Large ensemble modeling of last deglacial retreat of the West Antarctic Ice Sheet: Comparison of simple and advanced statistical techniques

David Pollard[1], Won Chang[2], Murali Haran[3], Patrick Applegate[1,4] and Robert DeConto[5]

[1] Earth and Environmental Systems Institute, Pennsylvania State University, University Park

[2] Department of Statistics, University of Chicago, Illinois

[3] Department of Statistics, Pennsylvania State University, University Park

[4] Earth Sciences Program, Pennsylvania State University, DuBois

[5] Department of Geosciences, University of Massachusetts, Amherst

Correspondence to: D. Pollard (pollard@essc.psu.edu)

Abstract

A 3-D hybrid ice-sheet model is applied to the last deglacial retreat of the West Antarctic Ice Sheet over the last ~20,000 years. A large ensemble of 625 model runs is used to calibrate the model to modern and geologic data, including reconstructed grounding lines, relative sea-level records, elevation-age data and uplift rates, with an aggregate score computed for each run that measures overall model-data misfit. Two types of statistical methods are used to analyze the large-ensemble results: simple averaging weighted by the aggregate score, and more advanced Bayesian techniques involving Gaussian process-based emulation and calibration, and Markov chain Monte Carlo. The analyses provide sea-level-rise envelopes with well defined parametric uncertainty bounds, but the simple averaging method only provides robust results with full-factorial parameter sampling in the large ensemble. Results for best-fit parameter ranges and envelopes of equivalent sea-level rise with the simple averaging method agree well with the more advanced techniques. Best-fit parameter ranges confirm earlier values expected from prior model tuning, including large basal sliding coefficients on modern ocean beds.

1. Introduction

Modeling studies of future variability of the Antarctic Ice Sheet have focused to date on the Amundsen Sea Embayment (ASE) sector of West Antarctica, including the Pine Island and Thwaites Glacier basins. These basins are currently undergoing rapid thinning and acceleration, producing the largest Antarctic contribution to sea level rise (Shepherd et al., 2012; Rignot et al., 2014). The main cause is thought to be increasing oceanic melt below their floating ice shelves, which reduces back pressure on the grounded inland ice (buttressing; Pritchard et al., 2012; Dutrieux et al., 2014). There is a danger of much more drastic grounding-line retreat and sea-level rise in the future, because bed elevations in the Pine Island and Thwaites Glacier basin interiors deepen to depths of a kilometer or more below sea level, potentially allowing Marine Ice Sheet Instability (MISI) due to the strong dependence of ice flux on grounding-line depth (Weertman, 1974; Mercer, 1978; Schoof, 2007; Vaughan, 2008; Rignot et al., 2014; Joughin et al., 2014).

Recent studies have mostly used high-resolution models and/or relatively detailed treatments of ice dynamics (higher order or full Stokes dynamical equations; Morlighem et al., 2010; Gladstone et al., 2012; Cornford et al., 2013; Parizek et al., 2013; Docquier et al., 2014; Favier et al., 2014; Joughin et al., 2014). Because of this dynamical and topographic detail, models with two horizontal dimensions have been confined spatially to limited regions of the ASE and temporally to durations on the order of centuries to one millennium. On the one hand, these types of models are desirable because highly resolved bed topography and accurate ice dynamics near the modern grounding line could well be important on timescales of the next few decades to century (references above, and Durand et al., 2011; Favier et al., 2012). On the other hand, the computational run-time demands of these models limit their applicability to small domains and short time scales, and they can only be calibrated against the modern observed state and decadal trends at most.

Here we take an alternate approach, using a relatively coarse-grid ice sheet model with hybrid dynamics. This allows run durations of several 10,000 years, so that model parameters can be calibrated against geologic data of major retreat across the continental shelf since the Last Glacial Maximum (LGM) over the last ~20,000 years. The approach is a trade-off between (i) less model resolution and dynamical fidelity, which degrades future predictions on the scale of ~10's km and the next few decades (sill-to-sill retreat immediately upstream from modern grounding lines), and (ii) more confidence on larger scales of 100's km and 1000's years (deeper into the interior basins, further into the future) provided by calibration versus LGM extents and deglacial retreat of the past 20,000 years. Also the approach allows more thorough exploration of uncertain parameter ranges and their interactions, such as sliding coefficients on modern ocean beds, oceanic melting strengths, and different Earth treatments of bedrock deformation.

A substantial body of geologic data is available for the last deglacial retreat in the ASE and other Antarctic sectors. Notably this includes recent reconstructions of grounding-line locations over the last 25 kyrs by the RAISED Consortium (RAISED, 2014). Other types of data at specific sites include relative sea-level records, cosmogenic elevation-age data, and modern uplift rates (compiled in RAISED, 2014; Briggs and Tarasov, 2013, Briggs et al., 2013, 2014; Whitehouse et al., 2012a,b). Following several recent Antarctic modeling studies (Briggs et al. and Whitehouse et al. as above; Golledge et al., 2014; Maris et al., 2015), we utilize these datasets in conjunction with large ensembles (LE), i.e., sets of hundreds of simulations over the last deglacial period with systematic variations of selected model parameters. LE studies have also been performed for past variations of the Greenland Ice Sheet, for instance by Applegate et al. (2012) and Stone et al. (2013).

This paper follows on from Chang et al. (2015, 2016), who apply relatively advanced Bayesian statistical techniques to LE's generated by our ice-sheet model. The statistical steps are described in detail in Chang et al. (2015, 2016), and include:

- Statistical emulators, used to interpolate results in parameter space, constructed using a new emulation technique based on principal components.

- Probability models, replacing raw square-error model-data misfits with formal likelihood functions, using a new approach for binary spatial data such as grounding-line maps.
- Markov Chain Monte Carlo (MCMC) methods, used to produce posterior distributions which are continuous probability density functions of parameter estimates and projected results based on formally combining the information from the above two steps in a Bayesian inferential framework.

Some of these techniques were applied to LE modeling for Greenland in Chang et al. (2014). McNeall et al. (2013) used a Gaussian process emulator, and scoring similar to our simple method, in their study of observational constraints for a Greenland ice sheet model ensemble. Tarasov et al. (2012) used Artificial Neural Nets in their LE calibration study of North American ice sheets, and have mentioned their potential application to Antarctica (Briggs and Tarasov, 2013). Apart from these examples, to our knowledge the statistical techniques in Chang et al. (2015, 2016) are considerably more advanced than the simpler averaging method used in most previous LE ice-sheet studies; these previous studies have involved

(i) Computing a single objective score for each LE member that measures the misfit between the model simulation and geologic and modern data, and

(ii) Calculating parameter ranges and envelopes of model results by straightforward averaging over all LE members, weighted by the scores.

The more advanced statistical techniques offer substantial advantages over the simple averaging method, such as providing robust and smooth probability density functions in parameter space. For instance, Applegate et al. (2012) and Chang et al. (2014) show that the simple averaging method fails to provide reasonable results for LE's with coarsely spaced Latin HyperCube sampling, whereas for the same LE, the advanced techniques successfully interpolate in parameter space and provide smooth and meaningful probability densities.

However, the advanced techniques in Chang et al. (2015, 2016) require statistical expertise not readily available to most ice-sheet modeling groups. It may be that the simple averaging method still gives reasonable results, especially for LE's with full factorial sampling, i.e., with

every possible combination of selected parameter values (also referred to as grid or Cartesian product; Urban and Fricker, 2010). The purpose of this paper is to apply both the advanced statistical and simple averaging methods to the same Antarctic LE, compare the results, and thus assess whether the simple (and commonly used) method is a viable alternative to the more advanced techniques, at least for full factorial LEs. The results include probabilistic ranges of model parameter values, and envelopes of model results such as equivalent sea-level rise. Further types of results related to specific glaciological problems (LGM ice volume, MeltWater Pulse 1A, future retreat) will be presented in Pollard et al. (2016) using the simple-averaging method, and do not modify or extend the comparisons of the two methods in this paper.

Sections 2.1 and 2.2 describe the model, the setup for the last deglacial simulations, and the model parameters chosen for the full factorial LE. Sections 2.3 to 2.4 describe the objective scoring vs. past and modern data used in the simple averaging method, and Sect. 2.5 provides an overview of the advanced statistical techniques. Results are shown for best-fit model parameter ranges and equivalent sea-level envelopes in Sects. 3 and 4, comparing simple and advanced techniques. Conclusions and steps for further work are described in Sect. 5.

2. Methods

2.1. Ice sheet model and simulations

The 3-D ice-sheet model has previously been applied to past Antarctic variations in Pollard and DeConto (2009), DeConto et al. (2012) and Pollard et al. (2015). The model predicts ice thickness and temperature distributions, evolving due to slow deformation under its own weight, and to mass addition and removal (precipitation, basal melt and runoff, oceanic melt, and calving of floating ice). Floating ice shelves and grounding-line migration are included. It uses hybrid ice dynamics and an internal condition on ice velocity at the grounding line (Schoof, 2007). The simplified dynamics (compared to full Stokes or higher-order) captures grounding-line migration reasonably well (Pattyn et al., 2013), while still allowing O(10,000's) year runs to be feasible. As in many long-term ice sheet models, bedrock deformation is

modeled as an elastic lithospheric plate above local isostatic relaxation. Details of the model formulation are described in Pollard and DeConto (2012a,b). The drastic ice-retreat mechanisms of hydrofracturing and ice-cliff failure proposed in Pollard et al. (2015) are only triggered in warmer-than-present climates and so do not play any role in the glacial-deglacial simulations here; in fact they are switched off in all runs. Tests show that they play no perceptible role in simulations over the last 40 kyears.

The model is applied to a limited area nested domain spanning all of West Antarctica, with a 20-km grid resolution. Lateral boundary conditions on ice thicknesses and velocities are provided by a previous continental-scale run. The model is run over the last 30,000 years, initialized appropriately at 30 ka (30,000 years before present, relative to 1950 AD) from a previous longer-term run. Atmospheric forcing is computed using a modern climatological Antarctic dataset (ALBMAP: Le Brocq, 2010), with uniform cooling perturbations proportional to a deep-sea core $\delta^{18}O$ record (as in Pollard and DeConto, 2009, 2012a). Oceanic forcing uses using archived ocean temperatures from a global climate model simulation of the last 22 kyr (Liu et al., 2009). Sea level variations versus time, which are controlled predominantly by Northern Hemispheric ice sheet variations, are prescribed from the ICE-5G dataset (Peltier, 2004). Modern bedrock elevations are obtained from the Bedmap2 dataset (Fretwell et al., 2013).

2.2. Large ensemble and model parameters

The large ensemble analyzed in this study uses full factorial sampling, i.e., a run for every possible combination of parameter values, with 4 parameters varied and with each parameter taking 5 values, requiring 625 ($=5^4$) runs. As discussed above, results are analyzed in two ways: (1) using the relatively advanced statistical techniques (emulators, likelihood functions, MCMC) in Chang et al. (2015, 2016), and (2) using the much simpler averaging method of calculating an aggregate score for each run that measures model-data misfit, and computing results as averages over all runs weighted by their score. Because the second method has no means of interpolating results between sparsely separated points in multi-dimensional parameter space, it is effectively limited to full factorial sampling with only a few parameters

and a small number of values each. The small number of values is nevertheless sufficient to span the full reasonable "prior" range for each parameter, and although the results are very coarse with the second method, they show the basic patterns adequately. Furthermore, envelopes of results of all model runs are compared in Appendix D with corresponding data, and demonstrate that the ensemble results do adequately "span" the data, i.e., there are no serious outliers of data far from the range of model results.

The 4 parameters and their 5 values are:

OCFAC: Sub-ice oceanic melt coefficient.
Values are 0.1, 0.3, 1, 3, 10 (non-dimensional).
Corresponds to K in Eq. 17 of Pollard and Deconto (2012a).

CALV: Factor in calving of icebergs at oceanic edge of floating ice shelves.
Values are 0.3, 0.7, 1, 1.3, 1.7 (non-dimensional).
Multiplies combined crevasse-depth-to-ice-thickness ratio r in Eq. B7 of Pollard et al. (2015).

CSHELF: Basal sliding coefficient for ice grounded on modern-ocean beds.
Values are10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} (m yr^{-1} Pa^{-2}).
Corresponds to C in Eq. 11 of Pollard and Deconto (2012a).

TAUAST: e-folding time of bedrock relaxation towards isostatic equilibrium.
Values are 1, 2, 3, 5, 7 kyrs.
Corresponds to τ in Eq. 33 of Pollard and Deconto (2012a).

The 4 parameters were chosen based on prior experience with the model; each has a strong effect on the results, yet their values are particularly uncertain. The first 3 involve oceanic processes or properties of modern ocean-bed areas. Parameters whose effects are limited to modern grounded-ice areas are reasonably well constrained by earlier work, such as basal sliding coefficients under modern grounded ice which are obtained by inverse methods (e.g., Pollard and DeConto, 2012b, for this model). More discussion of the physics and uncertainties associated with these parameters is given in Appendix A.

2.3. Individual data types and scoring

Following Whitehouse (2012a,b), Briggs and Tarasov (2013) and Briggs et al. (2013, 2014), we test the model against 3 types of data for the modern observed state, and 5 types of geologic

data relevant to ice-sheet variations of the last ~20,000 years, using straightforward mean squared or root-mean-square misfits in most cases. Each misfit (M_i, i = 1 to 8) is normalized into an individual score (S_i), which are then combined into one aggregate score (S) for each member of the LE. Only data within the domain of the model (West Antarctica) is used in the calculation of the misfits.

One approach to calculating misfits and scores is to borrow from Gaussian error distribution concepts, i.e., individual misfits M of the form $[(mod\text{-}obs) / \sigma]^2$ and overall scores of the form $e^{-\underline{M}/s}$, where *mod* is a model quantity, *obs* is a corresponding observation, σ is an observational or scaling uncertainty, $\underline{M}$ is an average of individual misfits over data sites and types of measurements, and s is another scaling value (Briggs and Tarasov, 2013; Briggs et al., 2014). However, the choice of these forms is somewhat heuristic, and different choices are also appropriate for complex model-data comparisons with widespread data points, very different types of data, and with many model-data error types not being strictly Gaussian. In order to determine the influence of these choices on the results, we compare two approaches: (a) with formulae adhering closely to Gaussian forms throughout, and (b) with some non-Gaussian aspects attempting to provide more straightforward and interpretable scalings between different data types. Both approaches are described fully below (next section, and Appendix B). They yield very similar results, with no significant differences between the two, as shown in Appendix C. The second more heuristic approach (b) is used for results in the main paper.

The 8 individual data types and model-data misfits are listed below, with basic information that applies to both of the above approaches. More details are given in Appendix B, including formulae for the two approaches, and "intra-data-type weighting" that is important for closely spaced sites (Briggs and Tarasov, 2013). The two approaches of combining the individual scores into one aggregate score S for the simple averaging method are described in the following Sect. 2.4. The more advanced statistical techniques (Chang et al., 2015, 2016) use elements of these calculations but differ fundamentally in some aspects, as outlined in Sect. 2.5.

The 8 individual data types are:

1. TOTE: Modern grounding-line locations.
Misfit M_1: based on total area of model-data mismatch for grounded ice.
Data: Bedmap2 (Fretwell et al., 2013).

2. TOTI: Modern floating ice-shelf locations.
Misfit M_2: based on total area of model-data mismatch for floating ice.
Data: Bedmap2 (Fretwell et al., 2013).

3. TOTDH: Modern grounded ice thicknesses.
Misfit M_3: based on model-data differences of grounded ice thicknesses.
Data: Bedmap2 (Fretwell et al., 2013).

4. TROUGH: Past grounding-line distance vs. time along the centerline trough of Pine Island Glacier. Centerline data for the Ross and Weddell basins can also be used, but not in this study.
Misfit M_4 : based on model-data differences over the period 20 to 0 ka.
Data: RAISED (2014) (Anderson et al., 2014, for the Ross; Hillenbrand et al., 2014, for the Weddell; Larter et al., 2014, for the Amundsen Sea).

5. GL2D: Past grounding-line locations (see Fig. 1). Only the Amundsen Sea region is used in this study.
Misfit M_5: based on model-data mismatches for 20, 15, 10, 5 ka.
Data: RAISED (2014) (Anderson et al., 2014; Hillenbrand et al., 2014; Larter et al., 2014; Mackintosh et al., 2014; O Cofaigh et al., 2014).
.

6. RSL: Past Relative Sea Level (RSL) records.
Misfit M_6 : based on χ-squared measure of model-data differences at individual sites.
Data: compilation in Briggs and Tarasov (2013).

7. ELEV/DSURF: Past cosmogenic elevation vs. age (ELEV) and thickness vs. age (DSURF).
Misfits M_{7a}, M_{7b}: based on model-data differences at individual sites, combined as in Appendix B.
Data: compilations in Briggs and Tarasov (2013) for ELEV; in RAISED (2014) with individual citations as above for DSURF.

8. UPL: Modern uplift rates on rock outcrops.
Misfit M_8: based on model-data difference at individual sites.
Data: compilation in Whitehouse et al. (2012b).

2.4. Combination into one aggregate score for simple averaging method

Each of the misfits above are first transformed into a normalized individual score for each data type i = 1 to 8. The transformations differ for the two approaches mentioned above.

(a) For approach (a), closely following Gaussian error forms, using misfits M_i as described in Appendix B:

- For a given data type i, the misfits M_i for all runs (1 to 625) are sorted, and normalized using the median value M_i^{50}, i.e., $M_i' = M_i \,/\, M_i^{50}$. This is somewhat analogous to the heuristic scaling for overall scores in Briggs et al., (2014, their sec. 2.3), but applied here to individual types.
- The individual score S_i for data type i and each run is set to $e^{-Mi'}$
- The aggregate score for each run is $S = S_1\, S_2\, S_3\, S_4\, S_5\, S_6\, S_7\, S_8$, i.e., $e^{-\Sigma Mi'}$

Of the two approaches, this most closely follows Briggs and Tarasov (2013) and Briggs et al. (2014), except for their inter-data-type weighting, which assigns very different weights to the individual types based on spatial and temporal volumes of influence (Briggs and Tarasov, 2013, their sec. 4.3.2; Briggs et al., 2014, their sec. 2.2). Here, we assume that each data type is of equal importance to the overall score, and that if any one individual score is very bad ($S_i \approx 0$), the overall score S should also be ≈ 0. This corresponds to the notion that if any single data type is completely mismatched, the run should be rejected as unrealistic, regardless of the fit to the other data types. The fits to past data, even if more uncertain and sparser than modern, seem equally important to the goal of obtaining the best calibration for future applications with very large departures from modern conditions.

(b) For the more heuristic approach (b), using misfits Mi as described in Appendix B:

- For a given data type i, a "cutoff" value C_i is set by taking the geometric mean (i.e., square root of the product) of (i) the minimum (best) misfits M_i over all runs 1 to 625, and (ii) the algebraic average of the 10 largest (worst) values. The 10 worst values are used to avoid a single outlier that could be unbounded; the single best value is used because it is bounded by zero, and is not an outlier but represents close-to-the-best possible simulation with the current model. The geometric mean and not the algebraic mean of these two numbers is more appropriate if the values range over many orders of magnitude.
- The normalized misfit M_i' for data type i and each run is set to M_i/C_i . We implicitly assume that M_i' values close to 0 ($M_i << C_i$) represent very good simulations of this data type, close to the best possible within the LE. M_i' values ≥ 1 ($M_i \geq C_i$) represent very

poor simulations, diverging from this data type so much that the run should be rejected no matter what the other scores are.

- The individual score S_i for data type i and each run is set to max [0, 1 - M_i']
- The aggregate score for each run is $S = (S_1 S_2 S_3 S_4 S_5 S_6 S_7 S_8)^{1/8}$

In both approaches, the formulae apply equal weights to the individual data types, and do not use "inter-data-type" weighting (Briggs and Tarasov 2013; Briggs et al. 2014). As in (a), if any individual score S_i is ≈ 0, then the overall score S is ≈ 0, and the discussion above also applies to approach (b). Both approaches have loose links to the calibration technique in Gladstone et al. (2012) and the more formal treatment in McNeall et al. (2013).

2.5. Advanced statistical techniques

The more advanced statistical techniques (Chang et al., 2015, 2016) consist of an emulation and a calibration stage, involving the same 4 model parameters and the 625-member LE as above. The aggregate scores S described in Sect. 2.4 are not used at all. The techniques are outlined here; full accounts are given in Chang et al. (2015, 2016).

Emulation phase:

Emulation is the statistical approach by which a computer model is approximated by a statistical model. This statistical approximation is obtained by running the model at many parameter settings and then "fitting" a Gaussian process model to the input-output combinations, analogous to fitting a regression model that relates independent variables (parameters) to dependent variables (model output) in order to make predictions of the dependent variable at new values of the independent variables. Of course, unlike basic regression, the model output may itself be multivariate. An emulator is useful because: (i) it provides a computationally inexpensive method for approximating the output of a computer model at any parameter setting without having to actually run the model each time, and (ii) it provides a statistical model relating parameter values to computer model output – this means the approximations automatically include uncertainties, with larger uncertainties at parameter

settings that are far from parameter values where the computer model has already been run. Specifically, the model output consisting of (i) modern grounding line maps, and (ii) past locations of grounding lines versus time along the centerline trough of Pine Island, are first reduced in dimensionality by computing Principal Components (PCs) over all LE runs. (Principal components are often referred to in the atmospheric science literature as empirical orthogonal functions or EOFs.) The first 10 PCs are used for modern maps, and the first 20 for past trough locations. Hence, we develop a Gaussian process emulator for each of the above PCs. Gaussian process emulators work especially well for model outputs that are scalars. The emulators are "fitted" to the PCs using a maximum likelihood estimation-based approach developed in Chang et al. (2015) that addresses the complications that arise due to the fact that the data are non-Gaussian. Details are available in (Chang et al., 2015, 2016). The emulators provide a statistical model that essentially replaces the data types TOTE, TROUGH and GL2D described in Sect. 2.3.

In an extension to Chang et al. (2016), Gaussian process emulators are also used here to estimate distributions of individual score values for the 5 data types TOTI, TOTDH, RSL, ELEV/DSURF and UPL (S_2, S_3, S_6, S_7, S_8, approach (b), Sect. 2.3 and Appendix B), one emulator per score. Again, emulators are developed for each of the scores by using the Gaussian process machinery and maximum likelihood estimation.

Calibration phase:

The calibration stage solves the following problem in a statistically rigorous fashion: given observations and model runs at various parameter settings, which parameters of the model are most likely? In a Bayesian inferential framework, this translates to learning about the posterior probability distribution of the parameter values given all the available computer model runs and observations. The approach may be sketched out as follows. The emulation phase provides a statistical model connecting the parameters to the model output. Suppose it is assumed that the model at a particular (ideal) set of parameter values produces output that resembles the observations of the process. We also allow for measurement error and systematic discrepancies between the computer model and the real physical system. We do this via a "discrepancy

function" that simultaneously accounts for both; this is reasonable because both sources of error are important while also being difficult to tease apart. Hence, one can think of our approach as assuming that the observations are modeled as the model output at an ideal parameter setting, added to a discrepancy function. Once we are able to specify a model in this fashion, Bayesian inference provides a a very standard approach to obtain the resulting posterior distribution of the parameters: we start with a prior distribution for the parameters, where we assume that all the values are equally likely before any observations are obtained, and then use Bayes theorem to find the posterior distribution given the data. The posterior distribution cannot be found in analytical form. Hence, in this second "calibration" stage, posterior densities of the model parameters are inferred via Markov Chain Monte Carlo (MCMC). The observation and model quantities used in emulation and calibration consist of the modern and past grounding-line locations, and 5 individual scores. The discrepancy function is accounted for in assessing model vs. observed grounding-line fits in our Bayesian approach. It is based in part on the locations and times in which grounded ice occurs in the model and not in the observations, or vice versa, in 50% or more of the LE runs (Chang et al., 2015, 2016). For the individual scores, we use exponential marginal densities, whose rate parameters receive gamma priors scaled in such a way that the 90^{th} percentile of the prior density coincides with each score's cutoff value C_i in Sect. 2.4.b.

In the above procedures, observational error enters for the individual scores RSL, ELEV/DSURF and UPL via the calculations described in Appendix B. It is implicitly taken into account by the discrepancy function for grounding-line locations. Observational error is considered to be negligible for modern TOTI and TOTDH scores.

3. Results: Aggregate scores with simple averaging method

Fig. 2 shows the aggregate scores S for all 625 members of the LE, over the 4-dimensional space of the parameters CSHELF, TAUAST, OCFAC and CALV. Each individual subpanel shows TAUAST versus CSHELF, and the subpanels are arranged left-to-right for varying CALV, and bottom-to-top for varying OCFAC.

3.1. "Outer" variations, CALV and OCFAC

All scores with the largest CALV value of 1.7 (right-hand column of subpanels) are 0. In these runs, excessive calving results in very little floating ice shelves and far too much grounding line-retreat. Conversely, with the smallest CALV value of 0.3 (left-hand column of subpanels) , most runs have too much floating ice and too advanced grounding lines during the runs, so most of this column also has zero scores. However, small CALV can be partially compensated by large OCFAC (strong ocean melting), so there are some non-zero scores in the upper-left subpanels.

3.2. "Inner" variations, CSHELF and TAUAST

For mid-range CALV and OCFAC (subpanels near the center of the figure), the best scores require high CSHELF (inner x axis) values, i.e., slippery ocean-bed coefficients of 10^{-6} to 10^{-5} m a^{-1} Pa^{-2}.This is the most prominent signal in Fig. 2, and is consistent with the widespread extent of deformable sediments on continental shelves noted above. Ideally the LE should have included CSHELF values greater than 10^{-5}. However, we note that values of 10^{-5} to 10^{-6} have been found to well represent active Siple Coast ice-stream beds in model inversions (Pollard and DeConto, 2012b). Subsequent work with wider CSHELF ranges confirms that values around 10^{-5} are in fact optimal, with unrealistic behavior for larger values (Pollard et al., 2016).

Somewhat lower but still reasonable scores exist for lower CSHELF values of 10^{-7}, but only for higher OCFAC (3 to 10) and smaller TAUAST (1 to 2 kyr). This is of interest because smaller CSHELF values support thicker ice thicknesses at LGM where grounded ice has expanded over continental shelves, producing greater equivalent sea-level lowering and alleviating the LGM "missing-ice" problem (Clark and Tarasov, 2014). In order for the extra ice to be melted by present day, ocean melting needs to be more aggressive (higher OCFAC), and to recover in time from the greater bedrock depression at LGM, TAUAST has to be smaller (more rapid bedrock rebound). This glaciological aspect is explored in Pollard et al. (2016).

Scores are quite insensitive to the asthenospheric rebound time scale TAUAST (inner y axis), although there is a tendency to cluster around 2 to 3 kyr and to disfavor higher values (5 to 7 kyr) especially for high OCFAC.

4. Results: Comparisons of simple averaging vs. advanced statistical techniques

4.1. Single parameter ranges

The main results seen in Fig. 2 are borne out in Fig. 3. The left-hand panels show results using the simple averaging method, i.e., the average score for all runs in the LE with a particular parameter value. Triangles in these panels show the mean parameter value $V_m = \Sigma\ (S^{(n)}\ V^{(n)}) / \Sigma\ S^{(n)}$, where $S^{(n)}$ is the aggregate score and $V^{(n)}$ is the value of this parameter for run n (1 to 625), and whiskers show the standard deviation. The prominent signal of high CSHELF values (sticky ocean beds) is evident, along with the absence (near absence) of positive scores for the extreme CALV values of 1.7 (0.3), and the more subtle trends for OCFAC and TAUAST.

The right-hand panels of Fig. 3 show the same single-parameter "marginal" probably density functions for this LE, using the advanced statistical techniques described in Chang et al. (2015, 2016) and summarized above. For OCFAC, CSHELF and TAUAST, there is substantial agreement with the simple-averaging results in both the peak "best-fit" values and the width of the ranges. For CALV, the peak values agree quite well, but the simple-averaging distribution has a significant tail for lower CALV values that is not present in the advanced results; this might be due to the discrepancy function in the advanced method (Sect. 2.5), which has no counterpart in the simple averaging method.

4.2. Paired parameter ranges

Probability densities for pairs of parameter values are useful in evaluating the quality of LE analysis, and can display offsetting physical processes that together maintain realistic results, e.g., greater OCFAC and lesser CALV (Chang et al., 2014; 2015, 2016). In Fig. 4, the left-hand

panels show mean scores for pairs of the 4 parameters, using the simple averaging method and averaged over all LE runs with a particular pair of values. The right-hand panels show corresponding densities for the same parameter pairs using the advanced statistical techniques. Overall the same encouraging agreement is seen as for the single-parameter densities in Fi**g** 3, with the locations of the main maxima being roughly the same for each parameter pair. There are some differences in the extents of the maxima, notable for CALV where the zone of high scores with the simple averaging method extends to lower CALV values than with the advanced techniques, as seen for the individual parameters in Fig. 3. In general, though, there is good agreement between the two methods regarding parameter ranges in Figs. 3 and 4, suggesting that the simple averaging method is viable, at least for LE's with full factorial sampling of parameter space.

4.3. Equivalent-sea-level contribution

Fig. 5 illustrates the use of the LE to produce past envelopes of model simulations. Fig. 5a,b show equivalent sea-level (ESL) scatter plots for all 625 runs. Early in the runs around LGM (20 to 15 ka), the curves cluster into noticeable groups with the same CSHELF values, due to the relatively weak effects of the other parameters (OCFAC, CALV and TAUAST) for cold climates and ice sheets in near equilibrium. Fig. 5c,d show the mean and one-sided standard deviations for the simple method. Most of the retreat and sea-level rise occurs between ~14 to 10 ka. Glaciological aspects of the retreat will be discussed in more detail in Pollard et al. (2016).

Fig. 5e,f shows the equivalent mean and standard deviations derived from the advanced statistical techniques. There is substantial agreement with the simple-method curves in Fig. 5c,d, for most of the duration of the runs. The largest difference is around the Last Glacial Maximum ~20 to 15 ka, when mean sea levels are nearly ~2 m lower (larger LGM ice volumes) in the simpler method compared to the advanced. This may be due to the simpler method's scores using past 2-D grounding-line reconstructions (data type GL2D), which are not used in the advanced techniques.

Fig. 6 shows probability densities of equivalent sea level rise at particular times in the runs. Fig. 6a-d show results with the simple averaging method, computed using score-weighted densities and 0.2-m wide ESL bins (see caption). The uneven noise in this figure is due to the small number of parameter values in our LE. The separate peaks for LGM (-15000 yr) in Fig. 6a and b are due to the widely separated CSHELF values, and the relatively weak effects of the other parameters (OCFAC, CALV and TAUAST) for cold climates and ice sheets in near equilibrium. Fig. 6e shows the equivalent but much smoother probability densities using the advanced statistical techniques. All major aspects agree reasonably well with the simple averaging results, and the separate peaks for -15000 yr are smoothed into a single broad range.

5. Conclusions and further work

1. The simple averaging method, with quantities weighted by aggregate scores, produces results that are reasonably compatible with relatively sophisticated statistical techniques involving emulation, probability model/likelihood functions, and MCMC (Chang et al., 2015, 2016; Sect. 2.5). They are applied to the same LE with full factorial sampling in parameter space, for which both techniques yield smooth and robust results, and the advanced technique acts as a benchmark against which the simple method can be compared.

Unlike the advanced techniques, the simple averaging method cannot interpolate in parameter space, and so is limited practically to relatively few parameters (4 here) and a small number of values for each (5 here). Previous work using LE's with Latin HyperCube sampling (Applegate et al., 2012; Chang et al., 2014, 2015) has shown that the simple averaging method can fail if the sampling is too coarse, whereas the advanced technique provides smooth and meaningful results. This is primarily due to emulation and MCMC in the advanced techniques, which still interpolate successfully in the coarsely sampled parameter space. Of course, this distinction depends on the size of the LE and the coarseness of the sampling; somewhat larger LE's with Latin HyperCube sampling and fewer parameters can be amenable to the simple method. Note that this is not addressed in this paper; where just one full-factorial LE is used.

2. The best-fit parameter ranges deduced from the LE analysis generally fit prior expectations. In particular, the results strongly confirm that large basal sliding coefficients (i.e., slippery beds) are appropriate for modern continental-shelf oceanic areas. In further work we will assess heterogeneous bed properties such as the inner region of hard outcropping basement observed in the ASE (Gohl et al., 2013).The best-fit range for the asthenospheric relaxation time scale TAUAST values is quite broad, including the prior reference value ~3 kyr but extending to shorter times ~1 kyr. This may be connected with low upper-mantle viscosities and thin crustal thicknesses suggested in recent work (Whitehouse et al., 2012b; Chaput et al., 2014), which will be examined in further work with full Earth models (Gomez et al., 2013, 2015; Konrad et al., 2015).

3. The total Antarctic ice amount at the Last Glacial Maximum is equivalent to ~5 to 10 meters of global equivalent sea level below modern (Fig. 5). This is consistent with the trend in recent modeling studies (Ritz et al., 2001; Huybrechts, 2002; Philippon et al., 2006; Briggs et al., 2014; Whitehouse et al., 2012a,b; Golledge et al., 2012,2013,2014, whose LGM amounts are generally less than in older papers. (Note that Fig. 5 shows contributions only from our limited West Antarctic domain, but as shown in Mackintosh et al., 2011, the contribution from East Antarctica at LGM is much smaller, ~1 mesl). This suggests that Antarctic expansion is insufficient to explain the "missing ice" problem, i.e., the total volume of reconstructed ice sheets worldwide is less than the equivalent fall in sea-level records at that time by 15 to 20 meters (Clark and Tarasov, 2014). A subsequent paper (Pollard et al., 2016) examines this glaciological aspect in more detail but does not alter the conclusions here.

4. There are only minor episodes of accelerated WAIS retreat and equivalent sea-level rise in the simulations (Fig. 5), and none with magnitudes comparable to Melt Water Pulse 1A for instance, with ~15 mesl rise in ~350 years around ~14.5 ka (Deschamps et al., 2012), in apparent conflict with significant Antarctic contribution implied by sea-level fingerprinting studies (Bassett et al., 2005; Deschamps et al., 2012) and IRD-core analysis (Weber et al.,

2014). Model retreat rates are examined in more detail in Pollard et al. (2016), again without altering the findings here.

A natural extension of this work is to extend the Antarctic model simulations and LE methods into the future, using climates and ocean warming following Representative Concentration Pathway scenarios (Meinshausen et al., 2011). In these warmer climates we expect Marine Ice Sheet Instability to occur in WAIS basins, consistent with past retreats simulated in Pollard and DeConto (2009). Also drastic retreat mechanisms of hydrofracture and ice-cliff failure, not triggered in the colder-than-present simulations of this paper, may play a role, as found for the Pliocene in Pollard et al. (2015). Future applications with simple-average LE's are described in Pollard et al. (2016), and detailed future scenarios with another type of LE are described in DeConto and Pollard (2016).

Acknowledgements

We thank Zhengyu Liu and his group at U. Wisconsin for providing output of their coupled GCM simulation (TraCE-21ka; Liu et al., 2009; He et al., 2013; www.cgd.ucar.edu/ccr/TraCE) used for ocean forcing over the last 20,000 years. This work was supported in part by the following grants from the National Science Foundation: NSF-DMS-1418090 and the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO1240507 (DP, MH); PLIOMAX OCE-1202632, OPP-1341394, ANT-1443190 (DP); NSF Statistical Methods in the Atmospheric Sciences Network 1106862, 1106974, and 1107046 (WC).

Code availability

The code for the ice-sheet model (PSUICE-3D) is available on request from the corresponding author. The post-processing codes for the large-ensemble statistical analyses are highly tailored to specific sets of model output and are not made available; however, modules that compute scores for the individual data types are also available on request.

Appendix A: Model parameters varied in the large ensemble

The four model parameters (OCFAC, CALV, CSHELF and TAUAST) and their ranges in the large ensemble are summarized in Sect. 2.2. Their physical effects in the model and associated uncertainties are discussed in more detail here.

OCFAC is the main coefficient in the parameterization of sub-ice-shelf oceanic melt, which is proportional to the square of the difference between nearby water temperature at 400 m, and the pressure-melting point of ice. Oceanic melting (or freezing) erodes (or grows on) the base of floating ice shelves, as warm waters at intermediate depths flow into the cavities below the shelves. The resulting ice-shelf thinning reduces pinning points and lateral friction, and thus back stress on grounded interior ice. As mentioned above, recent increases in ocean melt rates are considered to be the main cause of ongoing downdraw and acceleration of interior ice in the ASE sector of WAIS (Pritchard et al., 2012; Dutrieux et al., 2014). High-resolution dynamical ocean models (Hellmer et al., 2012) are not yet practical on these time scales, and simple parameterizations of sub-ice-shelf melting such as the one used here are quite uncertain (eg., Holland et al., 2008). For small (large) OCFAC values, oceanic melting is reduced (increased), ice shelves thicken (thin), discharge of interior ice across the grounding line decreases (increases), and grounding lines tend to advance (retreat).

CALV is the main factor in the parameterization of iceberg calving at the oceanic edges of floating shelves. Calving has important effects on ice-shelf extent with strong feedback effects via buttressing of interior ice. However, the processes controlling calving are not well understood, probably depending on a combination of pre-existing fracture regime, large-scale stresses, and hydrofracturing by surface meltwater. There is little consensus on calving parameterizations. We use a common approach based on parameterized crevasse depths and their ratio to ice thickness (Benn et al. 2007; Nick et al., 2010). For small (large) CALV, calving is decreased (increased), producing more (less) extensive floating shelves, and greater (lesser) buttressing of interior ice.

CSHELF is the basal sliding coefficient for ice grounded on areas that are ocean bed today (and is not frozen to the bed). Coefficients under modern grounded ice are deduced by inverse methods (Pollard and DeConto, 2012b; Morlighem et al., 2013), but they are relatively unconstrained for modern oceanic beds, across which grounded ice advanced at the Last Glacial Maximum ~20 to 15 ka. Most oceanic beds around Antarctica are covered in deformable sediment today, due to Holocene marine sedimentation, and to earlier transport and deposition of till by previous ice advances. For these regions, coefficients are expected to be relatively high (i.e., slippery bed), but there is still a plausible range that has significant effects on model results, because it strongly controls the steepness of the ice-sheet surface profile and ice thicknesses, and thus the sensitivity to climate change. In this paper, we vary the sliding coefficient CSHELF uniformly for all modern-oceanic areas. (In further work, we will allow for heterogeneity such as the hard crystalline bedrock zone observed in the inner Amundsen Sea Embayment; Gohl et al., 2013).

TAUAST is the e-folding time of asthenosephic relaxation in the bedrock model component. Ice sheet evolution on long timescales is affected quite strongly by the bedrock response to varying ice loads, especially for marine ice sheets in contact with the ocean where bathymetry determines grounding-line depths. During deglacial retreat, the bedrock rebounds upwards due to reduced ice load, which slows down ice retreat due to shallower grounding-line depths and less discharge of interior ice. However, the $O(10^3)$-year lag in this process is important in reducing this negative feedback, and accelerates the positive feedback of Marine Ice Sheet Instability if the bed deepens into the ice-sheet interior. As in many large-scale ice-sheet models, our bedrock response is represented by a simple Earth model consisting of an elastic plate over a local e-folding relaxation towards isostatic equilibrium (Elastic Lithosphere Relaxing Asthenosphere). Based on more sophisticated global Earth models, the asthenospheric e-folding time scale is commonly set to 3 kyr (e.g., Gomez et al., 2013), but note that recent geophysical studies suggest considerably shorter time scales for some West Antarctic regions (Whitehouse et al., 2012b; Chaput et al., 2014). In further work we plan to perform large ensembles with the ice sheet model coupled to a full Earth model, extending Gomez et al (2013, 2015).

Appendix B: Data types and individual misfits

The 8 types of modern and past data used in evaluating the model simulations are summarized in Sect. 2.3. More details on the algorithms used to compute the individual mismatches M_1 to M_8 with model quantities are given below. The term "domain" refers to the nested model grid that spans all of West Antarctica, and we only compare with observational sites and data within this domain. Modern observed data is from the Bedmap2 dataset (Fretwell et al., 2013).

As discussed in Sects. 2.3 and 2.4, we use 2 approaches in scoring: (a) more closely following Gaussian error forms, and (b) with more heuristic forms. Some of the algorithms for individual misfits differ between the two, as indicated by bullets (a) and (b) below. For most data types, approach (a) uses mean-square errors, and (b) uses root-mean-square errors. For some data types, the errors are normalized not by observational uncertainty, but by an "acceptable model error magnitude" representing typical model departures from observations in reasonably realistic runs, if this is larger than observational error. Note that if this scaling uncertainty is the same for all data of a given type, it cancels out in the normalization of individual misfits (M_i to M_i' in Sect. 2.4), so has no effect on the further results.

1. TOTE: Modern grounding-line locations.

A' = total area of mismatch where model is grounded and observed is floating ice or ocean, or vice versa. A_{tot} = total area of the domain.

Approach (a): Misfit $M_1 = (A' / B)^2$, where $B = (A_{tot})^{1/2}\ \sigma_w$. Here B is the product of the linear domain size, and σ_w = 30 km representing the typical size of modern grounding-line location errors in "reasonable" model runs.

Approach (b): Misfit $M_1 = A' / A_{tot}$

2. TOTI: Modern floating ice-shelf locations.

A' = total area of mismatch where model has floating ice and observed does not, or vice versa. A_{tot} = total area of the domain.

Approach (a): Misfit $M_1 = (A' / B)^2$, where $B = (A_{tot})^{1/2}\ \sigma_w$. Here B is the product of the linear domain size, and σ_w = 30 km representing the typical size of modern floating-ice extent errors in "reasonable" model runs.

Approach (b): Misfit $M_1 = A' / A_{tot}$

3. TOTDH: Modern grounded ice thicknesses.

Approach (a): Misfit M_3 is the mean of $((h - h_{obs}) / \sigma_h)^2$, where h is model ice thickness, h_{obs} is observed ice thickness, and σ_h = 10 m represents the typical size of modern ice thickness errors in "reasonable" model runs. The mean is taken over areas with observed modern grounded ice.

Approach (b): Misfit M_3 is the root mean square of $(h - h_{obs})$, over areas with observed modern grounded ice.

4. TROUGH: Past grounding-line distance vs. time along centerline troughs of Pine Island Glacier, and optionally the Ross and Weddell basins. Observed distances at ages 20, 15, 10 and 5 ka are obtained from grounding-line reconstructions of the RAISED Consortium (2014): Anderson et al. (2014) for the Ross; Larter et al. (2014) for the Amundsen Sea, and Hillenbrand et al. (2014) for the Weddell, using their Scenario A of most retreated Weddell ice. Distances are then linearly interpolated in time between these dates. The centerline trough for Pine Island Glacier is extended across the continental shelf following the paleo-ice-stream trough shown in Jakkobsen et al. (2011). The resulting Pine Island Glacier transect vs. time is similar to that in Smith et al. (2014).

Approach (a): Misfit M_4 is the mean of $((x - x_{obs}) / \sigma_x)^2$, where x is model grounding-line position on the transect at a given time, x_{obs} is the reconstructed position, and σ_x = 30 km represents a typical difference in "reasonable" model runs, and is also midway between 'measured' and 'inferred' uncertainties in the reconstructed data (RAISED, 2014). The mean is taken over the period 20 to 0 ka.

Approach (b): Misfit M_4 is the root-mean-square of $(x - x_{obs})$, over the period 20 to 0 ka.

In this study just the Pine Island Glacier trough is used, but if the Ross and Weddell are used also, the means are taken over all 3 troughs.

5. GL2D: Past grounding-line locations. This uses reconstructed grounding-line maps for 20, 15, 10, 5 ka by the RAISED Consortium (RAISED, 2014; Anderson et al., 2014; Hillenbrand et al., 2014; Larter et al., 2014; Mackintosh et al., 2014; O Cofaigh et al., 2014), with vertices provided by S. Jamieson, pers. comm., and choosing their Scenario A for the Weddell embayment (Hillenbrand et al., 2014). The modern grounding line (0 ka) is derived from the Bedmap2 dataset (Fretwell et al., 2013). For this study only the Amundsen Sea region is considered. We allow for uncertainty in the past reconstructions by setting a probability of reconstructed floating ice or open ocean at each point P_{obs} as follows:

(i) Computing the distance D_1 from the reconstructed grounding line.

(ii) Dividing this distance by the sum D_2 of the (Kriged) reported uncertainty of nearby vertices (interpreting their "measured"= 10 km, "inferred"=50 km, "speculative"=100 km) and a distance that ramps up to 100 km depending on distance to the nearest vertex *dv* (i.e., 100 max [0, min [1, (*dv*-100)/200]]), to obtain a scaled distance $D_s = D_1/D_2$.

(iii) Setting the probability P_{obs} to a value decaying upwards or downwards from 0.5, i.e., to 0.5 e^{-Ds} if on the grounded side of the grounding line, or to 1 - 0.5 e^{-Ds} if on the non-grounded side.

Then the "mismatch probability" P_{mis} at each model grid point is set to 2 (0.5 - P_{obs}) if $P_{obs} <$ 0.5 and the model is not grounded, or 2 (P_{obs} - 0.5) if $P_{obs} > 0.5$ and the model is grounded. P_{mis} is zero if the model is not grounded anywhere on the non-grounded side of the observed grounding line, or if it is grounded anywhere on the grounded side. Thus, if the model and observed grounding lines coincide exactly everywhere, then P_{mis} is zero at all points, regardless of the observational uncertainty reflected in P_{obs} (which seems a desirable feature).

Approach (a): Misfit M_5 is the mean of the squared mismatch probabilities $(P_{mis})^2$, with means computed over3 separate subdomains: Ross Sea, Amundsen Sea, and Weddell Sea embayments (defined crudely by intervals of longitude: 150E to 120W, 120W to 90W, and 90W to 0, respectively). In this study we only use the mean for the Amundsen Sea sector. Similarly to

TOTE and TOTI, the areal mean is increased by a factor $(A_{tot})^{1/2} / \sigma_w$, where A_{tot} is the total subdomain area and σ_w = 100 km is a representative width scale of reasonable past grounding-zone mismatches. Finally, the mean values for each of the reconstructed past times (20, 15, 10 and 5 ka) are averaged together equally.

Approach (b): Misfit M_5 is the mean of P_{mis} over the Amundsen Sea sector subdomain, with no adjustment factor to A_{tot}, and otherwise as for (a) above.

6. RSL: Past Relative Sea Level (RSL) records. This uses the compilation by Briggs and Tarasov (2013) of published RSL data vs. time at sites close to the modern coastline. Following those authors, the model RSL = [$SL(t)$ - $h_b(t)$] - [$SL(0)$ - $h_b(0)$], where $SL(t)$ is global sea level (with t=0 at modern) and h_b is bed elevation, at the closest model grid point to the observed site. The minimum model-minus-observed difference δRSL for each observed datum is used, i.e., the minimum elevation difference value over all model times within the range of the observational time uncertainty ($t_{obs} \pm \sigma_{to}$).

Approach (a): Misfit M_6 is the weighted mean of $(\delta RSL / \sigma_{zo})^2$, where σ_{zo} is the observational RSL uncertainty. Just as in Briggs and Tarasov (2013), the default for σ_{zo} is much larger for one-sided constraints (50 m) than absolute constraints (2 m). To reduce the influence of many nearby (and presumably correlated) data, we closely follow Briggs and Tarasov (2013) and apply "intra-data-type weighting" in calculating the mean. The weights are inversely proportional to the number of measurements within a distance L of each other, where L is equivalent to 5^o latitude (~550 km), so that each ~L-sized cluster of data contributes ~equally to the overall mean.

Approach (b): Misfit M_6 is the weighted mean of max [0, |RSL| - σ_{zo}]. The uncertainties σ_{zo} and the intra-data-type weights are the same as in (a).

7. ELEV/DSURF: This uses a combination of two compilations of cosmogenic data: elevation vs. age in Briggs and Tarasov (2013) for ELEV, and thickness change from modern vs. age in RAISED (2014) (with individual citations as above) for DSURF.

For ELEV, the calculations closely follow Briggs and Tarasov (2013, their sec. 4.2):

(i) a time series of model ice surface is used, with sea level and bedrock elevation changes subtracted out, for the closest model grid point to each ELEV datum.

(ii) Only model elevations with a "deglaciating trend" are used, i.e., the model elevation for each time is replaced by the maximum elevation between that time and the present, if the latter is greater, allowing for an uncertainty $\Delta h = \sqrt{2}\,\sigma_h$, as in Briggs and Tarasov (2013).

(iii) The mismatch for each datum is the minimum of $(\delta h/\,\sigma_h)^2 + (\delta t/\,\sigma_t)^2$ over the time series, where δh is the elevation difference from observed and δt is the time difference, $\sigma_h = [\sigma_{hobs}{}^2 + (100\text{ meters})^2]^{1/2}$, and σ_{hobs} and σ_t are the observational uncertainties in elevation and time respectively.

Approach (a): Misfit M_7 is the weighted mean of the mismatches for ELEV above, with intra-data-type weighting exactly as described for RSL above. The DSURF type is not used in approach (a).

Approach (b): For approach (b), ELEV calculations as above are combined with DSURF calculations.

The DSURF calculations are simpler: for each datum, the time series of model surface elevations h_s at the closest model grid point is used. The minimum model-minus-observed difference $\delta h_s^{\min}$ is found, i.e., the minimum difference over all model times within the range of the observational time uncertainty ($t_{obs} \pm \sigma_{to}$). The mismatch for the datum is max [0, $\delta h_s^{\min}$ - σ_h] where σ_h is the observational elevation uncertainty. The mean over all data is taken, weighted by intra-data-type weighting as described above. Finally, the ELEV and DSURF misfits are converted into separate normalized scores (S_{7a}, S_{7b}) as in Sect. 2.4(b), which are then combined into one individual score $S_7 = (S_{7a}\, S_{7b})^{1/2}$.

8. UPL: This uses modern uplift rates on rock outcrops, using the compilation in Whitehouse et al. (2012b). For each observed site, the model's modern $\partial h_b / \partial t$ at the closest model grid point is used.

Approach (a): The mismatch at each datum is $[(U_{mod} - U_{obs}) / \sigma_{uobs}]^2$, where U_{mod} and U_{obs} are model and observed uplift rates respectively, and σ_{uobs} is the observed 1-σ uncertainty. The misfit M_8 is the mean over all data points, using intra-data-type weighting as above.

Approach (b): The mismatch at each datum is $(U_{mod} - U_{obs})^2$, and the misfit M_8 is the root-mean-square over all data points, with no intra-data-type weighting (justified by the relatively uniform distribution of data points).

Appendix C: Comparison of results with two scoring approaches

As discussed in Sect. 2.3, the choice of formulae and algorithms to calculate model vs. data misfits and scores in the simple averaging method is somewhat heuristic, and different choices are also appropriate for complex model-data comparisons with widespread data points, very different types of data, and with many model-data error types not being strictly Gaussian. Two possible approaches are described above (Sect. 2.4, Appendix B): Approach (a) uses formulae closely following Gaussian error distribution forms, and approach (b) uses more heuristic forms. Approach (b) is used for all results in the main paper. In this appendix the simple-averaging results (Figs. 2-5) are compared using both approaches. No significant differences are found, especially in the LE-averaged results, which suggests that different reasonable approaches to misfits and scoring yield robust statistical results for the ensemble.

In Fig. C1, the individual scores have much the same patterns over 4-D parameter space. There are some minor differences in the relative magnitudes of very good, vs. poor but still meaningful scores, which we have compensated for to some extent in the two color scales, but these do not lead to any significant differences in the averaged results in the following figures.

In the parameter-pair scores (Fig. C2), the overall patterns are very similar. The biggest difference is for CALV vs. TAUAST, where the scores for approach (a) are higher and more tightly concentrated.

In the plots of equivalent sea level versus time (Fig. C3), approach (a) generally favors runs with less ice volume during LGM and retreat, compared to approach (b) (red curves, Figs. C3c vs. d). On the other hand, the single best-scoring run in approach (a) retreats later than the corresponding run in approach (b) (black curves, Fig. C3a vs. b). Generally, these differences are minor compared to the overall model behavior through the deglaciation.

In the density distributions of equivalent sea level at particular times (Fig. C4), there is very little difference between the 2 approaches. The size of the ~5 m peak at 15 ka is larger in approach (b), but as discussed in Sect. 4.3, these separate peaks at 15 ka are due to the widely spaced CSHELF parameter values in the ensemble, and their relative sizes have little significance.

Appendix D: Span of data by the Large Ensemble

This appendix compares envelopes of model results with corresponding types of geologic data used in the LE scoring. The main goal is to demonstrate that the envelopes of the 625-member ensemble adequately spans the data; i.e., at least some runs yield results that fall on both sides of each type of data, so that ensemble averages may potentially represent reasonably realistic ice sheet behavior (even if no single model run is close to all data types).

For modern data (grounded and floating ice extents, grounded ice thicknesses), the standard model has previously been shown to yield quite realistic simulations, both for perpetual modern climate and at the end of long-term glacial-interglacial runs (Pollard and DeConto, 2012a). Modern grounded ice thicknesses are close to observed mainly because of the inverse procedure in specifying the distribution of basal sliding coefficients (Pollard and DeConto, 2012b). Here we concentrate on fits to geologic data.

Fig. D1 compares scatter plots of Relative Sea Level in all 625 runs with RSL records, for the 3 sites within the model's West Antarctic domain (Briggs and Tarasov, 2013). The data for each site fall well within the overall model envelope, and in most cases within the envelopes of the top 120-scoring runs (colored curves). Similar comparisons for single runs are shown in Gomez et al. (2013), both using the simple bedrock model as here (their "uncoupled" runs), and coupled to a global Earth-sea level model.

Similarly, Fig. D2 compares elevation vs. age time series for all 625 runs with cosmogenic data at the 18 sites within the model domain (Briggs and Tarasov, 2013). With a few exceptions, the data lie within the LE model envelopes, although elevations at many of the sites are lower than in most of the model runs. At Reedy Glacier, the model exhibits oscillations of ~200 m amplitude and several hundred year period; these might be due to internal variability of ice streams as seen elsewhere in West Antarctica in Pollard and DeConto (2009).

Fig. D3 shows modern uplift rates for all model runs, at the 26 sites in the Whitehouse et al. (2012b) compilation that lie within the mode domain. Again, nearly all of the observed values lie within the overall model envelope. The geographic distribution for single runs is compared with observed in Gomez et al. (2013), both using a simple bedrock model ("uncoupled"), and coupled to a global Earth-sea level model.

The remaining past data types (GL2D and TROUGH) concern grounding-line locations during last deglacial retreat, and are less amenable to scatter plots, but can be compared with model averaged results. Fig. D4 shows maps of probability (0-1) of the presence of grounded ice at particular times, deduced by score-weighted averages over the ensemble. The thick black lines at 20, 15, 10 and 5 ka show grounding-line positions in the reconstructions of the RAISED Consortium (RAISED, 2014). (The figures do not show the uncertainty information associated with the data, which is used in the scoring; Appendix B). At all of these times, the envelopes of the model "grounding zone", i.e., the areas with intermediate probability values, span or are close to the observed positions.

Similarly, Fig. D5 shows model probabilities (0-1) of grounded ice vs. time along the centerline transects of the major West Antarctic embayments. Again, the model envelopes mostly span the various observed estimates for each transect (from RAISED, 2014, and various earlier studies).

Taken together, the various model vs. data comparisons in this Appendix show that the model's ensemble envelopes do encompass the ranges of data satisfactorily, as necessary for meaningful interpretations of the statistical results.

References

Anderson, J.B., Conway, H., Bart, P.J., Witus, A.E., Greenwood, S.L., McKay, R.M., Hall, B.L., Ackert, R.P., Licht, K., Jakobsson M., and Stone, J.O.: Ross Sea paleo-ice sheet drainage and deglacial history during and since the LGM. Quat. Sci. Res., 100, 31-54, 2014.

Applegate, P.J., Kirchner, N., Stone, E.J., Keller, K., and Greve, R.: An assessment of key model parametric uncertainties in projections of Greenland Ice Sheet behavior. The Cryosphere, 6, 589-606. 2012.

Bassett, S.E., Milne, G.A., Mitrovica, J.X., and Clark, P.U.: Ice sheet and solid Earth influences on far-field sea-level histories, Science, 309, 925-928, 2005.

Benn, D.I., Warren, C.R., and Mottram, R.H.: Calving processes and the dynamics of calving glaciers. Earth-Sci. Rev.82, 143–179, 2007.

Briggs, R.D. and Tarasov, L.: How to evaluate model-derived deglaciation chronologies: a case study using Antarctica. Quat. Sci. Rev., 63, 109-127, 2013.

Briggs, R.D., Pollard, D., and Tarasov, L.: A glacial systems model configured for large ensemble analysis of Antarctic deglaciation. The Cryosphere, 7, 1949-1970, 2013.

Briggs, R.D., Pollard, D., and Tarasov, L.: A data-constrained large-ensemble analysis of Antarctic evolution since the Eemian. Quat. Sci. Rev., 103, 91-115, 2014.

Chang, W., Applegate, P., Haran, M., and Keller, K: Probabilistic calibration of a Greenland Ice Sheet model using spatially-resolved synthetic observations: Toward projections of ice mass loss with uncertainties. Geosci. Model Devel., 7, 1933-1943, 2014.

Chang, W., Haran, M., Applegate, P., and Pollard, D.: Calibrating an ice sheet model using high-dimensional non-Gaussian spatial data. J. Amer. Stat. Assoc., doi:10.1080/01621459.2015.1108199, 2015.

Chang, W., Haran, M., Applegate, P., and Pollard, D.: Improving ice sheet model calibration using paleo and modern observations: A reduced dimensional approach. Available at: http://arxiv.org/abs/1510.01676, 2016.

Chaput, J., Aster, R., Huerta, A., Sun, X., Lloyd, A., Wiens, D., Nyblade, A., Anandakishnan, S., Winberry, P., and Wilson, T.: The crustal thickness of West Antarctica. J. Geophys. Res., 119, 1–18, doi:10.1002/2013JB010642, 2014.

Clark, P.U. and Tarasov, L.: Closing the sea level budget at the Last Glacial Maximum. Proc. Nat. Acad. Sci., 111, 15861–15862, 2014.

Cornford, S.L. et al.: Adaptive mesh, finite volume modeling of marine ice sheets. J. Comp. Phys., 232, 529-549, 201.

DeConto, R.M., Pollard, D., and Kowalewski, D.: Modeling Antarctic ice sheet and climate variations during Marine Isotope Stage 31. Glob. Plan. Change, 88-89, 45-52, 2012.

DeConto, R.M. and Pollard, D.: Antarctica's contribution to Last Interglacial and future sea-level rise. Nature, 531, 591-597, 2016.

Deschamps, P. et al.: Ice-sheet collapse and sea-level rise at the Bolling warming 14,600 years ago. Nature, 483, 559-564, 2012.

Docquier, D., Pollard, D., and Pattyn, F.: Thwaites Glacier grounding-line retreat: Influence of width and buttressing parameterizations. J. Glaciol., 60, 305-313, 2014.

Durand, G., Gagliardini, O., Favier, L., Zwinger, T., and le Meur, E.: Impact of bedrock description on modeling ice sheet dynamics. Geophys. Res. Lett., 38, L20501, doi: 10.1029/2011GL048892, 2011.

Dutrieux, P., De Rydt, J., Jenkins, A., Holland, P.R., Ha, H.K., Lee, S.H., Steig, E.J., Ding, Q., Abrahamsen E.P., and Schröder, M.: Strong sensitivity of Pine Island ice-shelf melting to climatic variability. Science, 373, 174-178, 2014.

Favier, L., Gagliardini, O., Durand G., and Zwinger, T.: A three-dimensional full Stokes model of the grounding line dynamics: effect of a pinning point beneath the ice shelf. The Cryosphere, 6, 101-112, 2012.

Favier, L., Durand, G., Cornford, S.L., Gudmundsonn, G.H., Gagliardini, O.,Gillet-Chaulet, F., Zwinger, T., Payne, A.J., and Le Brocq, A.M.: Retreat of Pine Island Glacier controlled by marine ice-sheet instability. Nature Clim. Change, 4, 117-121, 2014.

Fretwell, P., et al.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. The Cryosphere, 7, 375-393, 2013.

Gohl, K., Uenzelmann-Neben, G., Larter, R.D., Hillenbrand, C.-D., Hochmuth, K., Kalberg, T., Weigelt, E., Davy, B., Kuhn, G., and Nitsche, F.O.: Seismic stratigraphic record of the Amundsen Sea Embayment shelf from pre-glacial to recent times: Evidence for a dynamic West Antarctic ice sheet. Marine Geol., 344, 115-131, 2013.

Gladstone, R.P., Lee, V., Rougier, J., Payne, A.J., Hellmer, H., LeBrocq, A.M., Shepherd, A., Edwards, T.L., Gregory, J., and Cornford, S.: 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.

Golledge, N.R., Fogwill, C.J., Mackintosh, A.N., and Buckley. K.M.: Dynamics of the last glacial maximum Antarctic ice-sheet and its response to ocean forcing. Proc. Nat. Acad. Sci., 109, 16052-16056, 2012.

Golledge, N.R., Levy, R.H., McKay, R.M., Fogwill, C.J., White, D.A., Graham, A.G.C., Smith, J.A., Hillenbrand, C.-D., Licht, K.J., Denton, G.H., Ackert Jr., R.P., Maas, S.M., and Hall, B.L.: Glaciology and geological signature of the Last Glacial Maximum Antarctic ice sheet. Quat. Sci. Rev., 78, 225-247, 2013.

Golledge, N.R., Menviel, L., Carter, L., Fogwill, C.J., England, M.H., Cortese, G., and Levy, R.H.: Antarctic contribution to meltwater pulse 1A from reduced Southern Ocean overturning. Nature Comm., doi: 10.1038/ncomms6107, 2014.

Gomez, N., Pollard, D., and Mitrovica, J.X.: A 3-D coupled ice sheet-sea level model applied to Antarctica through the last 40 ky. Earth Plan. Sci. Lett., 384, 88-99, 2013.

Gomez, N., Pollard, D., and Holland, D.: Sea level feedback lowers projections of future Antarctic Ice Sheet mass loss. Nature Commun., 6, 8798, doi:10.1038/ncomms9798, 2015.

He, F., Shakun, J.D., Clark, P.U., Carlson, A.E., Liu, Z., Otto-Bliesner, B.L., and Kutzbach, J.E.: Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation. Nature, 494, 81-85, 2013.

Hellmer, H., Kauker, F., Timmermann, R., Determann, J., and Rae, J.: Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. Nature, 485, 225–228. 2012.

Hillenbrand, C.-D., Bentley, M.J., Stolldorf, T., Hein, A.S., Kuhn, G., Graham, A.C.G., Fogwill, C.J., Kristoffersen, Y., Smith, J.A., Anderson, J.B., Larter, R.D., Melles, M., Hodgson, D.A., Mulvaney, R., and Sugden, D.E.: Reconstruction of changes in the Weddell Sea sector of the Antarctic Ice Sheet since the Last Glacial Maximum. Quat. Sci. Rev., 100, 111-136, 2014.

Hjort, C., Ingolfsson, O., M€oller, P., Lirio, J.M., 1997.

Holland, P.R., Jenkins, A., and Holland, D.M.: The response of ice shelf basal melting to variations in ocean temperature. J. Clim., 21, 2558-2572, 2008.

Holschuh, N., Pollard, D., Anandakrishnan, S., and Alley, R.: Evaluating Marie Byrd Land stability using an improved basal topography. Earth Plan. Sci. Lett., 408, 362-369, 2014.

Huybrechts, P.: Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. Quat Sci Rev: 21, 203-231, 2002.

Jakobsson, M. et al.: Geological record of ice shelf break-up and grounding line retreat, Pine Island Bay, West Antarctica. Geology, 39, 691-694, 2011.

Joughin, I., Smith, B.E., and Medley, B.: Marine ice sheet collapse potentially under way for the Thwaites Glacier basin, West Antarctica. Science, 344, 735–738, 2014.

Konrad, H., Sasgen, S., Pollard, D., and Klemann, V.: Potential of the solid-Earth response for limiting long-term West Antarctic Ice Sheet retreat in a warming climate. Earth Plan. Sci. Lett., 432, 254-264, 2015.

Larter , R.D., Anderson, J.B., Graham, A.G.C., Gohl, K., Hillenbrand, C.-D., Jakobsson, M., Johnson, J.S., Kuhn, G., Nitsche, F.O., Smith, J.A., Witus, A.E., Bentley, M.J., Dowdeswell, J.A., Ehrmann, W., Klages, J.P., Lindow, J., Ó Cofaigh, C., and Spiegel, C.: Reconstruction of changes in the Amundsen Sea and Bellingshausen Sea sector of the West Antarctic Ice Sheet since the Last Glacial Maximum. Quat. Sci. Rev., 100, 55-86, 2014.

Le Brocq, A.M., Payne, A.J., and Vieli, A.: An improved Antarctic dataset for high resolution numerical ice sheet models (ALBMAP v1). Earth Sys. Sci. Data, 2, 247-260, 2010.

Liu, Z., Otto-Bliesner, B.L., He, F., Brady, E.C., Tomas, R., Clark, P.U., Carlson, A.E., Lynch-Stieglitz, J., Curry, W., Brook, E., Erickson, D., Jacob, R., Kutzbach, J., and Cheng, J.: Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. Science, 325, 310-314, 2009. Extended to modern: www.cgd.ucar.edu/ccr/TraCE.

Mackintosh, A., Golledge, N., Domack, E., Dunbar, R., Leventer, A., White, D., Pollard, D., DeConto, R., Fink, D., Zwartz, D., Gore, D., and Lavoie, C.: Retreat of the East Antarctic ice sheet during the last glacial termination. Nature Geosci., 4, 195-202, 2011.

Mackintosh, A.N., Verleyen, E., O'Brien, P.E., White, D.A., Jones, R.S., McKay, R., Dunbar, R., Gore, D.B., Fink, D., Post, A.L., Miura, H., Leventer, A., Goodwin, I., Hodgson, D.A., Lilly, K., Crosta, X., Golledge, N.R., Wagner, B., Berg, S., van Ommen, T., Zwartz, D., Roberts, S.J., Vyverman, W., and Masse, G.: Retreat history of the East Antarctic Ice Sheet since the Last Glacial Maximum. Quat. Sci. Rev. 100, 10-30, 2014.

Maris, M.N.A., van Wessem, J.M., van de Berg, W.J., de Boer, B., and Oerlemans, J.: A model study of the effect of climate and sea-level change on the evolution of the Antarctic Ice Sheet from the Last Glacial Maximum to 2100. Clim. Dyn., 45, 837-851, 2015.

McNeall, D.J., Challenor, P.G., Gattiker, J.R., and Stone, E.J.: The potential of an observational data set for calibration of a computationally expensive computer model. Geosci. Model Dev., 6, 1715-1728, 2013.

Meinshausen, M., Smith, S.J., Calvin, K., Daniel, J.S., Kainuma, M.L.T., Lamarque, J.-F., Matsumoto, K., Montzka, S.A., Raper, S.C.B., Riahi, K., Thomson, A., Velders, G.J.M., and van Vuuren, D.P.P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Clim. Change, 109, 213-241, 2011.

Mercer, J.H.: West Antarctic ice sheet and CO_2 greenhouse effect: a threat of disaster. Nature, 271, 321–325, 1978.

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., and Aubry, D: Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models

for Pine Island Glacier, West Antarctica. Geophys. Res. Lett., 37, L14502, 1-6, doi: 10.1029/2010GL043853, 2010.

Morlighem, M., Seroussi, H., Larour, E., and Rignot, E.: Inversion of basal friction in Antarctica using exact and incomplete adjoints of a higher-order model. J. Geophys. Res.-Earth Surface, 3, 1746-1753, 2013.

Nick, F.M., van der Veen, C.J., Vieli, A., and Benn, D.I.: A physically based calving model applied to marine outlet glaciers and implications for the glacier dynamics. J. Glaciol.56, 781–794, 2010.

O Cofaigh, C., Davies, B.J., Livingstone, S.J., Smith, J.A., Johnson, J.S., Hocking, E.P., Hodgson, D.A., Anderson, J.B., Bentley, M.J., Canals, M., Domack, E., Dowdeswell, J.A., Evans, J., Glasser, N.F., Hillenbrand, C.-D., Larter, R.D., Roberts, S.J., and Simms, A.R.: Reconstruction of Ice-Sheet Changes in the Antarctic Peninsula Since the Last Glacial Maximum. Quat. Sci. Rev. 100, 87-110, 2014.

Parizek, B.R., Christianson, K., Anandakrishnan, S., Alley, R.B., Walker, R.T., Edwards, R.A., Wolfe, D.S., Bertini, G.T., Rinehart, S.K., Bindschadler, R.A., and Nowicki, S.M.J.: Dynamic (in)stability of Thwaites Glacier, West Antarctica. J. Geophys. Res.: Earth Surface, 118, 638–655, 2013.

Pattyn, F. and 27 co-authors: Grounding-line migration in plan-view marine ice-sheet models: results of the ice2sea MISMIP3d intercomparison. J. Glaciol., 59, 410-422, 2013.

Peltier, W.R. 2004. Global glacial isostacy and the surface of the ice-age Earth: The ICE-5G (VM2) model and GRACE. 2004. Ann. Rev. Earth Planet. Sci., 32, 111–149.

Philippon, G., Ramstein, G., Charbit, S. Kageyama, M., Ritz, C., and Dumas. C.: Evolution of the Antarctic ice sheet throughout the last deglaciation: A study with a new coupled climate - north and south hemisphere ice sheet model. Earth Plan. Sci. Lett., 248, 750-758, 2006.

Pollard, D. and DeConto, R.M.: Modelling West Antarctic ice sheet growth and collapse through the past five million years. Nature, 458, 329-332, 2009.

Pollard, D. and DeConto, R.M.: Description of a hybrid ice sheet-shelf model, and application to Antarctica. Geosci. Model Devel., 5, 1273-1295, 2012a.

Pollard, D. and DeConto, R.M.: A simple inverse method for the distribution of basal sliding coefficients under ice sheets, applied to Antarctica. The Cryosphere, 6, 953-971, 2012b.

Pollard, D., DeConto, R.M., and Alley, R.B.: Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. Earth Plan. Sci. Lett., 412, 112-121, 2015.

Pollard, D., DeConto, R., Chang, W., Haran, M., and Applegate, P.: Large ensemble modeling of the Antarctic Ice Sheet: Contributions to past and future sea-level rise. The Cryosphere Discuss., to be submitted, 2016.

Pritchard, H.D., Ligtenberg, S.R.M., Fricker, H.A., Vaughan, D.G., van den Broeke, M.R., and Padman, L.: Antarctic ice-sheet loss driven by basal melting of ice shelves. Nature, 484, 502-505, 2012.

Ritz, C, Rommelaere, V., and Dumas, C.: Modeling the evolution of Antarctic ice sheet over the last 420,000 years: Implications for altitude changes in the Vostok region. J Geophys Res-Atmos 106, 31943-31964, 2001.

The RAISED Consortium: A community-based geological reconstruction of Antarctic Ice Sheet deglaciation since the Last Glacial Maximum. Quat. Sci. Rev., 100, 1-9, 2014.

Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H., and Scheuchl, B: Widespread, rapid grounding line retreat of Pine Island,Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. 2014. Geophys. Res. Lett., doi: 10.1002/2014GL060140, 3502-3509, 2014.

Schoof, C.: Ice sheet grounding line dynamics: steady states, stability, and hysteresis. J. Geophys. Res. - Earth Surf., 112, F03S28, doi: 10.1029/2006JF000664, 2007.

Shepherd, A. et al.: A reconciled estimate of ice-sheet mass balance. Science, 338, 1193-1198, 2012.

Smith, J.A., Hillenbrand, C.-D., Kuhn, G., Klages, J.P., Graham, A.G.C., Larter, R.T., Ehrmann, W., Moreton, S.G., Wiers, S., and Frederichs, T.: New constraints on the timing of West Antarctic Ice Sheet retreat in the eastern Amundsen Sea since the last Glacial Maximum. Glob. Planet. Change, 122, 224-237, 2014.

Stone, E.J., Lunt, D.J., Annan, J.D., and Hargreaves, J.C.: Quantification of the Greenland ice sheet contribution to Last Interglacial sea level rise. Clim. Past, 9, 621-639, 2013.

Tarasov, L., Dyke, A.S., Neal, R.M., and Peltier, W.R.: A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. Earth Plan. Sci. Lett., 315-316, 30-40, 2012.

Urban, N.M. and Fricker, T.E.: A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth System model. Computers and Geosciences, 36, 746-755, 2010.

Vaughan, D.G.: West Antarctic Ice Sheet collapse: the fall and rise of a paradigm. Clim. Change, 91, 65–79, 2008.

Weber, M. et al.: Millennial-scale variability in Antarctic ice-sheet discharge during the last deglaciation. Nature, 510, 134-138, 2014.

Weertman, J.: Stability of the junction of an ice sheet and an ice shelf. J.Glaciol.13, 3–11, 1974.

Whitehouse, P.L., Bentley, M.J., and Le Brocq, A.M.: A deglacial model for Antarctica: geological constraints and glaciological modeling as a basis for a new model of Antarctic glacial isostatic adjustment. Quat. Sci. Res., 32, 1-24, 2012a.

Whitehouse, P.L., Bentley, M.J., Milne, G.A., King, M.A., and Thomas, I.D.: A new glacial isostatic model for Antarctica: calibrated and tested using observations of relative sea-level change and present-day uplifts. Geophys. J. Internat., 190, 1464-1482 2012b.

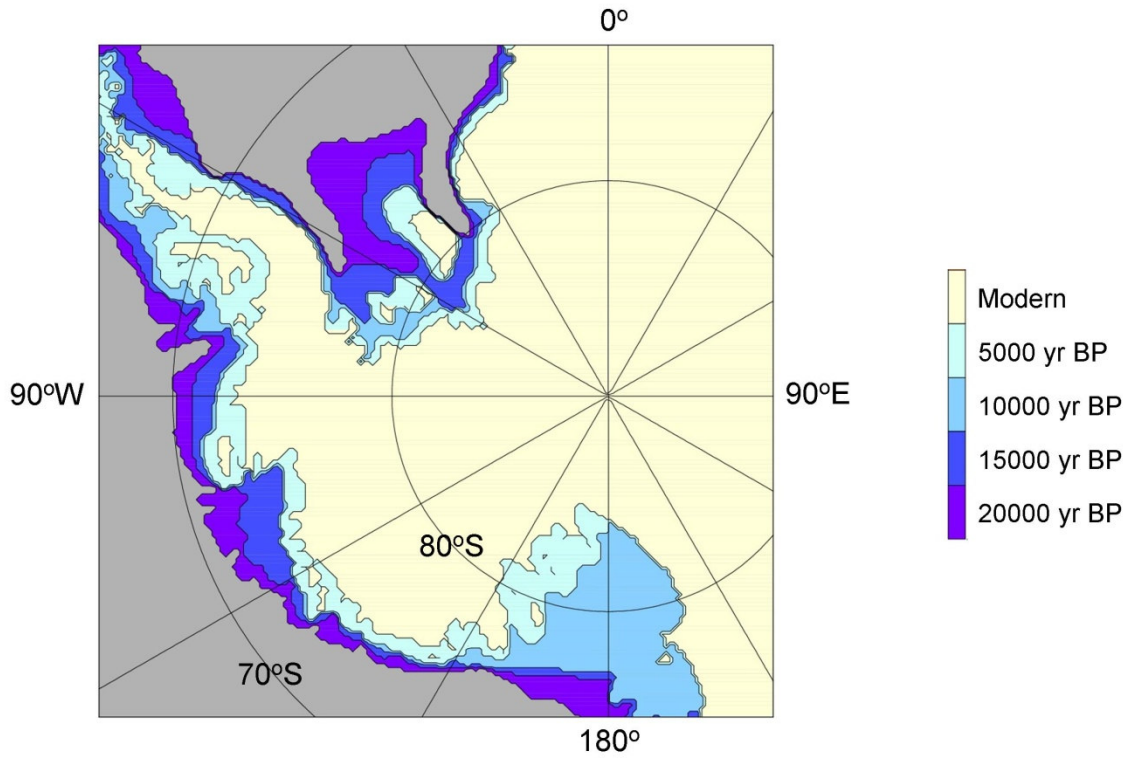

Figure 1. Geographical map of West Antarctica. Light yellow shows the modern extent of grounded ice (using Bedmap2 data; Fretwell et al., 2013). Blue and purple areas show expanded grounded-ice extents at 5, 10, 15 and 20 ka (thousands of years before present) reconstructed by the RAISED consortium (2014), plotted using their vertex information (S. Jamieson, pers. comm.), and choosing their Scenario A for the Weddell embayment (Hillenbrand et al., 2014). These maps are used in the large ensemble scoring (TOTE, TROUGH and GL2D data types, Sect. 2.3).

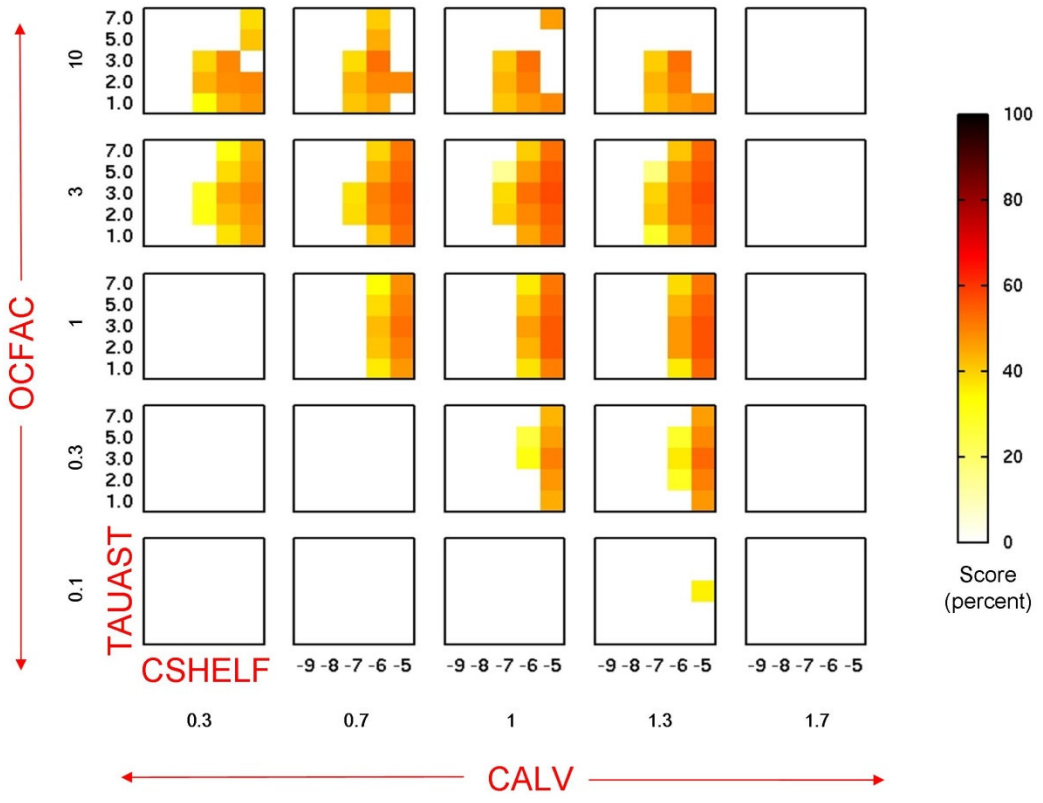

Figure 2. Aggregate scores for the complete large ensemble suite of runs (625 runs, 4 model parameters, 5 values each, Sect. 2.2), used in the simple method with score-weighted averaging. The score values range from 0 (white, no skill) to 100 (dark red, perfect fit). The figure is organized to show the scores in the four-dimensional space of parameter variations. The four parameters are: CSHELF = basal sliding coefficient in modern oceanic areas (exponent x, 10^{-x} m a^{-1} Pa^{-2}). TAUAST = e-folding time of bedrock-elevation isostatic relaxation (kyrs). OCFAC = oceanic-melt-rate coefficient at base of floating ice shelves (non-dimensional). CALV = calving-rate factor at edge of floating ice shelves (non-dimensional). Since each parameter only takes 5 values, the results are blocky, but effectively show the behavior of the score over the full range of plausible parameter values.

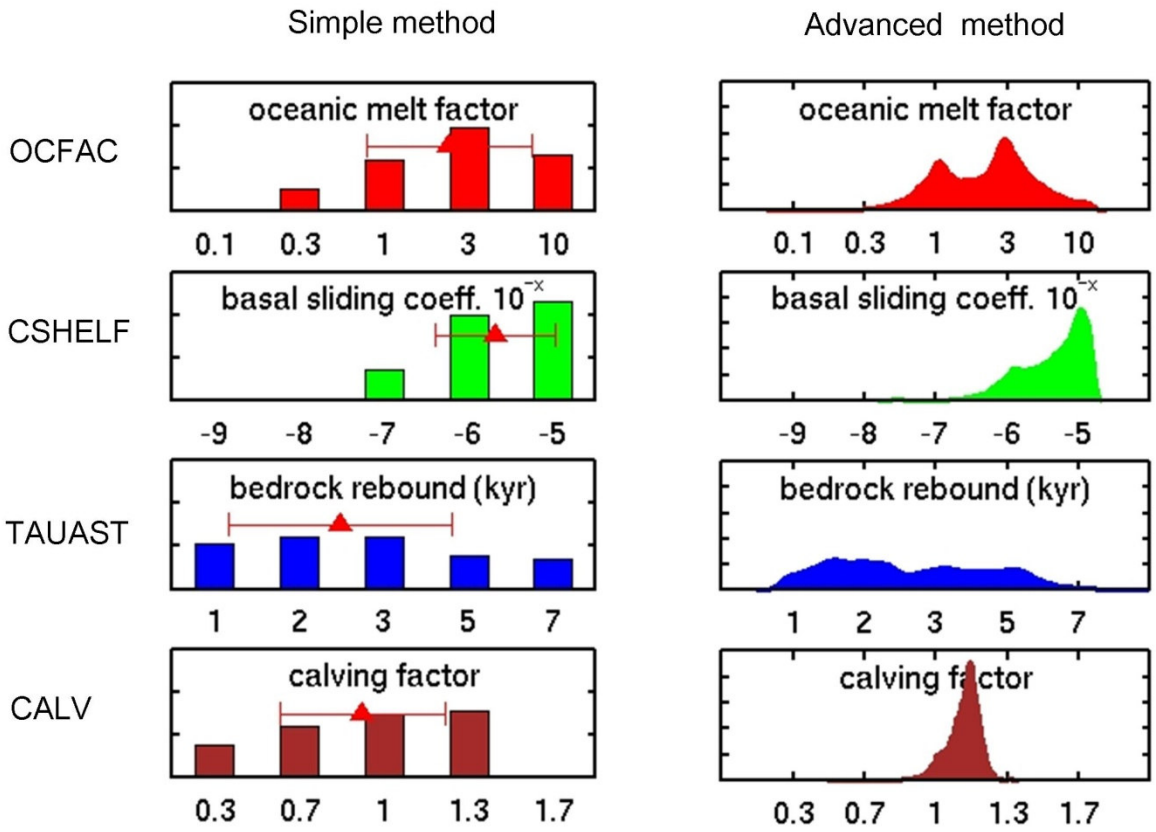

Figure 3. Left-hand panels: Ensemble-mean scores for individual parameter values, using the simple averaging method. The red triangle shows the mean, and whiskers show the 1-sigma standard deviations. **Right-hand panels:** Probability densities for individual parameters, using the advanced statistical techniques in Chang et al. (2016) extended as described in Sect. 2.5.

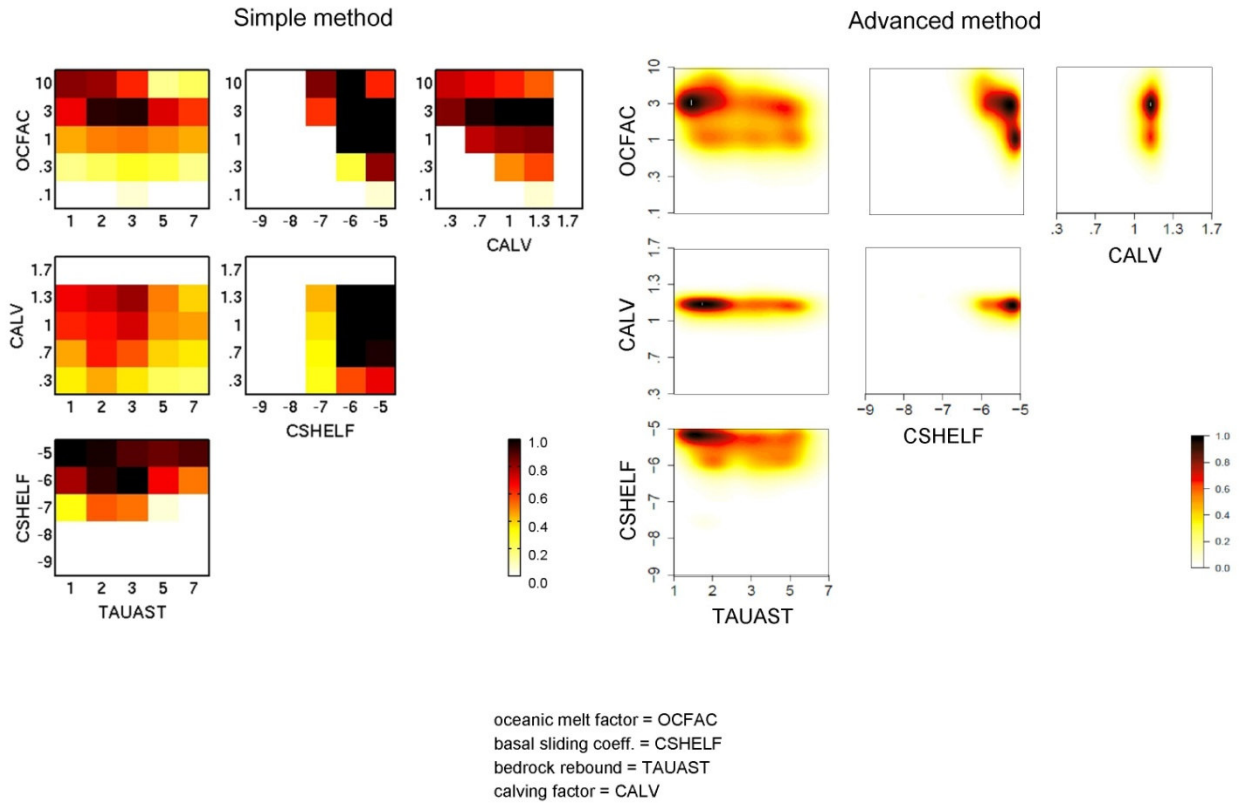

Figure 4. Left-hand panels: Ensemble-mean scores for pairs of parameters, using the simple averaging method. **Right-hand panels:** Probability densities for pairs of parameters, using the advanced statistical techniques in Chang et al. (2016) extended as described in Sect. 2.5.

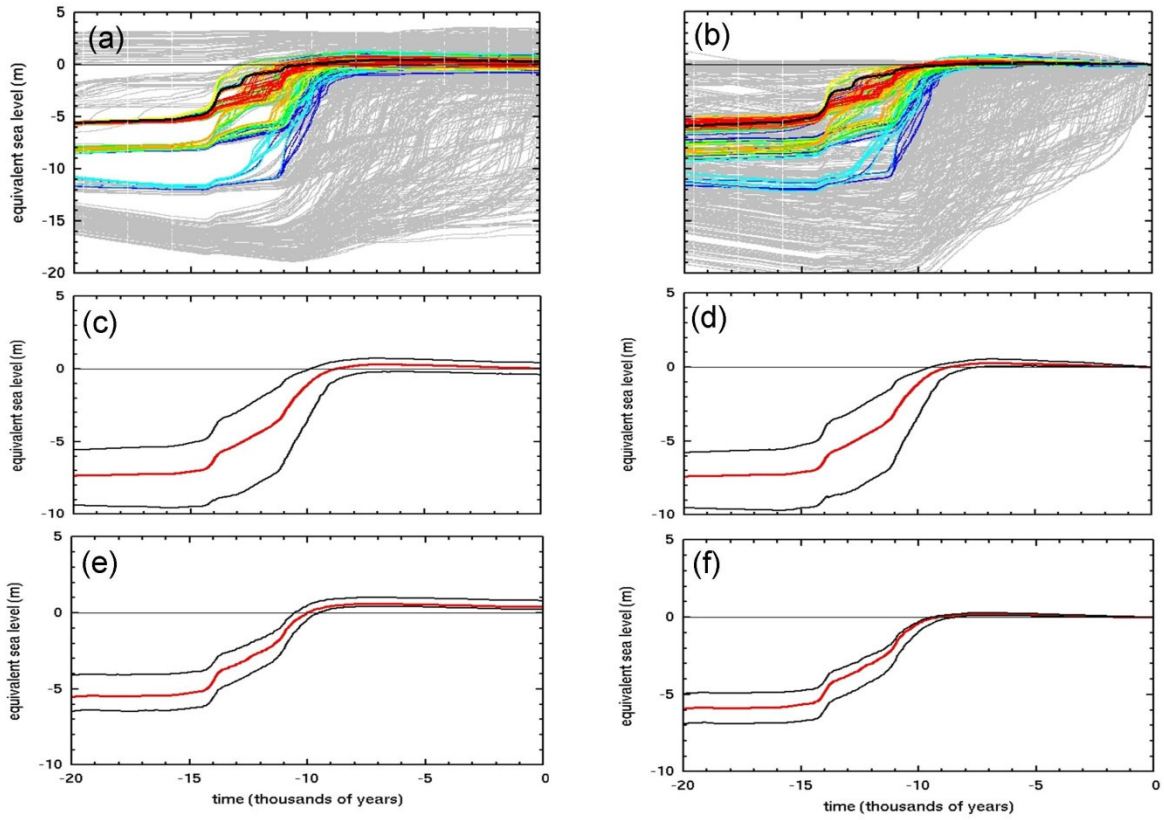

Figure 5. Equivalent global-mean sea level contribution (ESL) relative to modern vs. time. Time runs from 20,000 years before present to modern. ESL changes are calculated from the total ice amount in the domain divided by global ocean area, allowing for less contribution from ice grounded below sea level.

(a) Scatter plot of all 625 individual runs in the LE. ESL amounts are calculated relative to modern observed Antarctica, so non-zero values at time=0 imply departures from the observed ice state. Grey curves are for runs with aggregate score *S* equal to or very close to 0, and colored curves are for the 120 top-scoring runs in descending *S* order with 20 curves per color (red, orange, yellow, green, cyan, blue). The best scoring individual run is shown by a thick black curve (OCFAC=3, CALV=1, CSHELF=-5, TAUAST=3, with $S = 0.571$).

(b) As (a) but with ESL amounts relative to each run's modern value, so the curves pass exactly through zero at time=0.

(c) Score weighted curves over the whole LE, using the simple statistical method. Red curve is the score-weighted mean, i.e.,

$\Sigma\{S^{(n)}\, ESL^{(n)}(t)\} / \Sigma\{S^{(n)}\}$

where $S^{(n)}$ is the aggregate score for run n, $ESL^{(n)}(t)$ is the equivalent sea-level rise for run n at time t., and the sums are over all n (1 to 625) in the LE. Black curves show the one-sided standard deviations, i.e., the root mean square of deviations for $ESL^{(n)}$ above the mean (upper curve) or below the mean (lower curve) at each time t. $ESL^{(n)}(t)$ are relative to modern observed Antarctica, as in panel (a).

(d) As (c) but with $ESL^{(n)}(t)$ relative to each run's modern value as in (b).

(e) and **(f):** Corresponding results to (c) and (d) respectively, using the advanced statistical techniques in Chang et al. (2016) extended as described in Sect. 2.5.

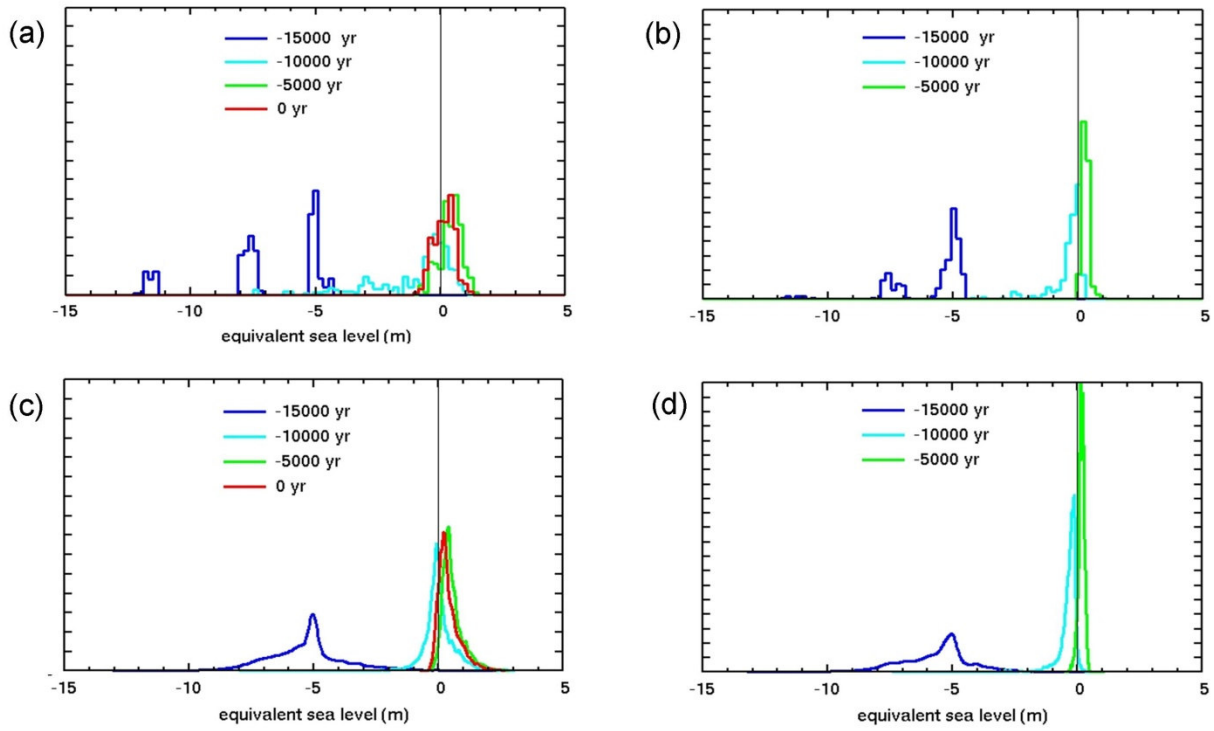

Figure 6. (a) Probability densities of equivalent sea level (ESL) rise at particular times in the LE simulations, computed with the simple averaging method. At a given time t, the density $P(E)$ is the sum of aggregate scores $S^{(n)}$ for all runs n with equivalent sea-level rise $ESL^{(n)}(t)$ within the bin E - 0.1 to E + 0.1 m, i.e., using equispaced bins 0.2 m wide. The resulting $P(E)$ are normalized so that the integral with respect to E is 1. $ESL^{(n)}(t)$ are relative to modern observed Antarctica, as in Fig. 5a.

(b) As (a) but with $ESL^{(n)}(t)$ relative to each run's modern value, as in Fig. 5b.

(c) and **(d):** Corresponding results to (a) and (b) respectively, using the advanced statistical techniques in Chang et al. (2016) extended as described in Sect. 2.5.

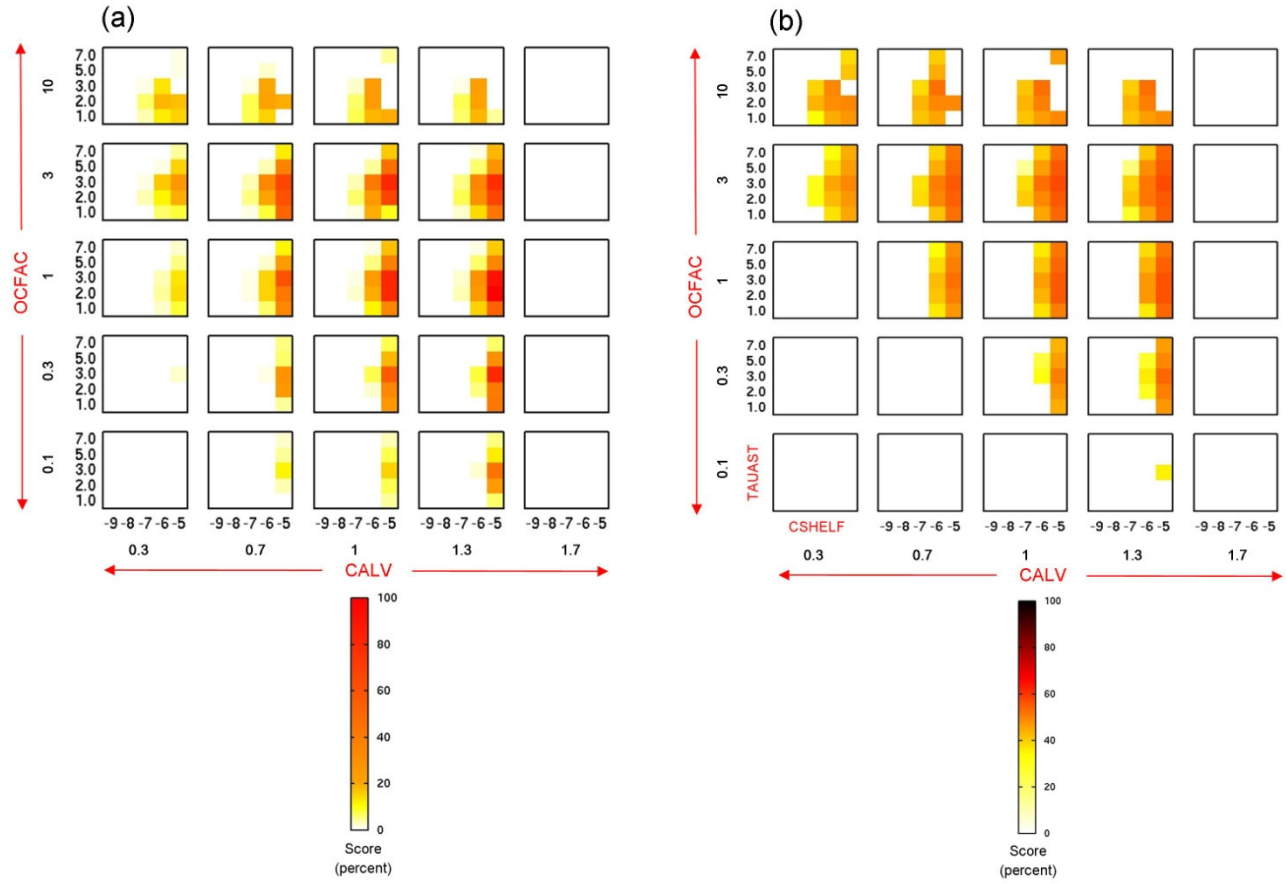

Figure C1. Aggregate scores for the complete large ensemble suite of runs (625 runs, 4 model parameters, 5 values each), used in the simple method with score-weighted averaging. The organization of the figure regarding the 4 parameter ranges is as described in Fig. 2.

(a) Using close-to-Gaussian scoring approach (a) (Sect. 2.4, Appendix B). The score values in this plot are normalized relative to the maximum score of the LE, and the color scale is adjusted to illustrate the similar qualitative distribution to (b).

(b) Using the more heuristic approach (b) (Sect. 2.4, Appendix B), exactly as in Fig. 2.

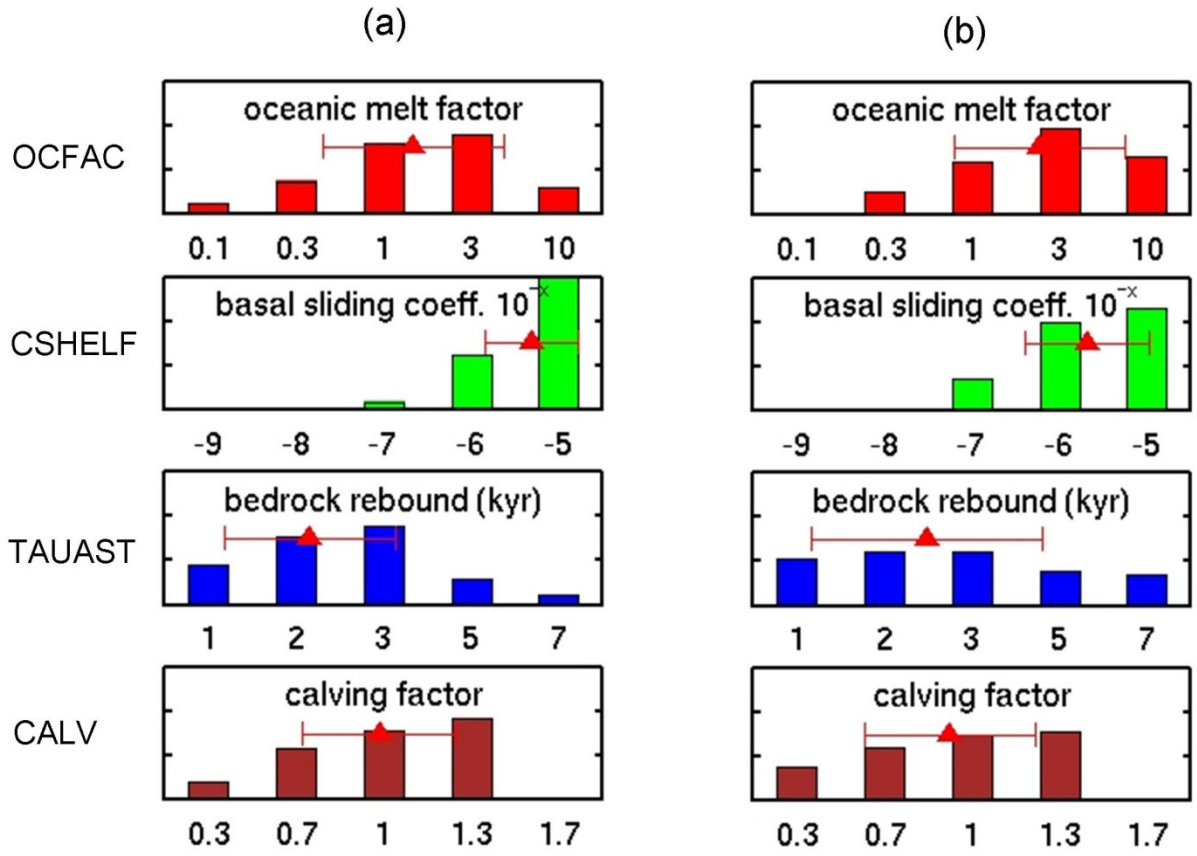

Figure C2. Ensemble-mean scores for individual parameter values, using the simple averaging method as in Fig. 3.

(a) Using close-to-Gaussian scoring approach (a) (Sect. 2.4, Appendix B).

(b) Using the more heuristic approach (b) (Sect. 2.4, Appendix B), exactly as in Fig. 3.

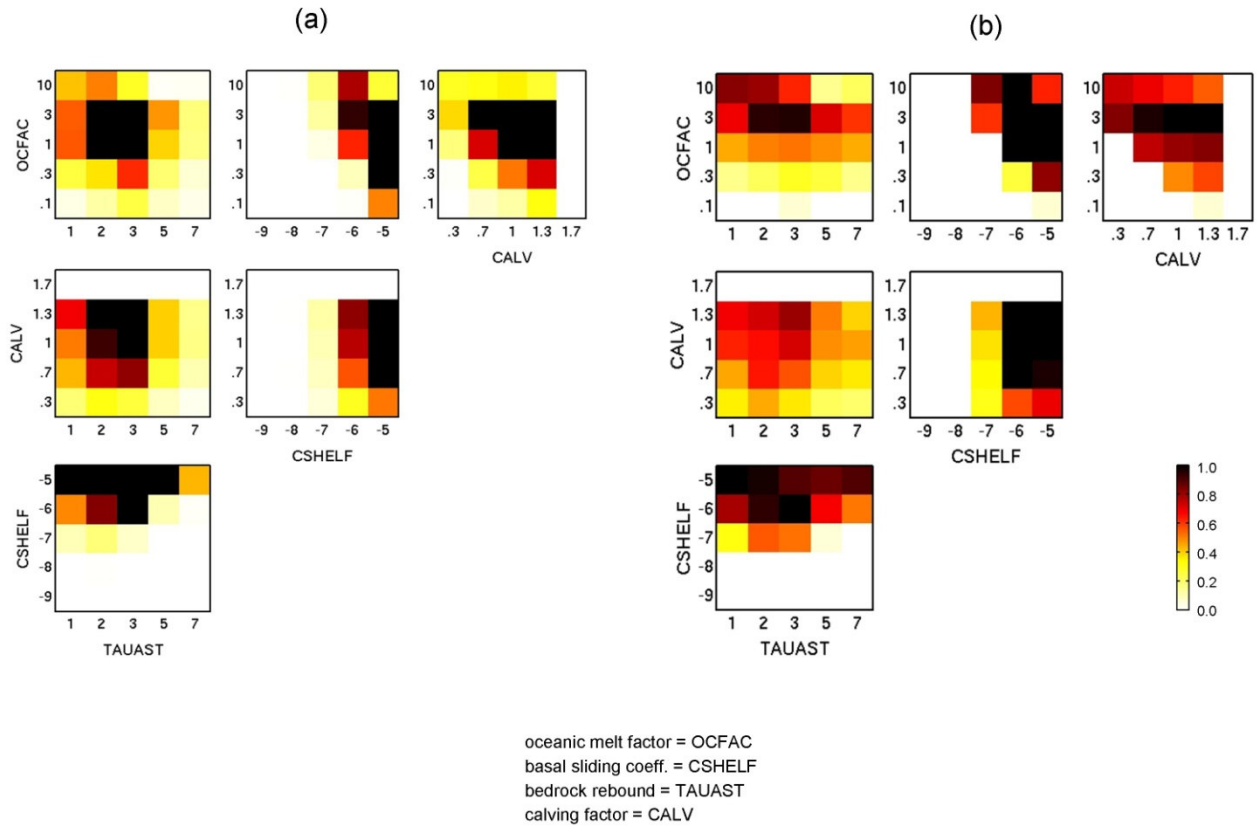

Figure C3. Ensemble-mean scores for pairs of parameters, using the simple averaging method as in Fig. 4.

(a) Using close-to-Gaussian scoring approach (a) (Sect. 2.4, Appendix B).

(b) Using the more heuristic approach (b) (Sect. 2.4, Appendix B), exactly as in Fig. 4.

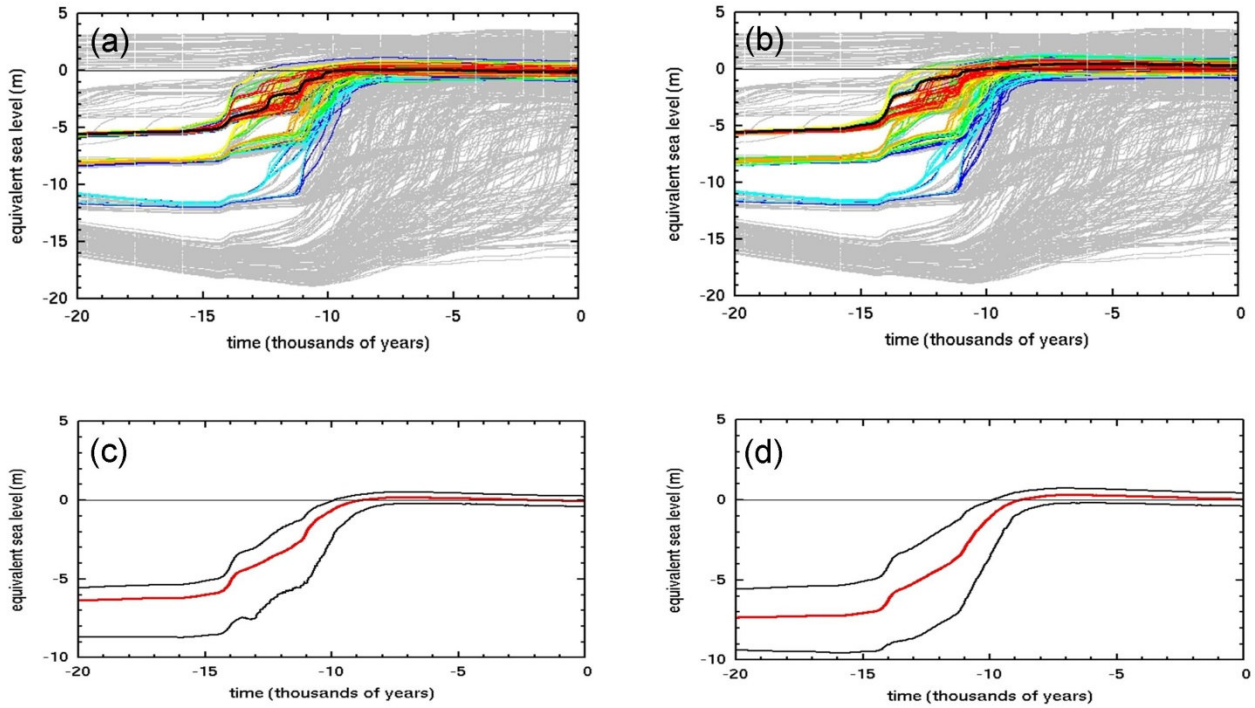

Figure C4. Equivalent global-mean sea level contribution (ESL) relative to modern vs. time as in Fig. 5.

(a) Scatter plot of all 625 individual runs in the LE, using close-to-Gaussian scoring approach (a) (Sect. 2.4, Appendix B).

(b) As (a) except using the more heuristic approach (b) (Sect. 2.4, Appendix B), exactly as in Fig. 5.

(c) Score weighted mean and one-sided standard deviations, using close-to-Gaussian scoring approach (a).

(d) As (c) except using the more heuristic approach (b), exactly as in Fig. 5.

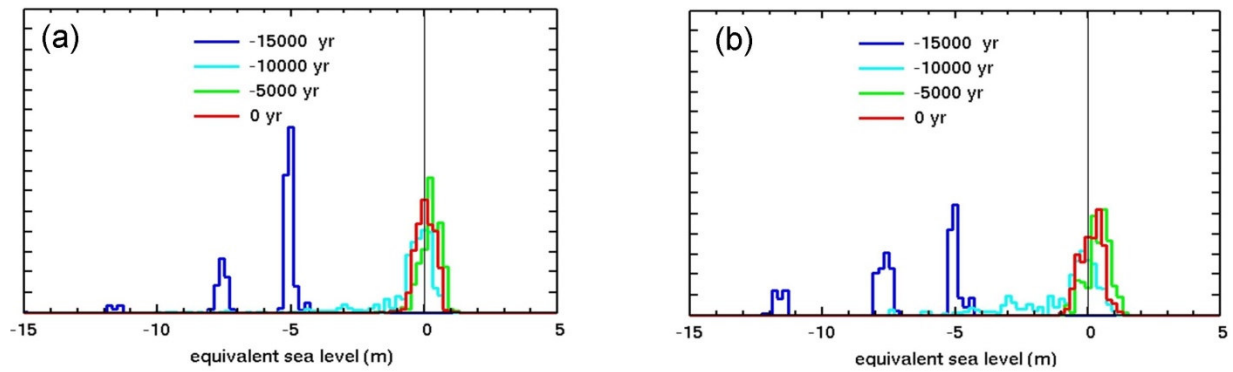

Figure C5. Probability densities of equivalent sea level (ESL) rise at particular times as in Fig. 6.

(a) Using close-to-Gaussian scoring approach (a) (Sect. 2.4, Appendix B).

(b) Using the more heuristic approach (b) (Sect. 2.4, Appendix B), exactly as in Fig. 6.

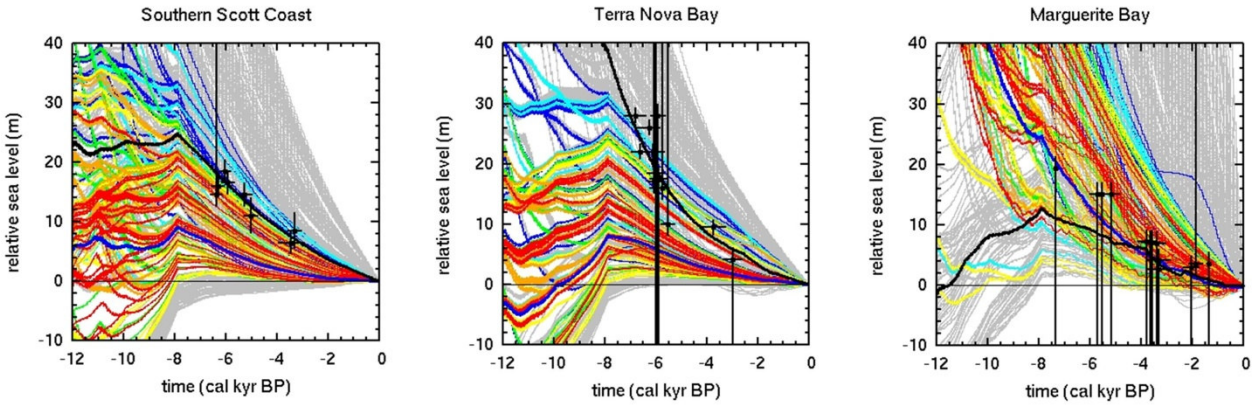

Figure D1. Model vs. observed Relative Sea Level (RSL) data, for the 3 RSL sites (Briggs and Tarasov, 2013) that lie within and away from the edges of the model's West Antarctic domain. The observations and uncertainty ranges are shown as black dots and whiskers. Model curves are shown for all 625 runs, with aggregate scores *S* indicated by colors as in Fig. 5. The run with the best individual score for each site is shown as a thick black line, and the run with best aggregate score *S* is shown as a thick blue line.

(a) Southern Scott Coast, ~77.3S, 163.6E.

(b) Terra Nova Bay, ~74.9N, 163.8E.

(c) Marguerite Bay, ~67.7S, 67.3W.

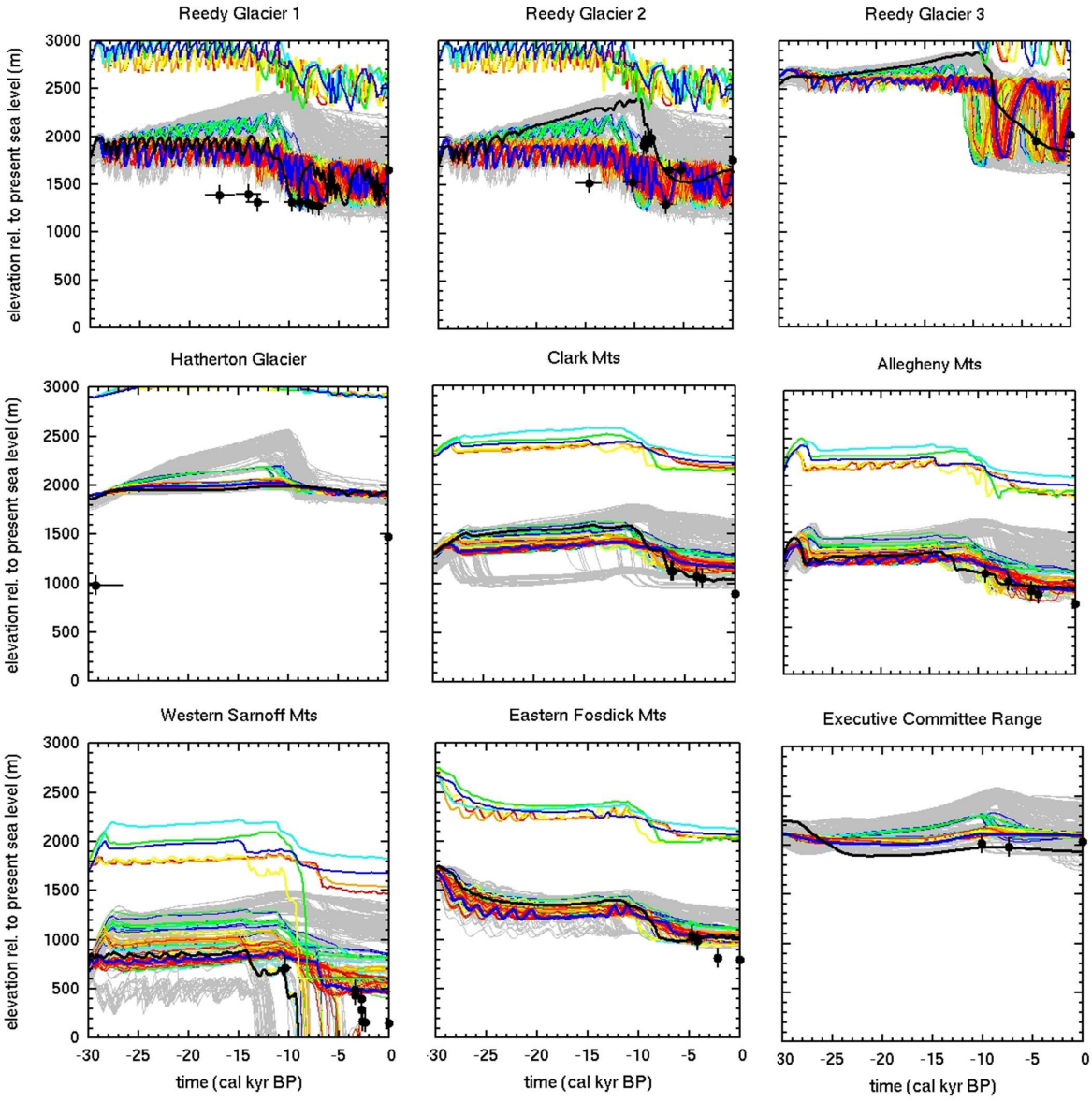
Reedy Glacier 1
Reedy Glacier 2
Reedy Glacier 3
Hatherton Glacier
Clark Mts
Allegheny Mts
Western Sarnoff Mts
Eastern Fosdick Mts
Executive Committee Range
elevation rel. to present sea level (m)
time (cal kyr BP)

Figure D2.

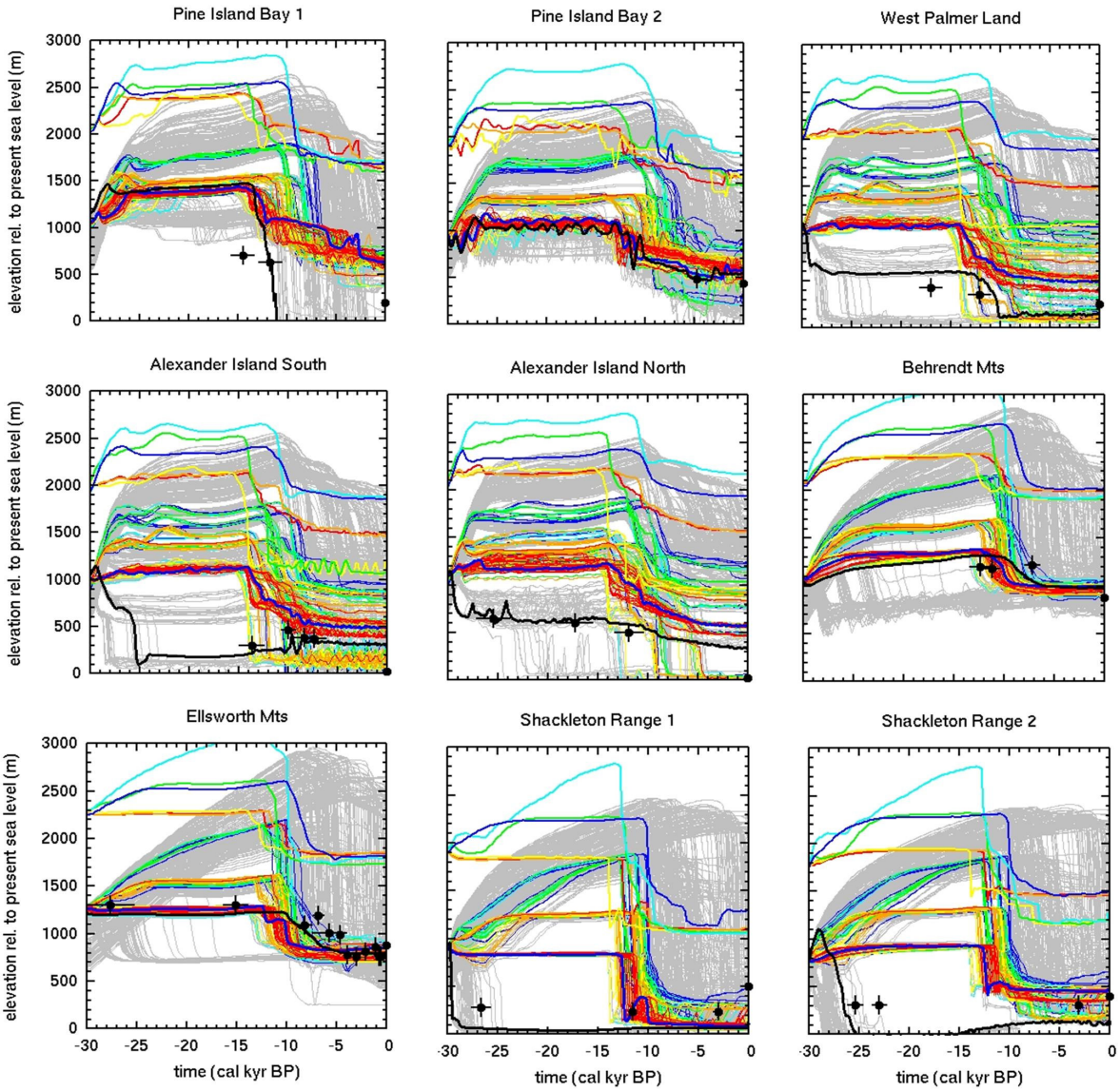

Figure D2 continued.

Figure D2 and D2 continued. Model vs. observed elevation vs. age data, for the 18 sites in the compilation (Briggs and Tarasov, 2013) that lie within and away from the edges of the model's West Antarctic domain, shown roughly in west-to-east order. The observations and uncertainty ranges are shown as black dots and whiskers. Model curves are shown for all 625 runs, with aggregate scores *S* indicated by colors as in Fig. 5. The run with the best individual score for

each site is shown as a thick black line, and the run with best aggregate score *S* is shown as a thick blue line. Sites shown (Briggs and Tarasov, 2013) are:

Reedy Glacier 1	~ 85.9S, 132.6W
Reedy Glacier 2,	~ 86.1S, 131.0W
Reedy Glacier 3,	~ 86.3S, 126.1W
Hatherton glacier	~ 79.9S, 156.8E
Clark Mts,	~ 77.3S, 142.1W
Allegheny Mts,	~ 77.3S, 143.3W
Western Sarnoff Mts,	~ 77.1S, 145.5W
Eastern Fosdick Mts,	~ 76.5S, 144.5W
Executive Committee Range,	~ 77.2S, 127.1W
Pine Island Bay 1,	~ 75.2S, 111.2W
Pine Island Bay 2,	~ 74.5S, 99.2W
West Palmer Land,	~ 71.6S, 67.4W
Alexander Island South,	~ 72.0S, 68.5W
Alexander Island North,	~ 70.9S, 68.4W
Behrendt Mts,	~ 75.3S, 72.3W
Ellsworth Mts,	~ 80.3S, 82.2W
Shackleton Range 1,	~ 80.4S, 30.1W
Shackleton Range 2,	~ 80.1S, 25.8W

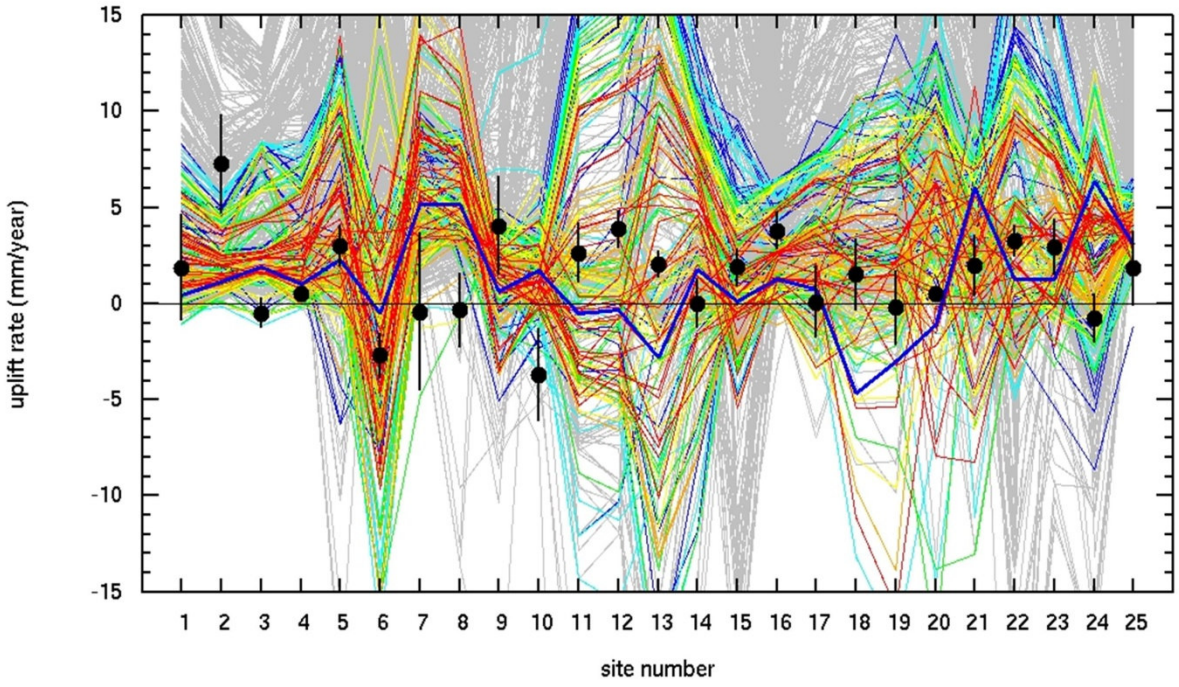

Figure D3. Model vs. observed modern uplift rates, for the 25 sites in the compilation (Whitehouse et al., 2012b) that lie within the model's West Antarctic domain, shown roughly in west-to-east order. The observations and uncertainty ranges are shown as black dots and whiskers. Model rates are shown for all 625 runs, with straight lines joining the sites, and aggregate scores *S* indicated by colors as in Fig. 5. The run with best aggregate score *S* is shown as a thick blue line. Sites shown, with labels as in Whitehouse et al. (2012b, Supp. Inf.), are:

1. FTP1, 78.93S, 162.57E
2. ROB1, 77.03S, 163.19E
3. TNB1, 74.70S, 164.10E
4. MCM4_AV, 77.85S, 166.76E
5. MBL1_AV, 78.03S, 155.02W
6. W01_AV, 87.42S, 149.43W
7. MBL2, 76.32S, 144.30W
8. MBL3, 77.34S, 141.87W
9. W09, 82.68S, 104.39W

10. W06A, 79.63S, 91.28W
11. W07_AV, 80.32S, 81.43W
12. W05_AV, 80.04S, 80.56W
13. HAAG, 77.04S, 78.29W
14. W08A/B, 75.28S, 72.18W
15. W02_AV, 85.61S, 68.55W
16. OHIG, 63.32S, 57.90W
17. PALM, 64.78S, 64.05W
18. ROTB, 67.57S, 68.13W
19. SMRT, 68.12S, 67.10W
20. FOS1, 71.31S, 68.32W
21. BREN, 72.67S, 63.03W
22. W04_AV, 82.86S, 53.20W
23. BELG, 77.86S, 34.62W
24. W03_AV, 81.58S, 28.40W
25. SVEA, 74.58S, 11.22W

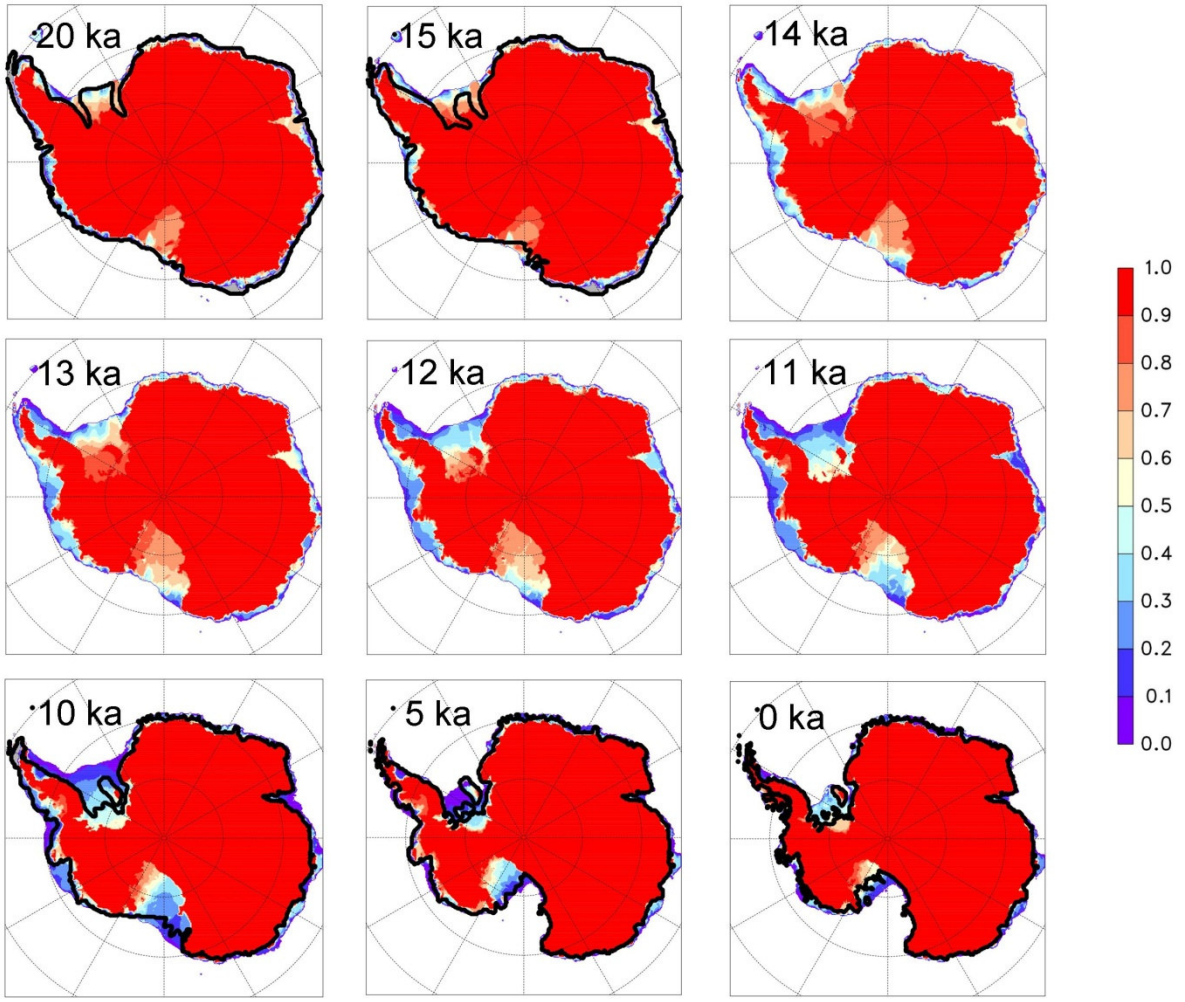

Figure D4. Score-weighted probability (0 to 1) of grounded ice vs. floating ice or open ocean at each grid point (see text), for various times over the last 20,000 years, concentrating on the period of rapid retreat between 15 and 10 ka. The LE and model version is essentially the same as above, except with all-Antarctic coverage to include East Antarctic variations. The quantity shown is the sum of scores $S(n)$ for runs n with grounded ice at each grid point and time, divided by the sum of scores for all runs in the ensemble. Thick black lines in the panels for 20, 15, 10 and 5 ka show grounding lines reconstructed for West Antarctica by the RAISED consortium (RAISED, 2014), plotted using their vertex information (S. Jamieson, pers. comm.), and choosing their Scenario A for the Weddell embayment (Hillenbrand et al., 2014). For 20 and 15 ka around East Antarctica, the black line is from the 20 ka RAISED timeslice which for EAIS is based on Livingston et al. (2012) and Mackintosh et al. (2014). Similarly the modern

grounding line (Fretwell et al., 2013) is shown by a thick black line for 0 ka, which is also used around East Antarctica for 10 and 5 ka.

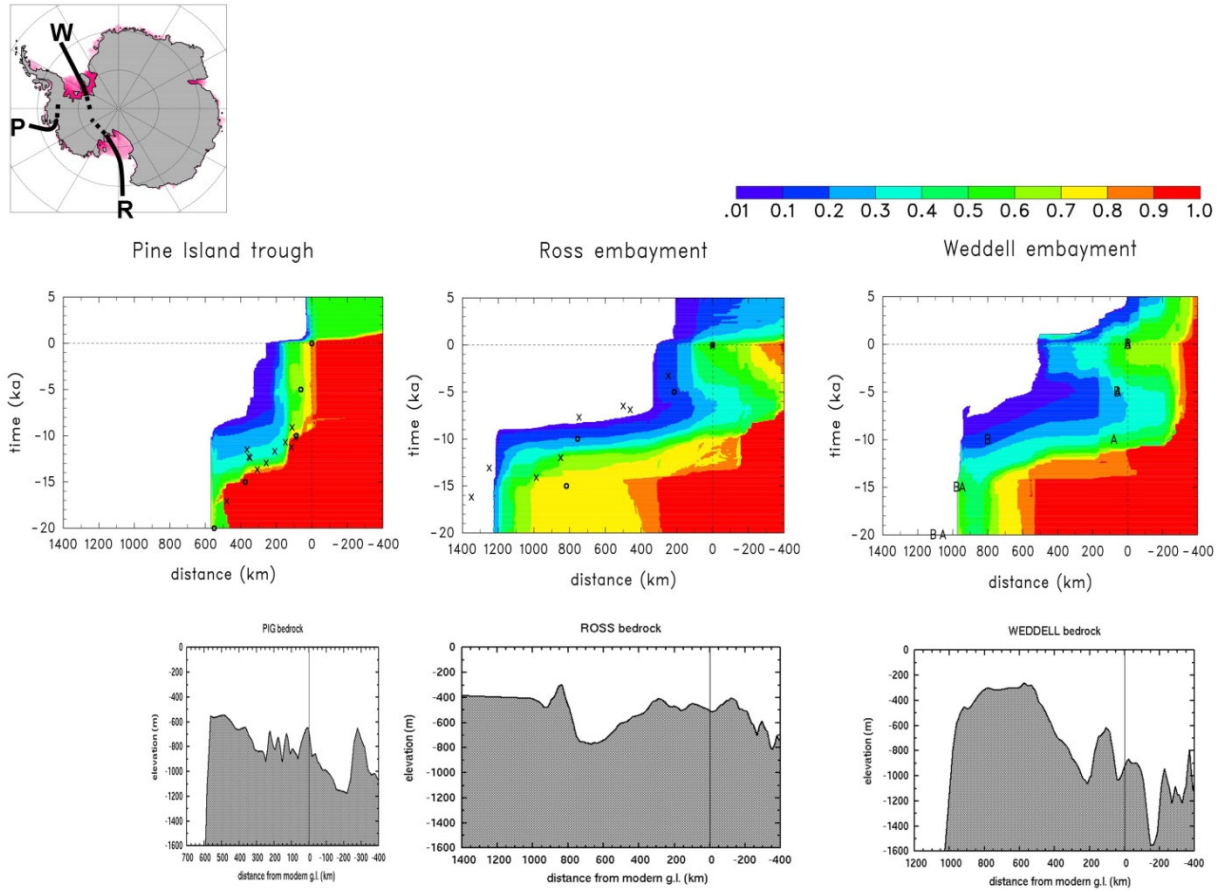

Figure D5. Upper panels: Score-weighted probability (0 to 1) of grounded ice vs. time, as in Fig. D4 but along centerline transects of (i) Pine Island Glacier and its paleo-trough, (ii) Ross embayment and (iii) Weddell embayment. Black symbols show various published data:

Pine Island, circles: Larter et al., 2014 (the RAISED Consortium).

Pine Island, crosses: Kirshner et al., 2012; Hillenbrand et al., 2013; Smith et al., 2014.

Ross, circles: Anderson et al., 2014 (the RAISED Consortium).

Ross, crosses: Conway et al., 1999; McKay et al., 2008.

Weddell, 'A' and 'B': Hillenbrand et al., 2014 (the RAISED Consortium), Scenarios A and B respectively.

Lower panels: Modern bathymetric profiles along each transect (from Bedmap2; Fretwell et al. 2013).