciences, Rutgers University, 71 Dudley Road, Suite 205, New Brunswick, NJ 08901, USA.

[2]Department of Earth and Planetary Sciences, Rutgers University, Piscataway, NJ, USA.

[3]Department of Geosciences, University of Massachusetts, Amherst, MA, USA.

[4]Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA.

[5]MARUM, Center for Marine Environmental Sciences, University of Bremen, Germany.

Contents of this file

Introduction

This supporting information provides underlying details on the ice-sheet model ensemble, emulator construction, validation, and sensitivity tests, as well as supplemental figures of timeseries color-coded by CREVLIQ, comparisons between the Last Interglacial and RCP8.5 ensembles across the model parameter space, conditional posterior distributions in 2150, and ice-sheet model parameter likelihoods as a function of LIG constraint distribution. We note that the ice-sheet model ensembles are constructed with a model version updated since DeConto and Pollard (2016), but predating that of DeConto et al. (2020). As such, the results herein are not representative of the most current results with the latest physical model, but are illustrative of how ice-sheet models may be combined with statistical/machine learning methods and paleoclimate evidence to (a) constrain projections of future Antarctic ice-sheet contributions to sea-level rise, and to (b) identify how improved paleo sea level estimates could inform projections. A glossary of key study terms is included at the end, for reference.

Table S1. Ice-sheet model parameter values used to construct a 14×14 grid composing 196 members for the Last Interglacial and RCP8.5 scenario ensembles.

CLIFVMAX ($\frac{\text{km}}{\text{yr}}$)	CREVLIQ ($\frac{\text{m}}{(\text{myr}^{-1})^2}$)
0	0
1	15
2	30
3	45
4	60
5	75
6	90
7	105
8	120
9	135
10	150
11	165
12	180
13	195

Table S2. Optimized hyperparameters of the GP models (Eqn. 1–3) found by maximizing the log-likelihoods, given the training ensembles.

Ensemble	α_1^2 (m^2)	ℓ_1^2	α_2^2 (m^2)	ℓ_2^2	τ (yr)
LIG	17.048	45.698	—	—	—
RCP8.5	2731.8	2.7567	1.830	0.50121	95.52198

Text S1. Emulator Leave-one-out Analyses

To assess whether the Gaussian process (GP) model emulator accurately mimics the ice-sheet simulator, we perform a leave-one-out (LOO) analysis following a modification of the methodology of Bastos and O'Hagan (2009). We calculate the individual standardized prediction errors as,

$$D_j^I = \frac{z_j - E[f(\theta_\mathbf{1}, \theta_\mathbf{2})_j | \mathbf{z}_{\backslash \mathbf{j}}]}{\sqrt{V[f(\theta_\mathbf{1}, \theta_\mathbf{2})_j | \mathbf{z}_{\backslash \mathbf{j}}]}} \tag{S1}$$

where $\mathbf{z_{\backslash j}}$ is the vector composed all of training ensemble members in $\mathbf{z}$ except z_j at the jth location in model parameter space (i.e. the value with a fixed [θ_1,θ_2] from Table S1, removed for the LOO process), and $E[\cdot]$ and $V[\cdot]$ are the expectation (mean function) and variance, respectively, of the optimized emulator conditioned on $\mathbf{z_{\backslash j}}$. For RCP8.5, f and $\mathbf{z_{RCP}}$ are a function of time, and hence D_j^I is also time-dependent. The LIG emulator, $\mathbf{z_{LIG}}$, and the LIG standardized prediction errors have no time dependency. Errors are shown for the LIG in Figure S3 and the RCP8.5 scenario (in 2000, 2050, 2100, and 2150) in Figure S4.

Standardized errors are expected to follow a standard Student-t distribution. Errors which consistently exceed ±2 (the 95% credibility interval) indicate a conflict between the emulator and simulator (Bastos and O'Hagan 2009). We find that the LIG and RCP8.5 emulators performs well, with nearly all errors falling within the confidence interval. Emulator skill degrades slightly over the time in the RCP8.5 scenario as the training data sea-levels disperse when instabilities drive mass loss (section 3.2), creating less densely packed training information in time and parameter space. 5/196, about 2.5%, of the errors exceed +2 in 2150. These poorly performing emulator estimates are located near the exterior of parameter space, where θ_1 and θ_2 are high, and there is less surrounding training information to constrain the emulator prediction (behavior which is typical of trained Gaussian process models, Rasmussen and Williams 2006).

Across time and both training ensembles, standardized emulator errors are less than ±2 in over 99% of points tested. One concern might be whether these errors indicate emulator variances are too large relative to the mean (i.e. whether the model is underconfident), driving low values of D_j^I. The RCP8.5 emulator very accurately predicts relatively small (near-zero) and broadly similar changes in mass loss across the whole parameter space; this contributes to the model's excellent standardized agreement through 2050 (Fig. S4, top panels). As the distribution of

ice-sheet mass begins to diverge around 2060 (Fig. 3) and emulator skill drops marginally (as discussed above), the model evolves toward an error distribution more consistent with the expected standard Student-t distribution. Overall, the time-independent variance of the RCP8.5 emulator is always <0.0004 m^2 across the model parameter space, such that the model standard deviation is always <2 cm.

The LIG emulator variance is plotted in Figure S7; values span over 0–0.016 m^2 across the ice-sheet model parameter space. The associated GP model standard deviation is 11 cm on average, $\sim$3% of the range of the LIG emulator output. The model may therefore be slightly underconfident, which could affect our study results/conclusions in two ways. First, a model with too high variance would result in less confidence in model parameters given a specific LIG constraint (i.e. less polarized likelihoods, Fig. S5), so that the LIG is less informative for the MICI process. Second, higher variance results in a broader emulated distribution than may be warranted. However, one of the strengths of the Bayesian approach (section 2.3) is the ability both include and quantify the uncertainty of the emulator (as in Fig. S7), so some variance in the final model is justified. Ultimately, the final model described in the main text captures the key behavior of the training data, and had the smallest variances of any model explored (cf. Text S2).

Overall, performance is consistent with that of another recently published ice-sheet model emulator (Edwards et al. 2019, their Extended Data Figure 6), which was trained on a different version of the same ice-sheet model (e.g., Pollard and DeConto 2012). We conclude the emulator is able to accurately predict simulator responses across the LIG and RCP8.5 scenarios with appropriate uncertainties.

Text S2. GP Model Selection and Sensitivity to Covariance Function

There are infinitely many possible model forms, specifications, optimization targets, etc. to consider for an emulator (Rasmussen and Williams 2006). The "final model" (described in the main text) represents the best model based on several metrics: model simplicity, likelihood maximization, and minimizing of prediction errors (described above in Text S2) and model uncertainty (i.e. posterior variance).

We assessed different covariance forms: squared-exponential functions (sometimes called the radial basis function), nonstationary linear (sometimes called dot-product) functions, and Matérn functions with shape parameters (ν) of $\frac{1}{2}$, $\frac{3}{2}$, and $\frac{5}{2}$. We also evaluated various combinations of these functions, and experimented with specifying them along the individual axes of the training data (θ_1, θ_2, and time). For instance, we considered the complex form,

$K(\theta_1, \theta_2) \sim Linear(\theta_1) * Matérn_{\nu=\frac{5}{2}}(\theta_1) + Linear(\theta_2) * Matérn_{\nu=\frac{5}{2}}(\theta_2) + Matérn_{\nu=\frac{5}{2}}(\theta_1, \theta_2)$.

Trained models produced log-likelihoods similar to (or sometimes even higher than) the final model. But when optimized, each of these models required a variance (fit uncertainty) larger than the final model (Fig. S7) in order to match the training data (cf. Text S2). Under such circumstances, the optimized model is underconfident, and a nugget of 10^{-6} m is a strong requirement that is inconsistent with the optimized model variance. We present one such model as an example below and discuss the implications.

To demonstrate the emulator sensitivity to the choice of covariance function, we specify an alternative set of covariance functions, f_1^* and f_2^*, which replace f_1 and f_2 in Eqn. (1):

$$f_1^*(\theta_1, \theta_2) \sim \mathcal{GP}(0, \alpha_1^2 K_{1,\theta_1}(\theta_1, \theta_1'; \ell_{1,\theta_1}) \cdot K_{1,\theta_2}(\theta_2, \theta_2'; \ell_{1,\theta_2})), \tag{S2}$$

$$f_2^*(\theta_1, \theta_2, t) \sim \mathcal{GP}(0, \alpha_2^2 K_{2,\theta_1}(\theta_1, \theta_1'; \ell_{2,\theta_1}) \cdot K_{2,\theta_2}(\theta_2, \theta_2'; \ell_{2,\theta_2}) \cdot K_t(t, t'; \tau)), \tag{S3}$$

where there are four distinct covariance functions, $K_{i,\theta}$, each with a unique and trainable length scale specified along either CLIFVMAX (θ_1) or CREVLIQ (θ_2), $\ell_{i,\theta}$. Because the model form is different, the hyperparameters which share an interpretation with Eqn. (1)—α_i and τ—need not have the same optimized values as those of the final model (Table S2). Following the procedure described in sections 2.2–2.3, this alternative model is optimized and conditioned on the training simulations, and its posterior distributions are found conditional on LIG constraints.

The unconstrained Last Interglacial distribution of this alternative model form is presented in Figure S8 alongside the final model distribution (reproduced from Fig. 4a) and the LIG training ensemble histogram. The alternative model distribution is broader than that of the final model distribution, driven by a larger variance. The LIG alternative model's average standard deviation is 25 cm, more than twice that of the final model, which smooths out some of the multi-modal features of the unconstrained LIG distribution. The training ensemble exhibits a multi-modal distribution (more consistent with the final model), suggesting the alternative model contributes less information about AIS mass loss from individual sectors than indicated by the original ice-sheet model simulations.

Likewise, the alternative model of the RCP8.5 emulator has greater uncertainty, with a time-constant standard deviation of $\sim$5 cm and a width of the 95% credibility interval between 2000 and 2060 of 20 cm (a period where the full range of simulated mass loss is 0–7.7 cm). Given these increased uncertainties, emulated behavior such as the instability-driven skew in Fig. 7b (given a relatively low LIG constraint) disappears, suggesting the alternative model is less physical. The alternative model posterior distributions of RCP8.5 AIS mass loss as a function of LIG constraints are shown in Figure S9. Comparing with Fig. 4b, posterior distributions have substantially broader projections if the LIG was known precisely (to within 10 cm). This degrades

the informative power of LIG constraints on the margins of the LIG distribution (i.e. high or low values, Fig. 7b), because the baseline uncertainty more than doubles. Hence, the final model more accurately captures the multi-modal behavior of the LIG training ensemble and is more precise in its predictions.

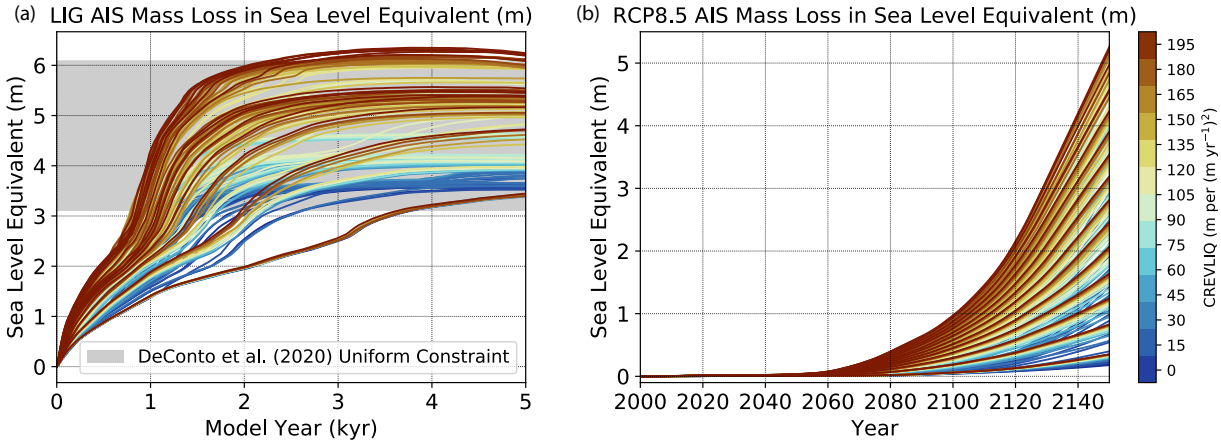

Figure S1. As in Fig. 1, except timeseries are color-coded by their CREVLIQ values over 0–195 $\frac{\text{m}}{(\text{myr}^{-1})^2}$.

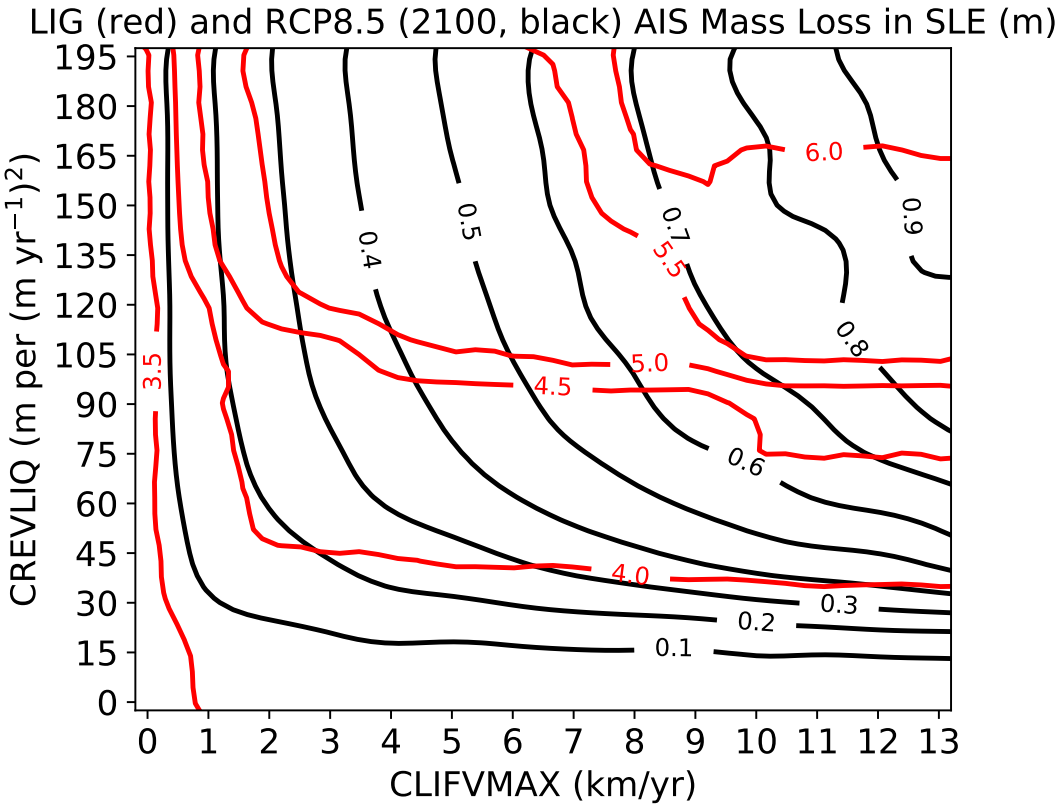

Figure S2. Contours are identical to the mean emulated sea-level contributions from the Antarctic ice sheet in Fig. 2, but with LIG and RCP8.5 contours overlapping for comparison.

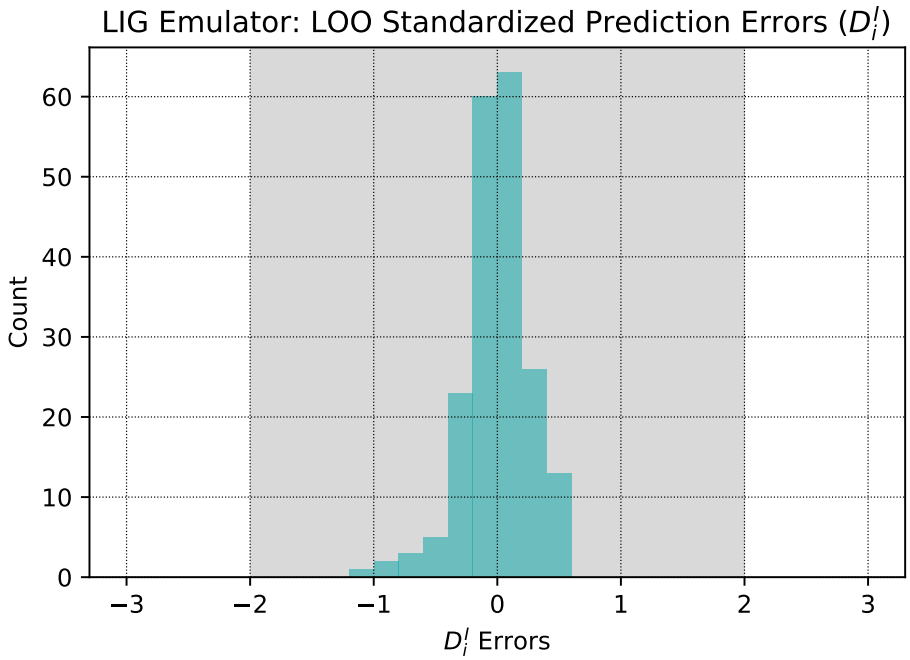

Figure S3. Histogram of standardized prediction errors (Eqn. S1) from leave-one out analyses performed with the Last Interglacial emulator. Errors $< \pm 2$ (gray shaded region) indicate the emulator is able to properly represent the ice-sheet model (cf. Bastos and O'Hagan, 2009).

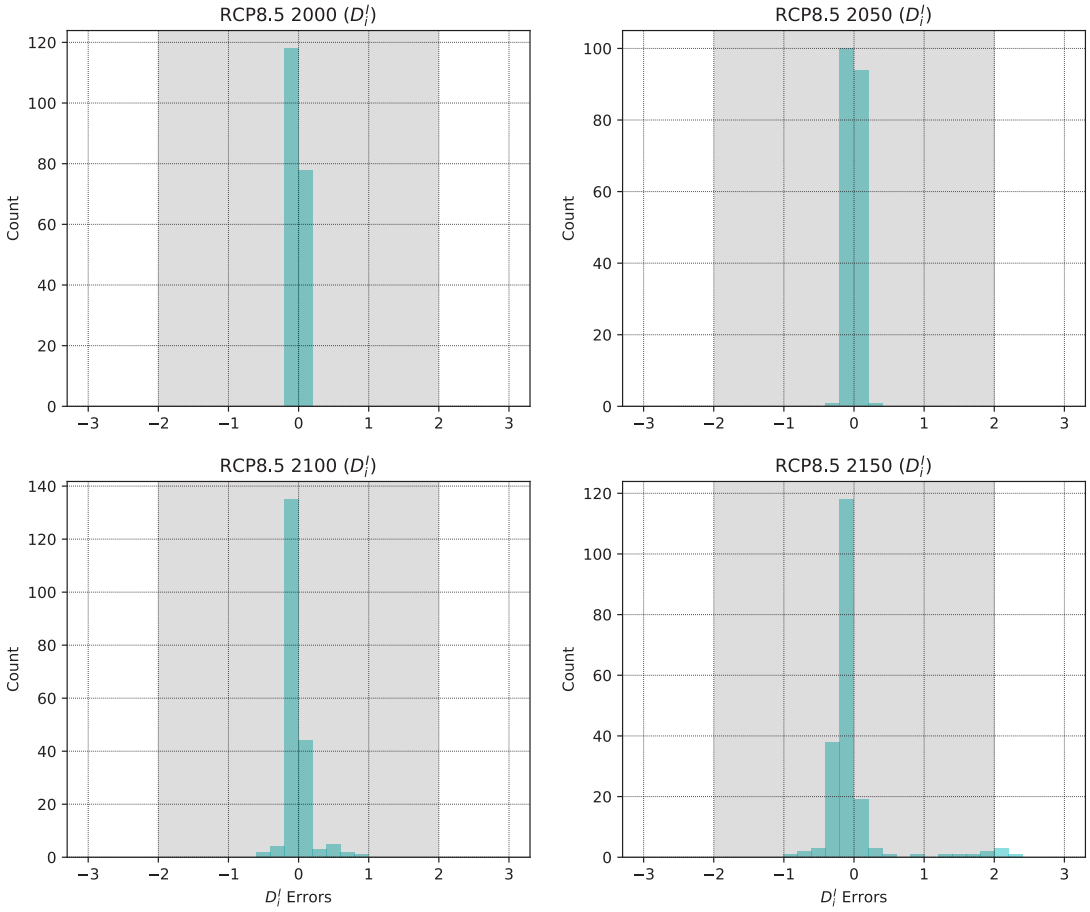

Figure S4. As in Fig. S3, but for the RCP8.5 emulator in 2000, 2050, 2100, and 2150.

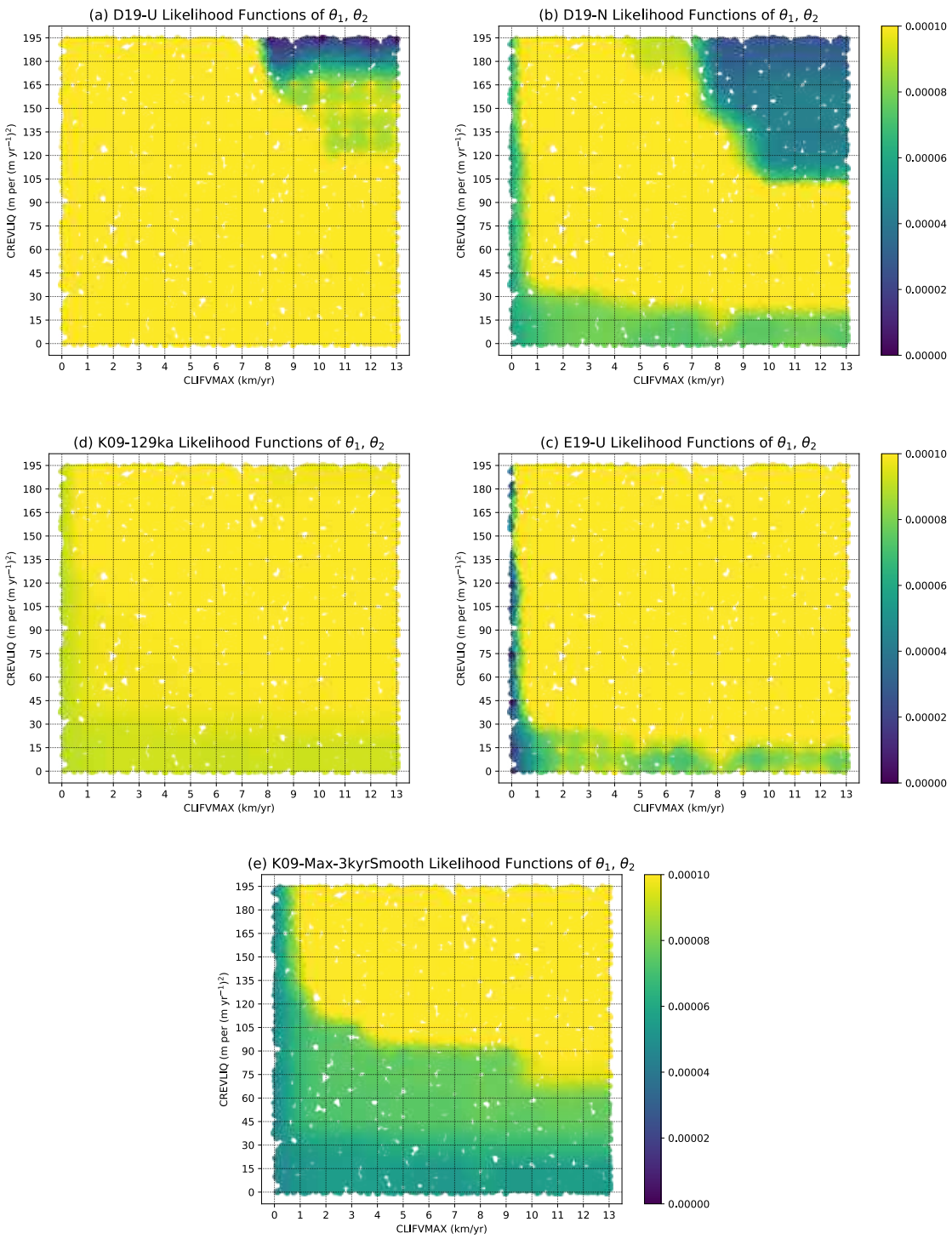

Figure S5. The posterior probabilities of CREVLIQ/CLIFVMAX latin-hypercube sampled pairs across the range of the model ensemble parameter space (cf. Table S1), conditional on specified constraints on Last Interglacial Antarctic Ice-sheet sea level contributions (cf. Figure 4a). The colorbar saturates at it upper extent.

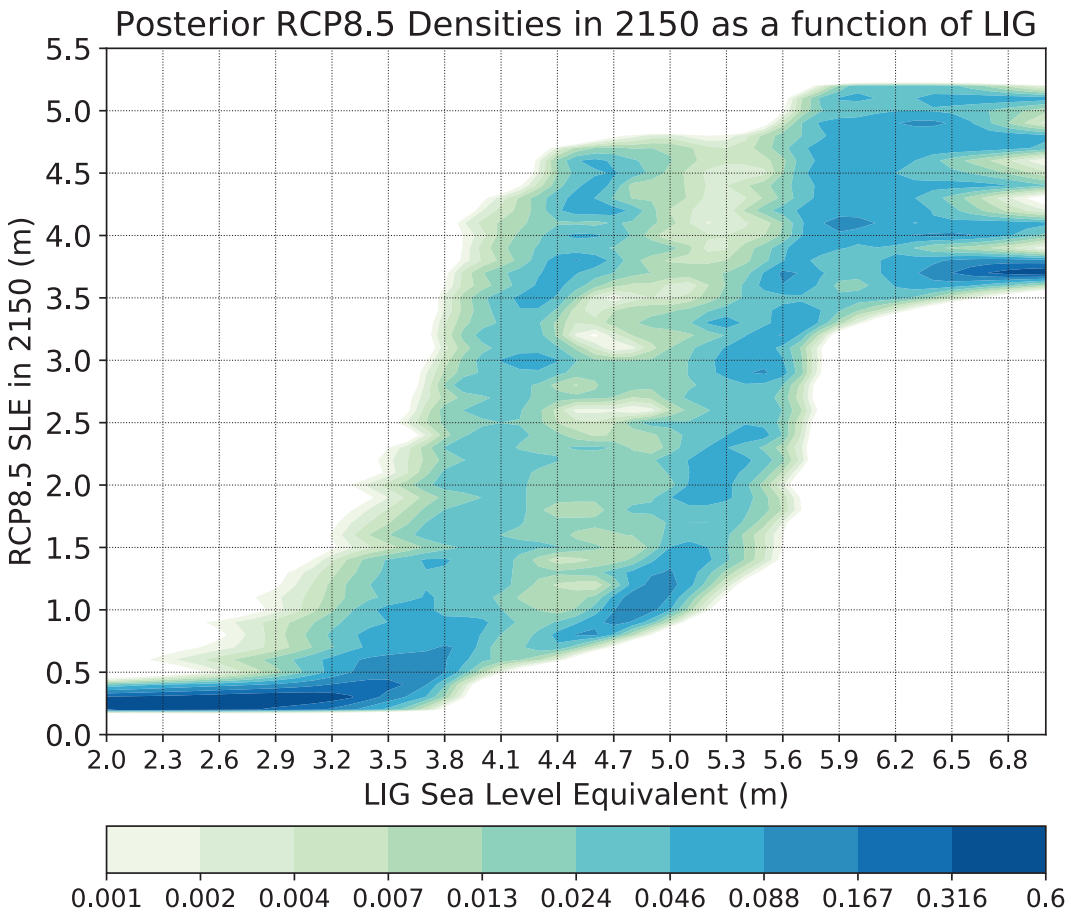

Figure S6. As in Fig. 4b, except for 2150.

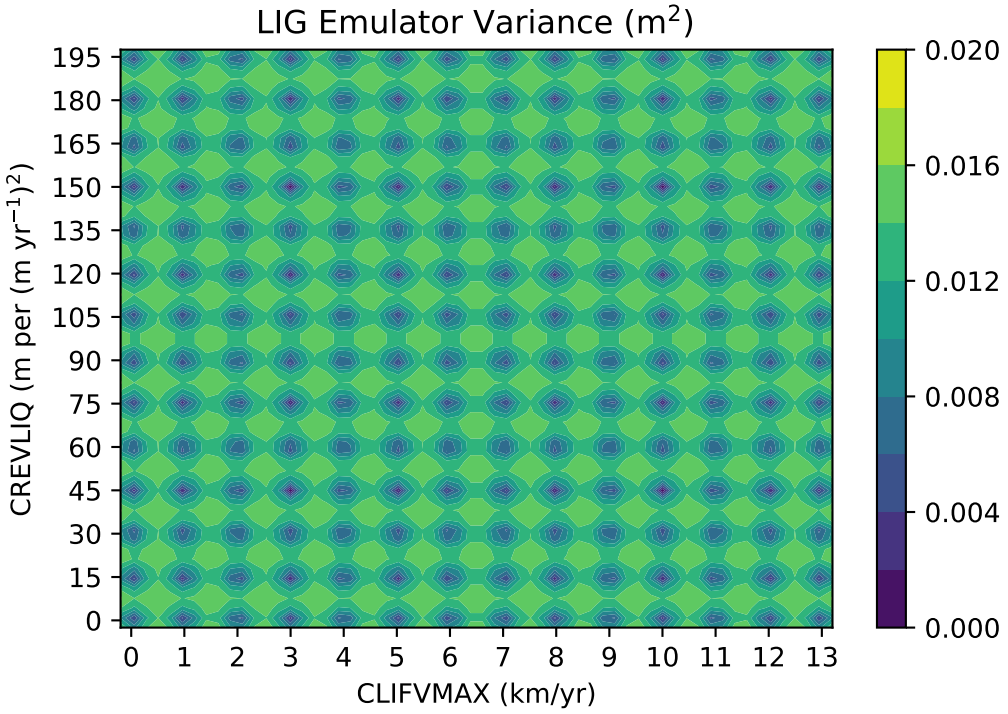

Figure S7. Last Interglacial emulator variance (m^2) over the ice-sheet model parameter space.

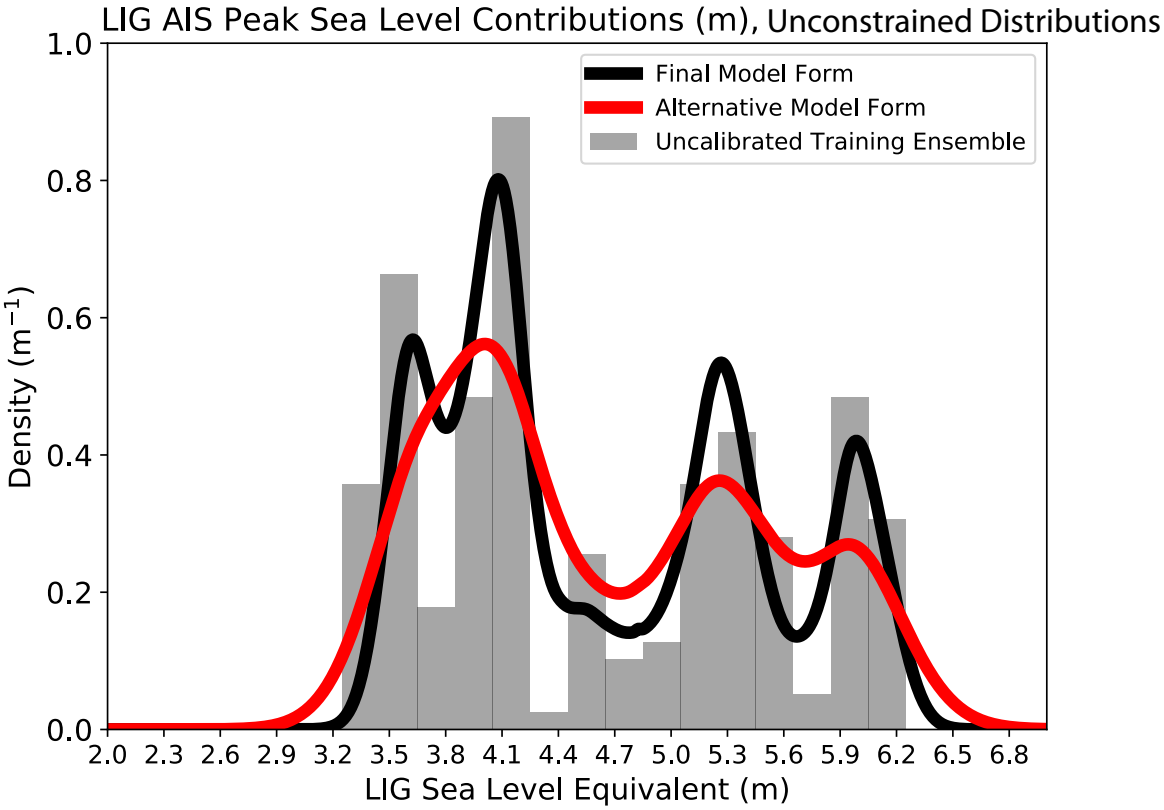

Figure S8. Unconstrained Last Interglacial emulated distribution reproduced from Fig. 4a (black curve), compared with the emulated distribution from an alternative model (red curve) defined with the covariance functions given in Eqn. (S2–S3). The training ensemble is shown as a histogram scaled for comparison.

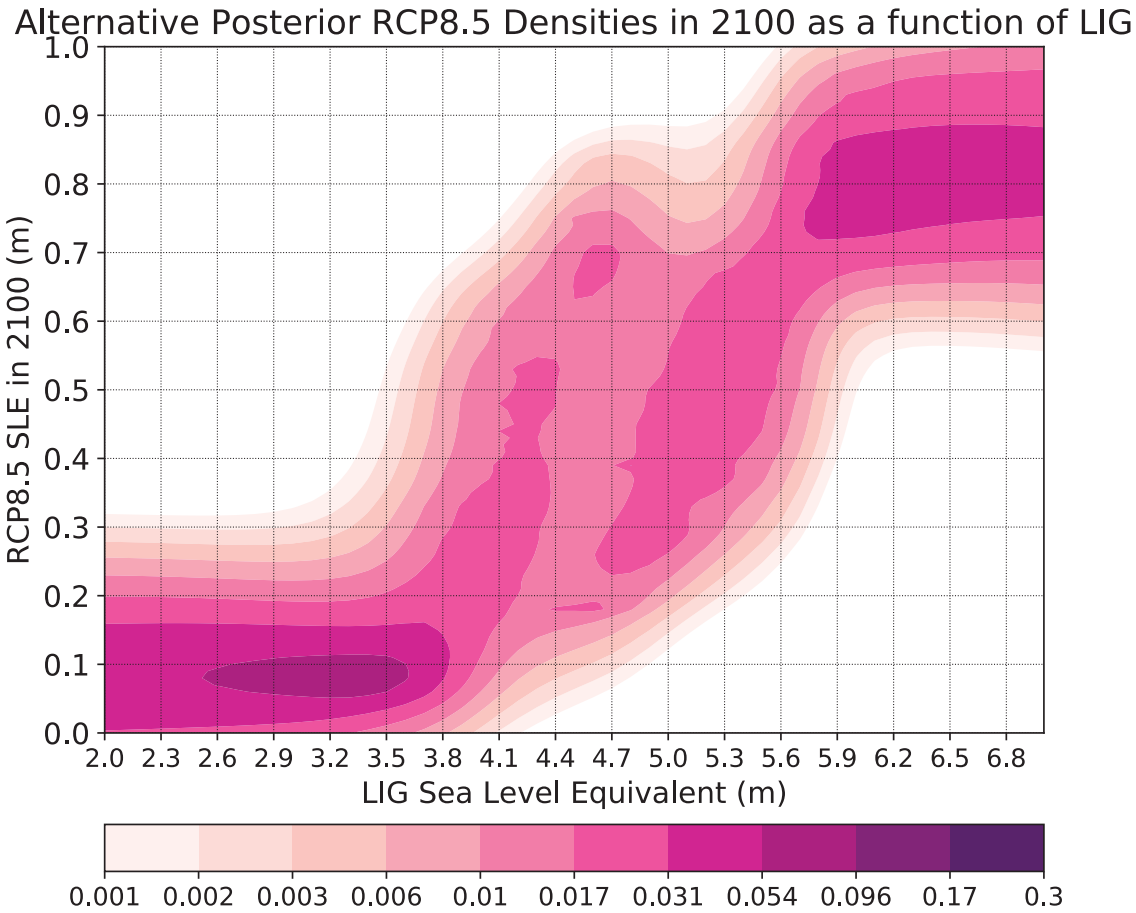

Figure S9. As in Fig. 4b, except normalized conditional posterior probability densities are plotted as a function of Last Interglacial AIS mass loss emulated with an alternative model defined with the covariance functions given in Eqn. (S2–S3).

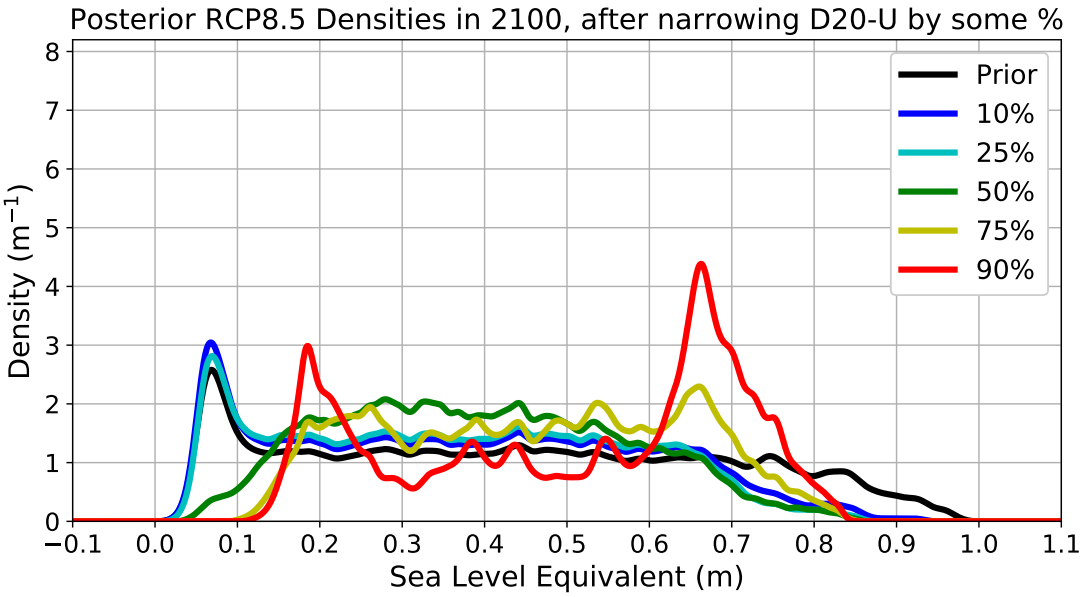

Figure S10. As in Fig. 6, except posteriors are constrained assuming the D20-U constraint was 10%, 25%, 50%, 75% or 90% narrower (blue, cyan, green, yellow and red curves, respectively).

Table S3. Glossary—Definitions of relevant terms.

Term	Meaning
Bayesian statistics	(in contrast to frequentist or classical statistical inference) is a theory based on the Bayesian interpretation of probability where probability expresses a degree of belief in an event. Bayesian methods compute a posterior probability of a model or parameter through the use of a prior probability distribution of the model or parameter times a likelihood function using Bayes' theorem
Bayesian updating	the process of using new information to improve on previous estimates. One uses the posterior distribution of one model as the likelihood of a new model. For example, the posterior distribution on the ice-sheet model parameters, $(\theta_1, \theta_2 \mid f_{LIG})$, is used as the likelihood distribution in the future projection model
Conditional probability	the distribution of a random quantity, given (assuming, or as a function of) a particular value of another random quantity
Covariance function	defines prior beliefs about the relationship between one or more variables or parameters in a Gaussian process, as a measure of how much they change together
Gaussian process (GP)	a generalization of the multi-variate Gaussian distribution to continuous parameter space, which is fully defined by its mean function and covariance function; GP regression provides an analytically-tractable solution when adopting the assumption of normality for all distributions
Hyperparameter	parameter of a GP model prior distribution
Likelihood	the probability of observing the data as described by the fitted model; also known as the sampling or data distribution; a conditional distribution that is a function of unknown parameters for observed data
Marginal distribution	unconditional probability distribution of a random quantity, found by integrating over all values of the conditional distribution

Table S3 ***(continued).***

Term	Meaning
Non-parametric	not involving any assumptions as to the functional form
Posterior probability	the probability distribution of an unknown quantity, conditional on (or assuming/given) observed data; In this study these are, 1) the future AIS sea-level contribution projections over time conditioned on a specified Last Interglacial estimate distribution, and 2) the distribution of the model parameters (CREVLIQ and CLIFVMAX) given specific LIG constraints
Prior probability	(of an uncertain quantity—e.g., parameter or model) uses a priori beliefs about the quantity before some evidence or data is taken into account; the prior is combined with the probability distribution of new data to yield a posterior distribution. The prior can be subjective or uninformative (such as a uniform distribution) to minimize the impact on Bayesian statistics