PNAS
www.pnas.org

Supplementary Information for

Ice sheet contributions to future sea level rise from structured expert judgement

by J. L. Bamber, M Oppenheimer, R. E. Kopp, W.P. Aspinall, R.M. Cooke

Corresponding author: J.L. Bamber
Email: j.bamber@bristol.ac.uk

This PDF file includes:

Supplementary Notes 1-9
Figs. S1 to S12
Tables S1 to S6
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Expert responses, introductory and preparatory materials, available at https://doi.org/10.5523/bris.10.5523/bris.23k1jbtan6sjv2huakf63cqgav

Please note that the results given in this Supplementary Information do not incorporate the GMSL contribution from continuation of the 2000-2010 baseline rates that were provided to the elicitation workshops. See Supplementary note 5.

www.pnas.org/cgi/doi/10.1073/pnas.1817205116

Supplementary note 1: Determining expert quantiles

Thirteen experts participated in the expert elicitation on contribution to sea level rise from ice sheets, held at RFF, Washington DC, USA on Jan 25-26. Nine experts participated in a similar elicitation held near London, UK on Feb 20-21, 2018. The two elicitations used the same elicitation protocol. The assessments concerned Accumulation, Runoff and Discharge for GrIS, WAIS and EAIS for the time × temperature scenarios shown in Figure S1. Experts were chosen based on whether they were research active in the topic, assessed on their publications over the last ~5 years and involvement in related initiatives such as NASA SeaRise, Delta Commission (Netherlands Govt), EU Ice2Sea project, Ice Sheet MIPS etc. A working minimum group size, from previous experience, is about six experts and more than 20 provides diminishing returns in terms of the performance of the synthetic pooled expert. We also wished to obtain a balance in age, gender and specialism within the broad field of ice sheets and SLR and to avoid accessing multiple experts from the same group. In addition to the 22 that participated, nine experts were invited who could not attend (4) or did not wish to (5).

The participating experts are listed below

US elicitation	**EU elicitation**
Robert Bindschadler	Gaël Durand
Rob DeConto	Johannes Fuerst
Natalya Gomez	Hilmar Gudmundsson
Ian Howat	Anders Levermann
Ian Joughin	Frank Pattyn
Shawn Marshall	Catherine Ritz
Sophie Nowicki	Ingo Sasgen
Stephen Price	Aimee Slangen
Eric Rignot	Bert Wouters
Ted Scambos	
Christian Schoof	
Helene Seroussi	
Ryan Walker	

The assessments were combined using equal weighting and performance-based weighting[1]. The combined assessments were convolved to obtain the overall ice sheet contribution to global sea level rise using dependence information provided by the experts.

[1] In the EU expert panel, one expert provided judgments based on a conceptual interpretation of the three processes, Accumulation, Runoff and Discharge, that differed significantly from the definitional framework outlined in the questionnaire; the expert acknowledged this to be the case upon enquiry, and their judgments were not included in subsequent processing. In the US panel one expert misinterpreted the baseline values, as a result, their uncertainty judgments contained systematic discrepancies in relation to others in the panel. Unfortunately, there was not an opportunity to re-visit and correct this expert's evaluations in a timely manner, and so the relevant inputs were removed from the analysis reported here.

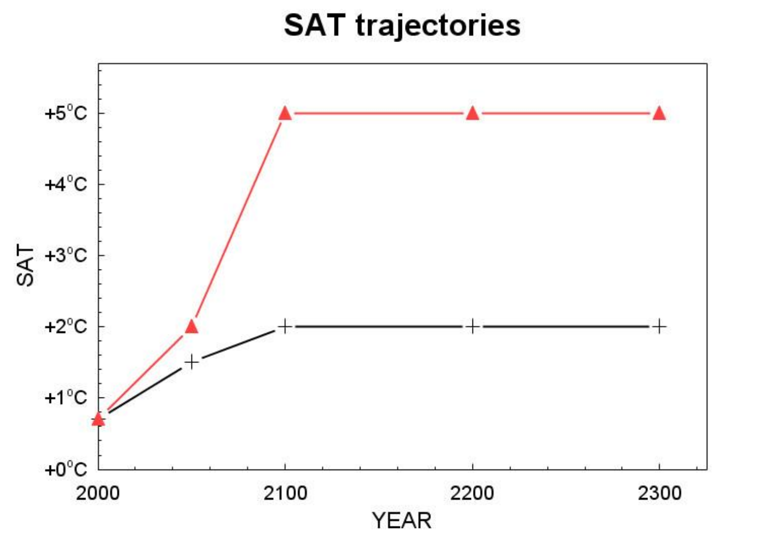


Figure S1. The two temperature scenarios prescribed: L (+2° C) and H (+5° C).

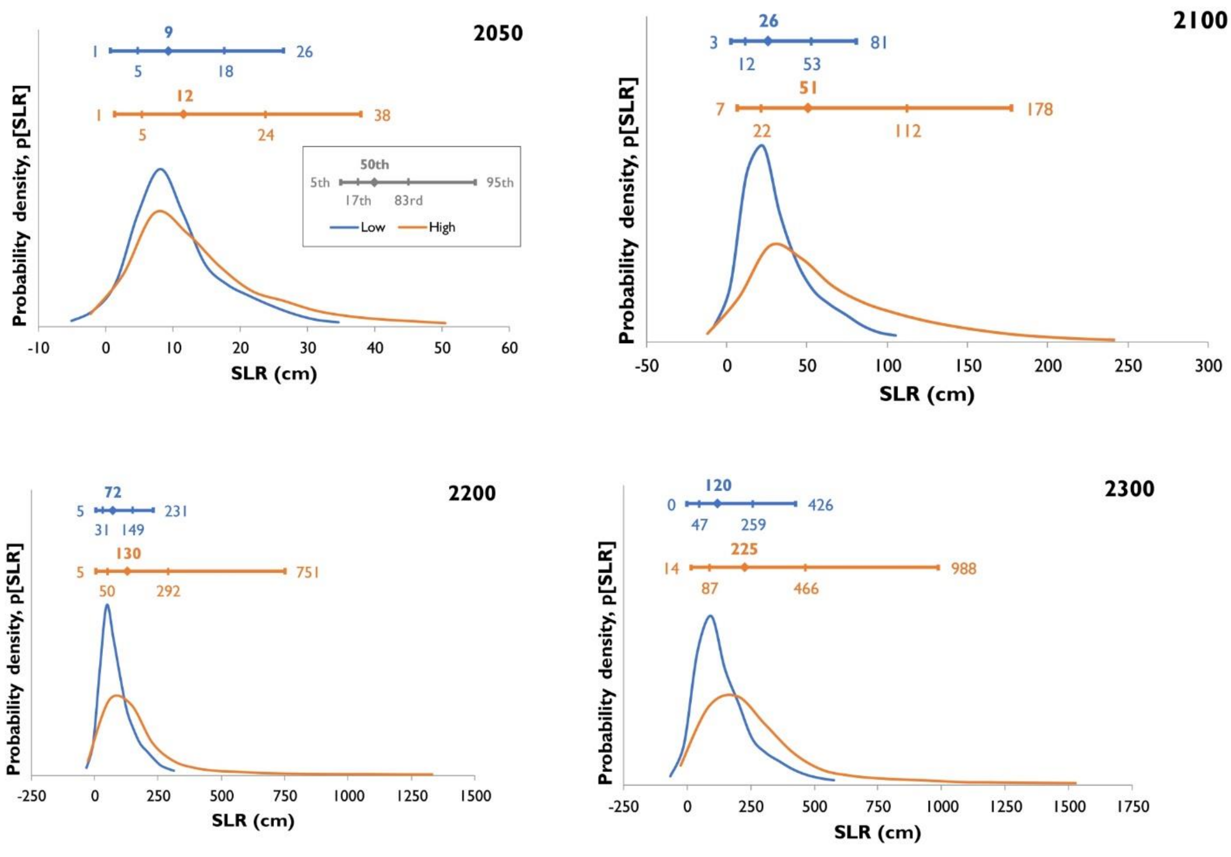


Figure S2. Probability density functions (PDFs) for the L (blue) and H (red) temperature scenarios for the combined ice sheet contribution at a) 2050; b) 2100; c) 2200; d) 2300. The horizontal bars show the 5th,17th, 50th (median), 83rd and 95th percentile values. The baseline rate of 0.76 mm a^{-1} is included. Note change of x-axis scales. (See Supplementary note 5).

1.1 Overall Results

In this exercise experts quantified their 5th, 50th and 95th percentiles for accumulation, for discharge and for runoff for each of GrIS, WAIS and EAIS as anomalies from the 2000-2010 baseline trend (see Supplementary note 5).

They also quantified their dependence between these quantities at 2100 with 5˚ C warming with respect to pre-industrial. This same dependence structure was applied for all other scenarios. As an extension, more articulated dependence structures could be elicited for the different scenarios and applied to the present assessments. In the terminology of SEJ, a Decision Maker (DM) is a "synthetic pooled expert" that is some weighted combination of experts. Equal Weights (EW) is sometime referred to as "one person one vote". Performance Weighting (PW) is where experts are weighted based on measures of their informative and accuracy quantified using a set of calibration questions or items (described in greater detail in SI Note 1.2).

The results with Performance Weighting (PW) are shown in Table S1 in yellow. For the final results, it was decided to use the performance weighted combination of all experts whose statistical accuracy (P-value) was greater than 0.01 (PW01). EW denotes Equal Weighted combinations.

Total ice-sheet SLR is the sum of SLR from all three ice sheets: however, this is a sum of stochastic variables. For 2300H the total mean of 287 cm is the sum of 63 cm, 113 cm and 111 cm, but the quantiles do not sum in this way. For 2300H, the total 95th percentile, 966cm, is smaller than 498 cm + 332 cm + 378 cm = 1208 cm. Adding stochastic variables requires knowledge of their joint distribution. The quantiles will add only if the variables are completely rank dependent (sometimes called co-monotonic). In this case one variable is at or above its 95th percentile if and only if the others are as well. The chance of that happening is then 5%, which means that the sum of the 95th percentiles is exceeded with probability 5%. If the variables are independent, then the chance that all three are at or above their respective 95th percentiles is $0.05^3 = 0.000125$. In this case the 95th percentile will be much lower than the sum of the separate 95th percentiles. In fact, if the three ice sheets are independent the 95th percentile of PW01 (Figure S9b) is 823 cm. The difference 966 cm − 823 cm reflects the effect of the dependence.

The choice of a cutoff for statistical accuracy (P-value) beneath which experts are unweighted is imposed by the theory of strictly proper scoring rules (see Supplementary Information section 1.2). The scoring rule theory does not say what this cutoff should be, only that there should be some positive lower bound to the admissible statistical accuracy scores. Optimal performance weighting (PWOpt) chooses a cutoff which optimizes the scores of the resulting combination. PW01 reflects the choice to include all those experts who have acceptable statistical accuracy so as to ensure wider representation. The distributions of PW01 are somewhat wider than those of PWOpt. With the optimal cutoff of 0.399, only experts 3 and 14 are weighted. Cutoff = 0.01 forms a weighted combination of eight experts whose statistical accuracy is above 0.01; these are experts 3,6,8,9,12,14,24 and 27. EW01 forms an equal weighted combination of these same eight experts. All combinations concern the experts' joint distributions based on the elicited dependence information.

Ice Sheets contribution to SLR [cm]; eight US & EU experts with P-value > 0.01, All DMs; 50k samples												
2300H	Mean	StDev	5%	50%	95%		2300L	Mean	StDev	5%	50%	95%

eu_pwopt	358	381	-10	245	982		eu_pwopt	154	167	-18	107	450
eu_pwopt_e	70	159	-23	4	268		eu_pwopt_e	4	30	-17	0	25
eu_pwopt_g	118	131	-19	71	367		eu_pwopt_g	53	65	-10	33	165
eu_pwopt_w	170	136	-7	146	408		eu_pwopt_w	98	94	-8	73	291
useu_ew01	316	346	-3	214	1070		useu_ew01	150	151	-13	105	448
useu_ew_e	88	210	-44	14	632		useu_ew_e	22	55	-25	2	169
useu_ew_g	102	111	-10	73	325		useu_ew_g	53	66	-8	37	199
useu_ew_w	127	137	-28	90	400		useu_ew_w	76	89	-35	50	258
useu_pw01	287	322	-9	202	966		useu_pw01	132	137	-22	97	403
useu_pw01_e	63	195	-53	10	498		useu_pw01_e	10	41	-29	3	96
useu_pw01_g	113	117	-10	81	332		useu_pw01_g	61	75	-10	38	220
useu_pw01_w	111	136	-42	77	378		useu_pw01_w	61	88	-53	38	242
us_ew01	312	345	3	209	1130		us_ew01	152	150	-11	105	451
us_pw01	278	318	-9	196	973		us_pw01	129	133	-23	95	395
us_pwopt	265	276	-15	211	750		us_pwopt	133	133	-32	109	394
us_pwopt_e	35	162	-56	8	105		us_pwopt_e	9	39	-30	4	56
us_pwopt_g	123	123	-15	89	339		us_pwopt_g	69	84	-12	38	236
us_pwopt_w	107	134	-53	85	355		us_pwopt_w	55	87	-64	44	209
2200H	Mean	StDev	5%	50%	95%		2200L	Mean	StDev	5%	50%	95%
eu_pwopt	220	270	-16	142	610		eu_pwopt	80	87	-16	58	237
eu_pwopt_e	43	122	-11	3	134		eu_pwopt_e	1	12	-13	0	17
eu_pwopt_g	59	61	-11	44	167		eu_pwopt_g	30	38	-7	20	94
eu_pwopt_w	117	117	-8	82	321		eu_pwopt_w	49	49	-7	38	144
useu_ew01	217	278	-6	129	844		useu_ew01	82	78	-6	61	225
useu_ew_e	66	169	-23	6	517		useu_ew_e	8	21	-14	1	58
useu_ew_g	59	61	-6	44	184		useu_ew_g	33	41	-5	23	126
useu_ew_w	91	119	-21	57	367		useu_ew_w	41	46	-20	28	133
useu_pw01	189	260	-10	115	735		useu_pw01	74	72	-10	57	216
useu_pw01_e	48	158	-29	6	398		useu_pw01_e	4	15	-15	2	34
useu_pw01_g	66	69	-8	44	205		useu_pw01_g	38	47	-6	23	138
useu_pw01_w	76	113	-29	47	320		useu_pw01_w	33	45	-28	22	124
us_ew01	216	281	-3	125	904		us_ew01	82	75	-5	61	222
us_pw01	184	258	-10	112	769		us_pw01	73	69	-9	56	212
us_pwopt	164	209	-12	118	505		us_pwopt	78	71	-12	64	220
us_pwopt_e	28	127	-30	6	45		us_pwopt_e	4	14	-15	4	24
us_pwopt_g	74	77	-11	46	216		us_pwopt_g	45	53	-7	24	149

us_pwopt_w	63	97	-36	48	191		us_pwopt_w	29	44	-33	24	105
2100H	Mean	StDev	5%	50%	95%		2100L	Mean	StDev	5%	50%	95%
eu_pwopt	65	70	-5	42	198		eu_pwopt	17	26	-10	11	60
eu_pwopt_e	10	18	-5	2	44		eu_pwopt_e	0	5	-6	0	7
eu_pwopt_g	20	20	-4	15	53		eu_pwopt_g	10	11	-3	7	29
eu_pwopt_w	35	38	-4	23	106		eu_pwopt_w	7	14	-4	3	25
useu_ew01	63	60	-1	46	179		useu_ew01	25	27	-5	17	75
useu_ew_e	10	20	-9	2	53		useu_ew_e	2	7	-7	1	15
useu_ew_g	23	25	-3	18	80		useu_ew_g	11	13	-3	8	42
useu_ew_w	30	35	-5	20	96		useu_ew_w	12	17	-5	6	44
useu_pw01	59	56	-1	43	170		useu_pw01	24	25	-5	18	73
useu_pw01_e	6	17	-11	2	46		useu_pw01_e	1	6	-8	0	12
useu_pw01_g	27	30	-3	17	93		useu_pw01_g	13	16	-3	7	51
useu_pw01_w	25	33	-7	16	91		useu_pw01_w	11	16	-5	6	42
us_ew01	63	57	1	48	171		us_ew01	28	26	-3	21	77
us_pw01	59	54	-1	44	166		us_pw01	25	25	-4	19	74
us_pwopt	60	52	0	48	157		us_pwopt	27	25	-5	21	75
us_pwopt_e	4	14	-12	3	21		us_pwopt_e	0	5	-8	1	9
us_pwopt_g	33	34	-4	18	100		us_pwopt_g	15	19	-4	7	57
us_pwopt_w	23	31	-9	18	68		us_pwopt_w	11	16	-5	8	34
2050H	Mean	StDev	5%	50%	95%		2050L	Mean	StDev	5%	50%	95%
eu_pwopt	7	12	-4	5	24		eu_pwopt	6	9	-4	4	21
eu_pwopt_e	0	3	-4	0	5		eu_pwopt_e	0	2	-4	0	4
eu_pwopt_g	4	5	-1	2	12		eu_pwopt_g	3	4	-2	2	11
eu_pwopt_w	3	5	-1	2	11		eu_pwopt_w	3	5	-1	1	9
useu_ew01	12	14	-2	7	41		useu_ew01	8	9	-3	5	25
useu_ew_e	1	5	-5	0	11		useu_ew_e	0	3	-4	0	5
useu_ew_g	5	6	-1	3	19		useu_ew_g	3	4	-1	2	12
useu_ew_w	6	8	-1	3	24		useu_ew_w	4	6	-2	2	17
useu_pw01	11	12	-3	8	34		useu_pw01	7	8	-3	6	23
useu_pw01_e	0	4	-6	0	7		useu_pw01_e	0	2	-4	0	4
useu_pw01_g	6	7	-1	3	24		useu_pw01_g	4	5	-1	2	15
useu_pw01_w	5	7	-2	3	21		useu_pw01_w	3	5	-2	2	16
us_ew01	13	14	-2	9	42		us_ew01	9	9	-3	6	25
us_pw01	11	11	-2	8	34		us_pw01	7	8	-3	6	23
us_pwopt	12	11	-3	9	34		us_pwopt	8	8	-3	6	22
us_pwopt_e	0	4	-7	0	5		us_pwopt_e	0	2	-4	0	3

us_pwopt_g	7	9	-1	3	26		us_pwopt_g	4	5	-2	2	17
us_pwopt_w	5	383	-2	4	14		us_pwopt_w	3	238	-2	3	10

Table S1. Combined results for ice sheet contribution to SLR [cm], with dependence for the EU and US elicitations. PW denotes Performance Weighted combination, EW denotes Equal Weighted combination and "Opt" indicates that the cut-off P-value is chosen to optimize the combined score of the resulting combination. The designation "01" indicates that all experts with p-value greater or equal to 0.01 are weighted. "eu" denotes European experts only, "us" denotes US experts only, "useu" denotes both European and US experts, jointly. "e" denotes East Antarctica, "w" denotes West Antarctica, "g" denotes Greenland. Without "e", "w" or "g" denotes sea level rise contribution computed expert-wise as the stochastic sum of "e", "w" and "g" using the elicited dependence information, combined with equal or performance weighting. (See Supplementary note 5 for quantitative definitions of the elicited ice sheet contributions).

Figures S3 and S4 show the SLR medians and 90% confidence ranges for the eight experts with P-value > 0.01 and for the combinations EW01, PW01 and PWOpt.

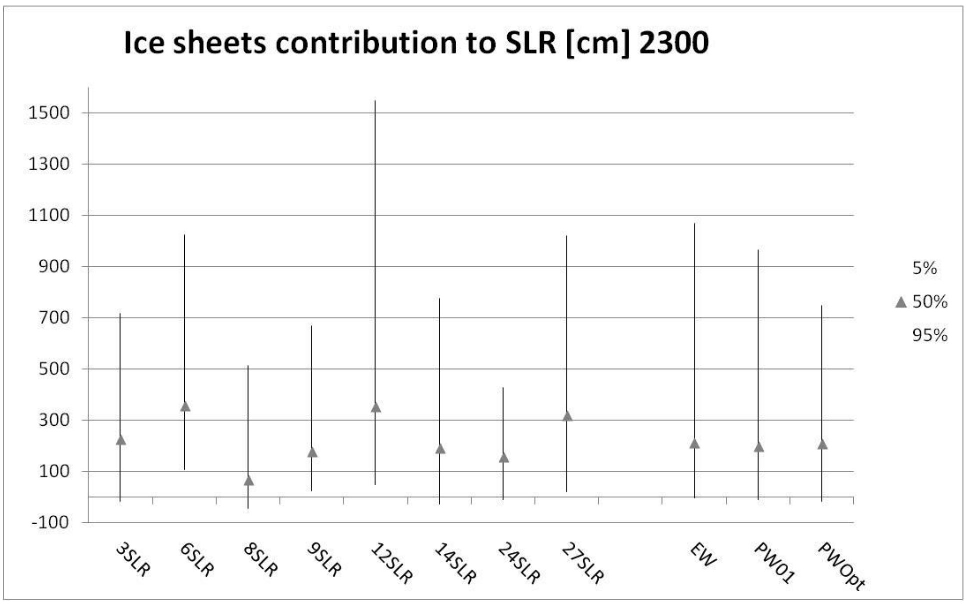


Figure S3: Experts with P-value > 0.01 and combinations: EW01 equal weight combination of all experts with P-value > 0.01; PW01 performance-based combination of all experts with P-value > 0.01, PWOpt, performance-based combinations of experts with P-values > 0.40, for ice sheet contribution to SLR [cm] in 2300, with 5° C warming.

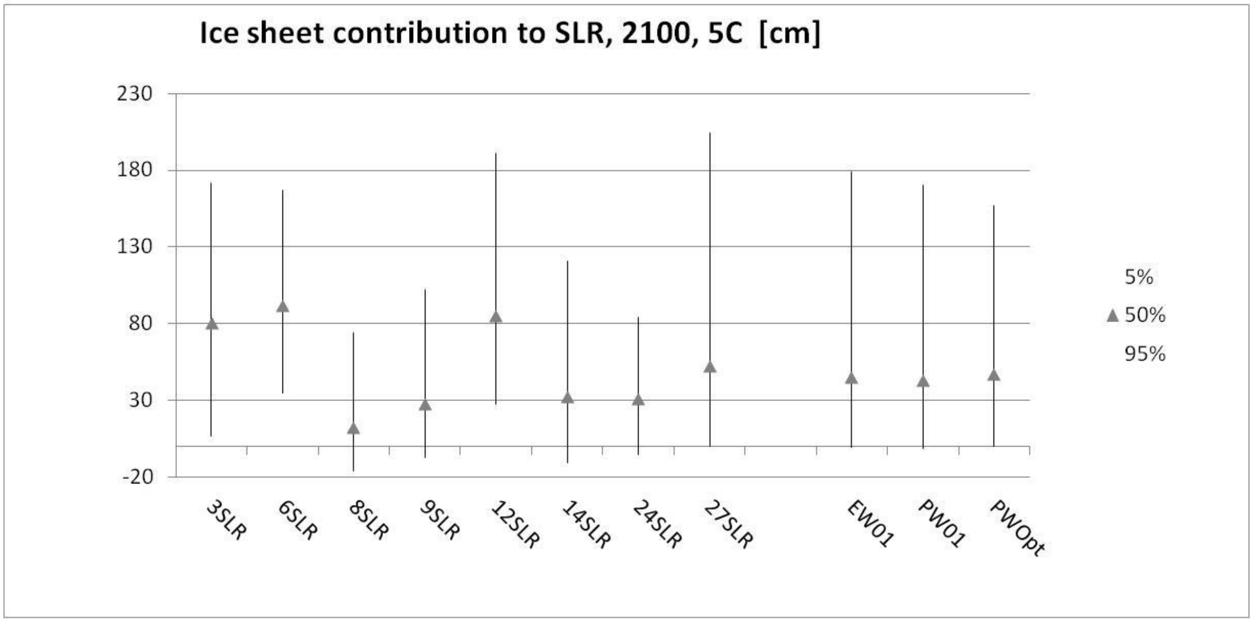


Figure S4: Experts with P-value > 0.01 and combinations: EW01 equal weight combination of experts with P-value > 0.01; PW01 performance-based combination of experts with P-value > 0.01, PWOpt, performance-based combinations of experts with P-values > 0.399, for ice sheet contribution to SLR [cm] in 2300, with 5° C warming.

From Figures S3 and S4 it is apparent the medians are not midway between the lower and upper quantiles but are closer to the lower values. This suggests a heavy right tail in the PDF: probability dies off more slowly above the median than below. Kurtosis is a common measure of "tailedness" with values above 3 being "above normal". Figure S5 shows the kurtosis of the three ice sheets over time for the 2C and 5C stabilization scenarios.

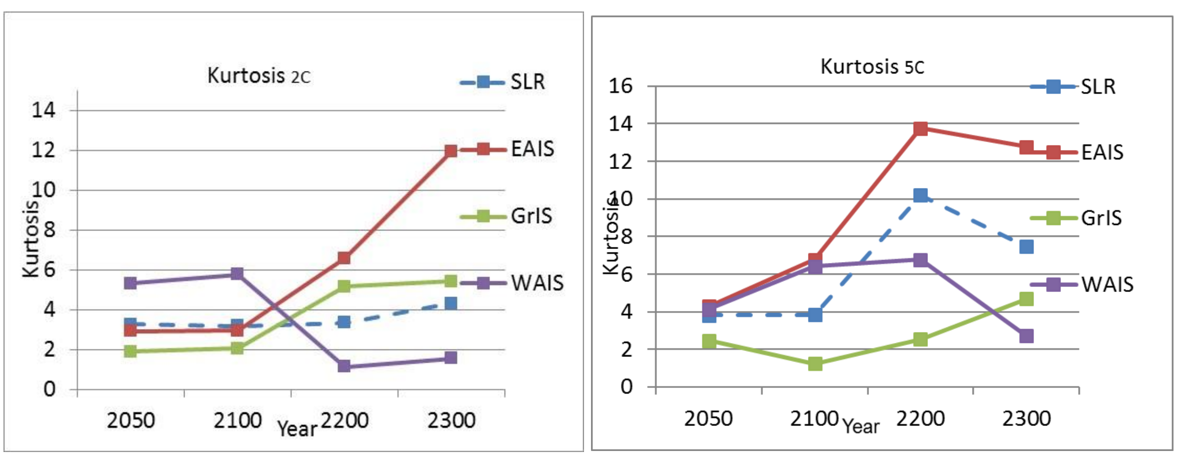


Figure S5: Kurtosis paths for the +2° C and +5° C scenarios.

In the +5° C scenario the tailedness in Antarctica peaks at 2200. In the +2° C scenario EAIS's tailedness peaks in 2300; WAIS is associated with some instability up to 2100 but little thereafter. This implies that experts believe that a substantial dynamic response is unlikely in the next century for the L scenario. This interpretation is supported by the rationale plots shown in Figures S9 and S10.

1.2 Expert Scoring

The expert judgment methodology applied here is termed the "*Classical Model*" because of its analogy to classical hypothesis testing (1). The key idea is that experts are treated as statistical hypotheses. Experts were given a PowerPoint presentation to explain the basic features of the method (see SI Section 8), on which this section is based. Expert scoring is shown in Table S2. For detailed explanations please refer to (2), especially the online supplementary material (Appendix A).

An expert's *statistical accuracy* is the P-value (column 2 in Table S2) at which we would falsely reject the hypothesis that an expert's probability assessments are statistically accurate. Roughly, an expert is statistically accurate if, in a statistical sense, 5% of the realizations fall beneath his/her 5^{th} percentile, 45% of the realizations fall between the 5^{th} and 50^{th} percentile, etc. High values (near 1) are good, low values (near 0) reflect low statistical accuracy. An expert's *informativenes*s is measured as the Shannon relative information in the expert's distribution relative to a uniform background measure over an interval containing all experts' percentile assessments and the realizations, variable-wise. Columns 3 and 4 give the average information scores for each expert for all variables (column 3) and all calibration variables (column 4). The number of calibration variables is shown in column 5 for each expert (in this case all experts assessed all 16 calibration variables). The product of columns 2 and 4 is the *combined score* for each expert. Note that statistical accuracy scores vary over seven orders of magnitude whereas information scores vary within a factor three. Therefore, by design, the ratios of the products of combined scores are dominated by the statistical accuracy. If an expert's P-value is above a cut-off value (in this case P=0.01) then the expert is weighted with weight proportional to the combined score. Normalized weights for weighted experts are shown in column 6.

A combination of the experts' distributions is termed a "decision maker" (DM). Column 7 gives each expert's Shannon relative information with respect to the equal weight (EW) DM (1). These dimensionless numbers indicate the divergence among the experts themselves and are compared with perturbations caused by dropping a single expert or a single calibration variable (Supplementary Tables 3 and 4). Note that the scores in column 7 are somewhat smaller than the scores in column 3. This suggests that EW is somewhat more informative than the background measure, relative to which the experts' informativeness is measured in column 3.

Other DMs in Table 2, besides EW, are PW01, the performance weighted combination of the eight weighted experts, and PWOpt, the performance weighted combination with the cutoff chosen to optimize the combined score of the DM. Indeed, the combined score of PWOpt (0.4914) is (only) slightly greater than that of PW01 (0.4795). As is typical in such studies, the information of EW is about half that of PWOpt. Very roughly, this translates to EW's average 90% confidence bands being twice as large as those of PWOpt. Similarly, EW's statistical accuracy (P-value) is inferior to that of PWOpt. This is an "in-sample" comparison since DM's are compared on the same set over which PWOpt is optimized. For "out-of-sample" comparisons see below.

Six of the 13 US experts had a statistical accuracy score above 0.01. This is a high number for SEJ studies, especially considering the fact that 16 calibration variables were used, constituting a more powerful statistical test than the traditional number of ten calibration items. Two of the eight EU experts had a statistical accuracy score above 0.01, which is in line with most SEJ studies. There is very little difference between the scores of PWOpt and PW01, though there are modest differences in SLR predictions (see Table S1). A scoring system is asymptotically strictly proper if and only if an expert obtains his/her highest expected score in the long run by, and only by, stating percentiles corresponding to his/her true beliefs. The combined score is an asymptotic strictly proper scoring rule if experts get zero weight when their P-value drops below some threshold (1). If (s)he tries to game the system to maximize his/her expected weight, (s)he will eventually figure out that (s)he must say exactly what (s)he thinks. Honesty is the only optimal strategy. The theory does not say what the cut-off value should be, so that is often chosen by optimization.

In the Classical Model, the optimization works as follows: starting with a cutoff beneath the lowest P-value includes all experts with weight proportional to their combined scores. The combined score of the resulting DM is stored. Taking the expert with the lowest P-value, we next exclude that expert, normalize the remaining combined scores, compute the resulting DM, apply this DM to the calibration variables and store the resulting DM's combined score. Then we remove the next lowest P-value expert and repeat. With N expert P-values this results in N-1 different DM's. We choose the DM whose combined score is the highest. In this case, setting the cut-off at 0.399 and retaining experts 3 and 14 produced the highest scoring DM. With this scoring system it is impossible that a weighted expert has a lower P-value than an unweighted expert, even though doing so might produce a higher DM score. This system can thus be regarded as optimal weighting under a strictly proper scoring rule constraint. The theory was developed in the 1980s and is detailed in (1) and (2).

The Classical Model has been applied in hundreds of expert panels and has been validated both in- and out-of-sample (2-5). In the absence of observations of the variables of interest, out-of-sample validation comes down to cross-validation whereby the calibration variables are repeatedly separated into subsets of training- and test variables. The PW model is initialized on the training variables and scored on the test variables. The superiority of PW over EW in terms of statistical accuracy and informativeness has been demonstrated using this approach.

Expert	**P-value**	**Rel Inf total**	**Rel Inf Calib vbls**	**Nr Calib vbls**	**normalized weight**	**Rel Inf to EW**
exprt01	7.33E-06	2.83	2.44	16		2.34
exprt02	1.37E-08	1.01	1.99	16		1.91
exprt03	5.17E-01	1.09	0.95	16	0.28	0.40
exprt04	8.88E-06	1.66	1.77	16		1.13
exprt05	5.97E-04	1.83	1.45	16		1.18
exprt06	2.62E-02	2.51	1.68	16	0.03	0.91
exprt08	1.47E-01	1.82	1.28	16	0.11	1.06
exprt09	8.97E-02	3.32	1.41	16	0.07	1.05
exprt10	3.91E-03	1.12	1.96	16		1.37
exprt11	1.50E-05	1.98	2.06	16		1.36
exprt12	8.97E-02	1.25	1.68	16	0.09	1.20
exprt14	3.99E-01	1.41	1.32	16	0.30	0.45
exprt22	2.81E-04	2.25	1.83	16		1.30
exprt21	2.81E-04	2.19	1.98	16		1.35
exprt24	3.56E-02	2.80	1.96	16	0.04	1.42
exprt25	3.57E-04	1.70	1.72	16		1.33
exprt26	7.58E-04	1.56	1.70	16		1.24
exprt23	8.66E-05	2.36	2.21	16		1.36
exprt27	8.97E-02	1.66	1.51	16	0.08	0.93

exprt29	3.57E-04	2.11	2.32	16		1.56
PW01	0.6286	1.10	0.76	16		
EW	0.1284	0.59	0.55	16		
PWOpt	0.5173	1.092	0.95	16		

Table S2: Scores and weights for all 21 experts when performance weights are not optimized but computed for all eight experts with statistical accuracy > 0.01. "P-value" denotes the significance level at which the hypothesis that an expert is statistically accurate would be falsely rejected. "Rel Inf Total/ Rel Inf calib vbls" denote the average Shannon relative information in an expert's assessments for all variables and for calibration variables only. "Nr calib vbls" denotes the number of calibration variables answered by an expert. "Normalized weight" for weighted experts is the normalized sum of the product of columns 2 and 4. "Rel Inf to EW" is an expert's relative information with respect to the EW combination of all experts.

1.3 Robustness on Experts

Robustness on experts examines the effect on the PW01 "decision maker" (i.e. the synthetic pooled expert) of losing individual experts. Experts are removed one at a time and PW01 is recomputed. Table S3 shows the resulting information and P-values of the "perturbed" PW01. The rightmost column of Table S4 shows the divergence among the experts themselves. Comparison with the rightmost column of table S3 shows that the scoring results are very robust against loss of a single expert.

Excluded expert	Rel Inf total	Rel Inf Calib vbls	P-value	Rel Inf Orig DM Total	Rel Inf Orig DM Calib vbls
expert 01	1.10	0.76	0.63	0.00	0.00
expert 02	1.07	0.71	0.63	0.02	0.01
expert 03	1.45	0.80	0.63	0.44	0.21
expert 04	1.09	0.76	0.63	0.00	0.00
expert 05	1.10	0.76	0.63	0.00	0.00
expert 06	1.06	0.76	0.52	0.08	0.02
expert 08	1.10	0.76	0.63	0.16	0.14
expert 09	0.95	0.79	0.63	0.18	0.07
expert 10	1.07	0.76	0.63	0.01	0.00
expert 11	1.09	0.76	0.63	0.00	0.00
expert 12	1.12	0.79	0.37	0.10	0.10
expert 14	1.26	0.73	0.52	0.35	0.18
expert 22	1.09	0.75	0.63	0.00	0.00
expert 21	1.10	0.76	0.63	0.00	0.00
expert 24	1.03	0.74	0.63	0.11	0.03
expert 25	1.08	0.76	0.37	0.01	0.00
expert 26	1.09	0.76	0.63	0.02	0.00
expert 23	1.09	0.76	0.63	0.00	0.00
expert 27	1.10	0.74	0.52	0.05	0.04
expert 29	1.10	0.76	0.63	0.00	0.00
None	1.10	0.76	0.63	0.00	0.00

Table S3. Robustness on experts; for explanations of columns see section 1.2

A more complete sense of robustness would examine the effect of the method of recruitment of experts and of the elicitation team. Before the Classical Model was adopted for the European uncertainty analysis of accident consequence codes for nuclear power plants (6), the authorities in Brussels required that parallel elicitations be carried out using the same elicitation protocol, but with different elicitation teams independently recruiting different experts. The findings in this case indicate a strong convergence of elicitation results from the two groups (6). Such an approach is generally far beyond the budgets of most applications. However, the results here (Table S1) show good general agreement on SLR between the US and European panels who were elicited separately. A different type of robustness is gleaned from the 14 year running expert judgment assessments of risks from the Montserrat volcano (7). Those assessments concerned a consistent elicitation method, applied to the same variables under changing conditions, with some exchanging of participating experts over elicitations. The approach showed good consistency of performance for volcanic hazard assessment purposes, over more than seventy repeat elicitations.

1.4 Robustness on Items

Seed variables are removed one at a time and PW01 is recomputed. These scores are extremely robust against loss of a seed variable. Comparing the rightmost columns of Supplementary Tables 3 and 5 shows that the perturbation caused by loss of a single calibration variable is very small relative to the divergence among the experts themselves.

Excluded vbl	Rel Inf total	Rel Inf Calib vbls	P-value	Rel Inf Orig DM Total	Rel Inf Orig DM Calib vbls
Calib vbl 1	0.94	0.71	0.60	0.22	0.10
Calib vbl 2	1.13	0.79	0.66	0.10	0.06
Calib vbl 3	1.17	0.76	0.66	0.12	0.06
Calib vbl 4	1.30	0.80	0.66	0.33	0.05
Calib vbl 5	1.34	0.76	0.39	0.29	0.12
Calib vbl 6	1.01	0.73	0.29	0.22	0.14
Calib vbl 7	1.04	0.81	0.39	0.15	0.07
Calib vbl 8	1.26	0.83	0.39	0.23	0.06
Calib vbl 9	1.02	0.79	0.66	0.17	0.08
Calib vbl 10	1.12	0.79	0.66	0.18	0.15
Calib vbl 11	1.02	0.77	0.60	0.17	0.05
Calib vbl 12	1.44	0.73	0.39	0.38	0.08
Calib vbl 13	1.45	0.77	0.30	0.38	0.07
Calib vbl 14	1.29	0.73	0.39	0.28	0.04
Calib vbl 15	1.39	0.77	0.66	0.38	0.13
Calib vbl 16	1.14	0.87	0.36	0.24	0.15
None	1.10	0.76	0.63		

Table S4. Robustness on calibration questions (calib vbl); for explanations of columns see section 1.2 and caption for Table S2.

1.5 Dependence Elicitation

Dependence and especially tail dependence are unfamiliar concepts for many scientists. A PowerPoint presentation was given to the experts, before the elicitation, to introduce these notions, where the reader can find precise definitions (see Supplementary Section 8 for links). Figure S6 from the presentation shows how aggregation affects uncertainty. 3σ or one in 1000 upper tail events are depicted for the sum of 10 zero mean normal variables. If the variables have a pairwise correlation of 0.5, the distribution dilates such that the 3σ event coincides with the 5σ event of the sum of independent Normals. If this pairwise correlation is realized with an upper tail dependent copula, the 3σ event coincides with the 7σ event for independent Normals. Thus an event whose probability is 1/1000 (3σ) will appear to be an event with probability 1.28×10^{-12}(7σ) when tail dependence is present but ignored.

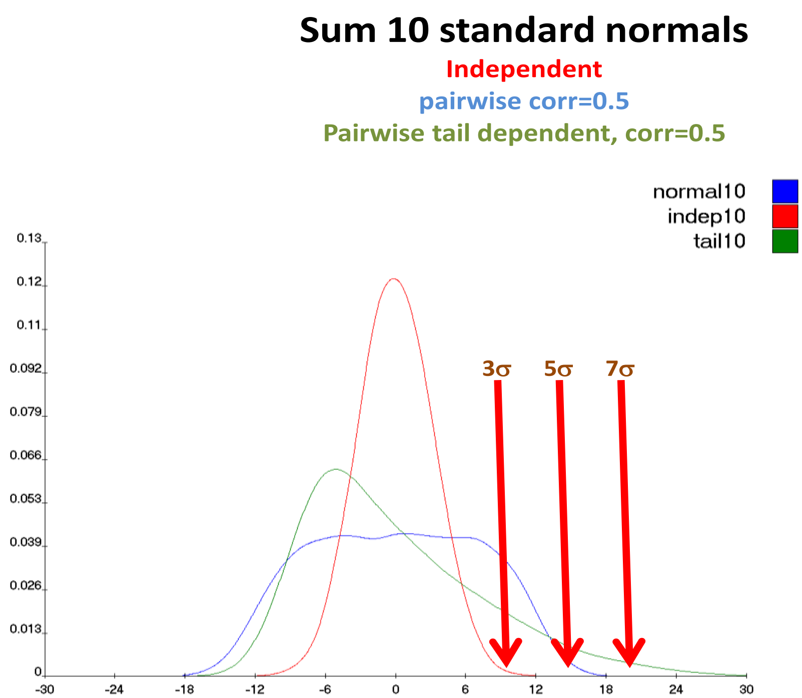


Figure S6: Effect of aggregation on uncertainty. 3σ or one in 1000 upper tail events are depicted for the sum of 10 zero mean normal variables. If the variables have a pairwise correlation of 0.5, the distribution dilates such that the 3σ event coincides with the 5σ event (2.87×10^{-7}) of the sum of independent Normals. If this pairwise correlation is realized with an upper tail dependent copula, the 3σ event coincides with the 7σ event for independent Normals (1.28×10^{-12}).

The dependence elicitation for pairs of variables was accomplished by eliciting conditional exceedance probabilities: for central correlation, experts answered: "*what is the probability that variable X exceeds its median given that variable Y has exceeded its median?*". Numerical and verbal answers were accepted (Table S8). For upper tail dependence, "median" was replaced by "95th percentile" in the above question and verbal responses were elicited as indicated in Table S8.

Three random variables (Runoff, Discharge and Accumulation) for each of the three ice sheets yield 36 pairs of variables. Potential dependences between ice sheets were also identified. Based on judgments of size and relevance, the analysis team pared this down to 10 pairs corresponding to the colored nodes in Figure S8, in addition to 3 inter-ice sheet relations. This structure is a "dependence vine" for determining a high dimensional joint distribution based on bivariate and conditional bivariate distributions. Unspecified (conditional) bivariate distributions are conditionally independent, making it easy to extend a partially specified structure to the minimally informative realization of the specified structure.

The basic “dependence vine” for expert 14, as an example, is shown in Figure S8. The ellipses represent the variables (GA = Greenland Accumulation; WD = West Antarctica Discharge; etc). The dependences, represented by arcs, are quantified by assessing exceedance probabilities. The colored nodes are those between which dependence is assessed. Conditional independence is assumed elsewhere. Calculations and sampling were performed with the freeware [UNINET](). This exposition of vine theory is necessarily incomplete; a [Wiki page]() provides more background and references. A full exposition is in (8, 9).

For each of the eight experts with P-value > 0.01, a comparable regular vine was constructed using the dependence information elicited from each individual expert. These eight joint distributions were combined with the various weighting schemes shown in Table S2.

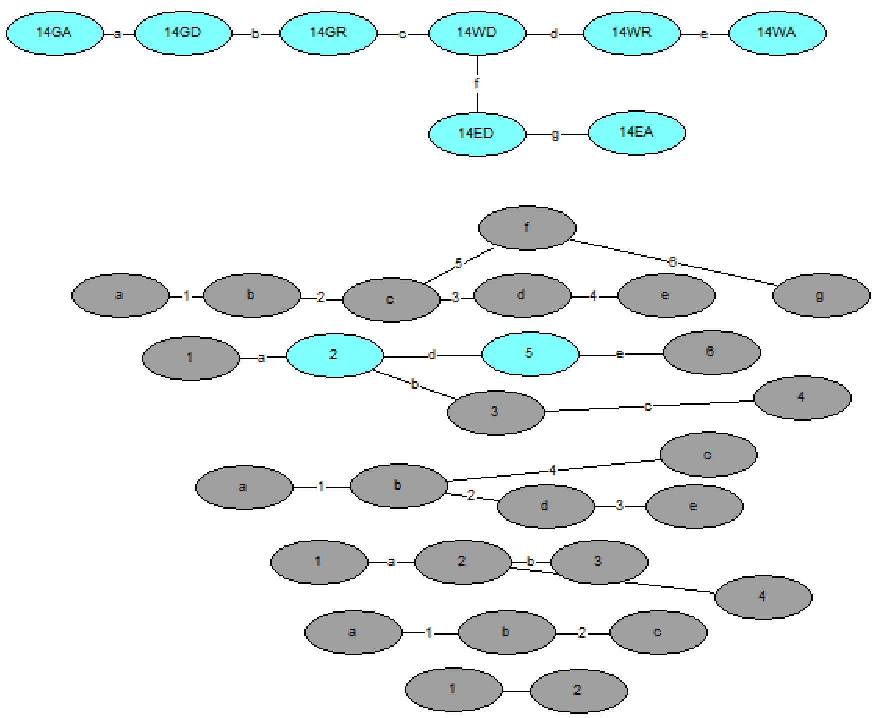


Figure S8: Regular vine depiction of the joint distribution for expert 14. The colored nodes correspond to information elicited from the expert. Grey nodes are given values by taking the minimal information extension of the information from the colored nodes.

1.6 Exceedance Results and Graphs with Independence

Ice Sheets contribution to SLR [cm]; 20 US & EU experts; 15k samples for EW and PWOpt; assuming INDEPENDENCE												
2300H	Mean	StDev	5%	50%	95%		2300L	Mean	StDev	5%	50%	95%
EWE	144	195	-37	69	432		EWE	39	58	-33	24	142
EWG	149	130	-18	128	397		EWG	83	75	-11	73	223
EWW	154	137	-38	135	397		EWW	83	92	-40	64	252
PWE	49	151	-53	23	122		PWE	15	40	-29	9	67
PWG	121	104	-4	95	311		PWG	72	67	-7	52	198
PWW	117	124	-47	103	337		PWW	58	88	-63	50	221
EW	447	271	72	414	936		EW	205	132	7	193	437
PW	286	220	19	250	663		PW	145	117	-25	133	355
2200H	Mean	StDev	5%	50%	95%		2200L	Mean	StDev	5%	50%	95%
EWE	104	152	-24	39	316		EWE	14	23	-20	11	56
EWG	81	71	-13	70	212		EWG	49	45	-8	43	132
EWW	113	117	-28	83	320		EWW	44	47	-18	33	131
PWE	35	119	-28	15	62		PWE	8	16	-15	6	35
PWG	69	62	-7	52	184		PWG	44	40	-5	35	119
PWW	68	92	-31	56	199		PWW	33	49	-35	28	123
EW	298	203	34	267	672		EW	106	69	3	100	229
PW	171	161	2	139	482		PW	84	65	-12	78	199
2100H	Mean	StDev	5%	50%	95%		2100L	Mean	StDev	5%	50%	95%
EWE	19	23	-11	13	62		EWE	4	9	-9	3	19
EWG	30	25	-5	26	76		EWG	14	13	-4	12	38
EWW	38	37	-9	29	106		EWW	16	19	-7	11	51
PWE	8	16	-13	5	31		PWE	1	6	-8	1	12
PWG	25	23	-4	21	69		PWG	13	13	-4	12	38
PWW	24	28	-9	19	65		PWW	12	15	-5	9	36
EW	86	51	12	81	177		EW	34	25	-2	31	78
PW	56	40	1	51	130		PW	27	21	-3	24	64
2050H	Mean	StDev	5%	50%	95%		2050L	Mean	StDev	5%	50%	95%
EWE	2	5	-6	2	11		EWE	1	4	-5	1	7
EWG	6	5	-1	6	16		EWG	4	4	-1	4	11
EWW	8	8	-2	5	23		EWW	5	6	-3	3	16
PWE	2	6	-7	1	13		PWE	0	3	-4	0	5
PWG	6	6	-2	5	18		PWG	5	4	-1	4	12
PWW	5	6	-2	4	15		PWW	3	4	-2	3	10
EW	16	11	0	15	36		EW	10	8	-2	9	24
PW	13	10	-2	13	31		PW	8	7	-2	8	19

Table S5: Results for SLR [cm] assuming independence, for EW and PWOpt excluding the baseline values for 2000-2010 (see Supplementary note 5).

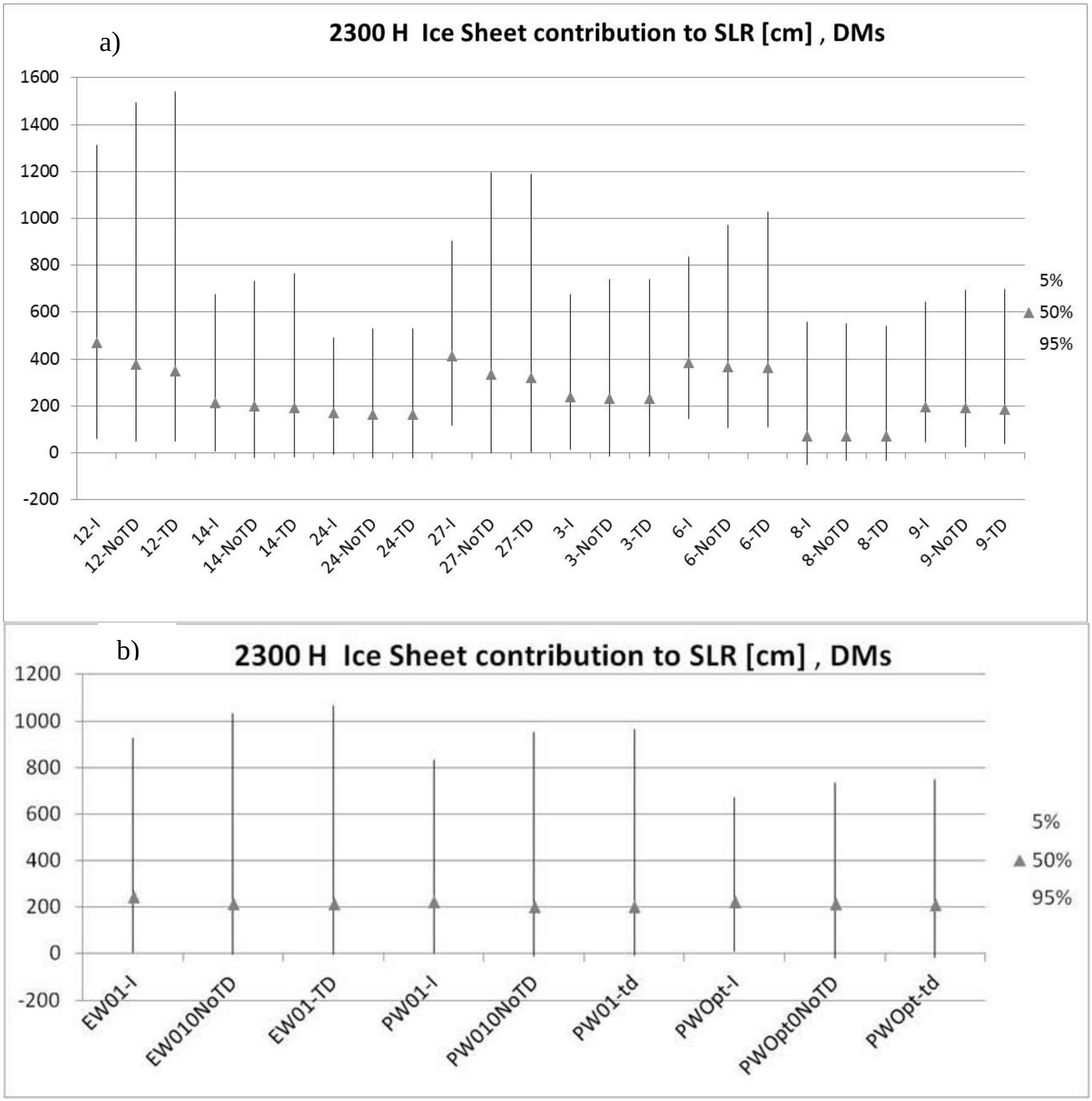


Figure S9: a) PW individual expert with independence (I), no tail dependence (NoTD), and including tail dependence (TD) for 2300 H contribution to SLR [cm]; b) EW and PW values with and without dependencies (as for a)).

Supplementary note 2: From expert quantiles to SLR

The procedure of going from expert quantiles to distributions for SLR is as follows (for detailed information see (2), especially the online supplementary material Appendix A):

1) For each variable we determine an "intrinsic range" (IR), the smallest interval that contains all expert assessments plus the realization (in case of seed variables) + a 10% overshoot below and above (10% is a parameter that can be adjusted in the code)
2) We put a background measure on each IR. In the code the user can choose between the uniform and log-uniform background measure. Log-uniform is indicated when experts reason in orders of magnitude. In this case all background measures are uniform. Other choices could be made but would require re-coding.
3) For each expert and each variable, we fit a density that is minimally informative with respect to the background measure and complies with the expert's quantile assessments. For the uniform background, this is a piecewise uniform density. This density "adds as little as possible" to the expert's assessment. Note that fitting a two-parameter family such as the Gaussian distribution will often be unable to match 3 quantiles.
4) ASSUMING INDEPENDENCE
    a. With N experts we form the EW combination by simple averaging of the experts' densities. DO NOT average the quantiles; that can give a very overconfident result.
    b. With PW, we take a weighted average of densities.
    c. Simple Monte Carlo sampling is used to build a distribution for SLR. For each ice sheet we sample D, R and A and store D+R-A.
    d. Monte Carlo sampling is used to build a distribution of total SLR as SLRGr + SLREAIS + SLRWAIS. Again, do not sum the quantiles.
5) WITH DEPENDENCE, we build a joint density for each expert based on the elicited exceedance probabilities. This cannot be done with generic software. XL add-ons like @Risk and Crystal Ball impose the assumption of the Gaussian copula. Based on a pilot elicitation with the 2012 experts, we anticipated that tail dependence could be significant, rendering the Gaussian copula inappropriate. For each expert we obtain a distribution for total SLR, and we take a weighted average of these densities to find the combined distribution for SLR. Each expert's total SLR distributions incorporates his/her dependence.

Steps (1) – (3) can be done with freeware EXCALIBUR (EXpert CALIBRation). Step (4) can be done with Freeware UNICORN (UNCertainty analysis wIth CORelatioNs)– which has limited dependence modeling capability). Step (5) uses freeware UNINET which is much more powerful. All these programs can be downloaded from http://www.lighttwist.net/wp/.

Supplementary note 3: Experts' rationales

This section summarizes and collates the expert responses to the rationale questionnaire that is reproduced in supplementary note 6. For descriptions of the processes considered see note 6.

GrIS IV	GrIS EF	WAIS IV	WAIS EF	EAIS No trend	EAIS No trend
0.5	0.5	0.5	0.5		
	1	1			
	1	1			
	1	0.3	0.7		
	1		1		
0.5	0.5	0.5	0.5		
0.5	0.5	0.5	0.5		
	1	0.5	0.5		
	1	0.5	0.5		
	1		1		
0.1	0.9	0.3	0.7		
	1	0.9	0.1		
	1	0.7	0.3		
	1	0.5	0.5		
0.4	0.6	0.5	0.5		
0.3	0.7	0.7	0.3		
	1	0.5	0.5		
	1	1			
0.5	0.5	1			
	1		1		
	1	0.3	0.7		
	1		1		
2.8	19.2	11.2	10.8		

Table S6. Expert response to the first question in the descriptive rationales (supplementary note 6): "Are the recent ~decadal trends in mass balance largely due to internal variability (IV) of the atmosphere/ice/ocean/climate system or external forcing (EF), for each ice sheet overall?

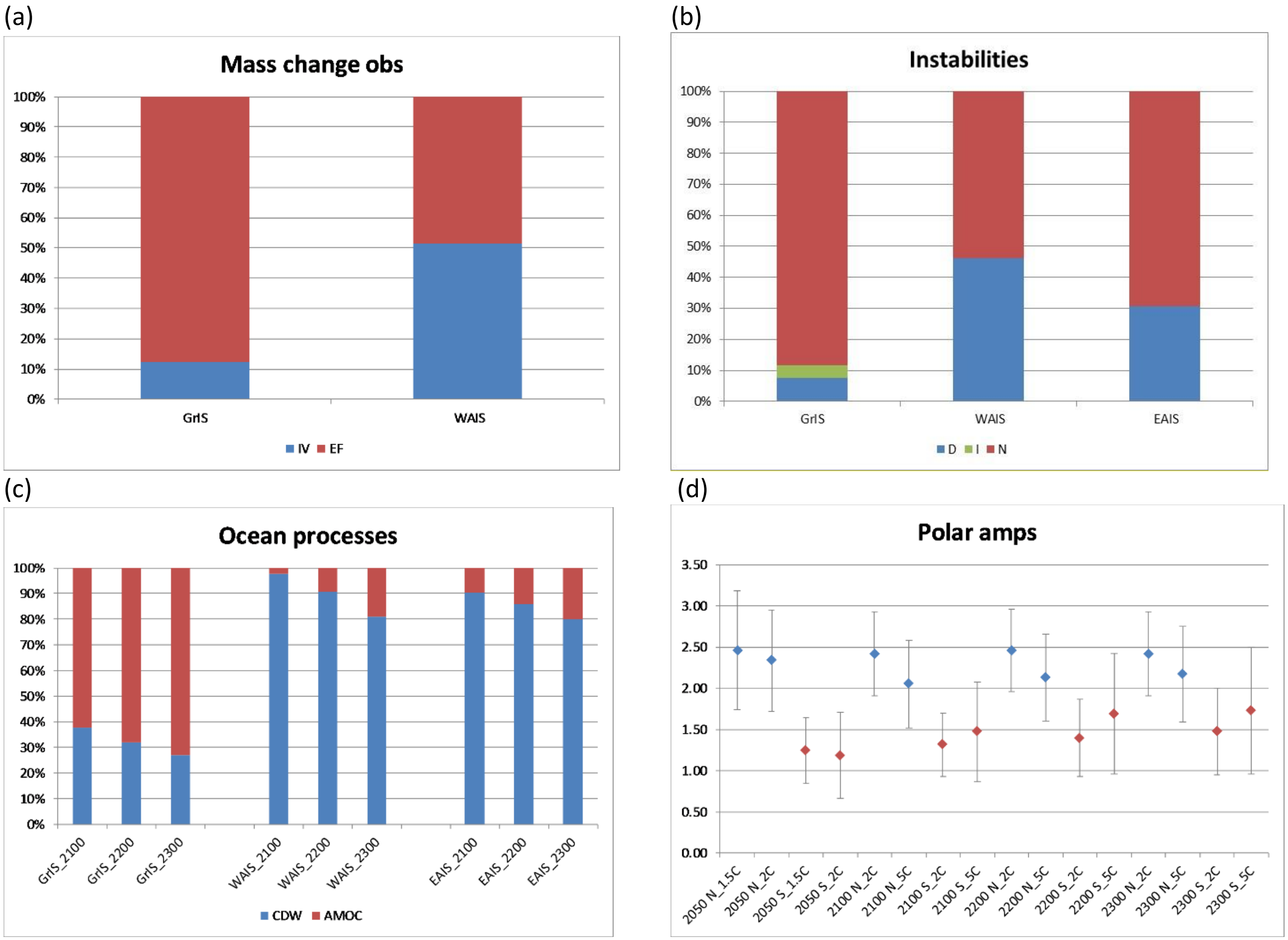
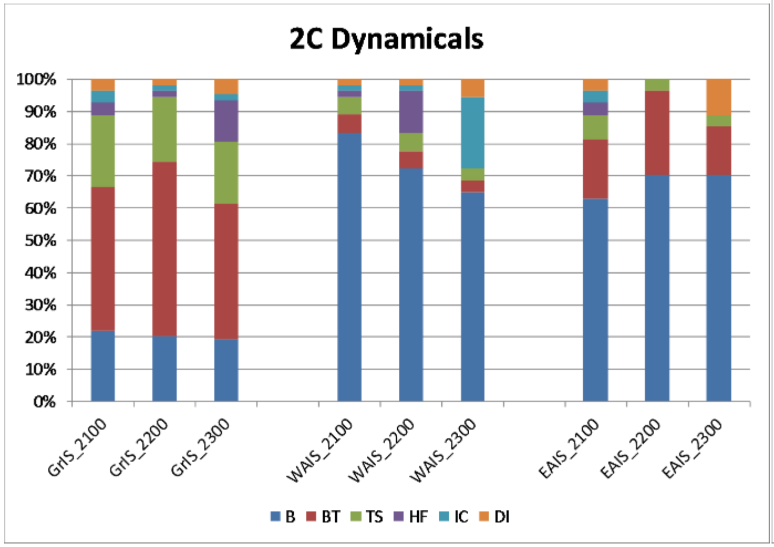
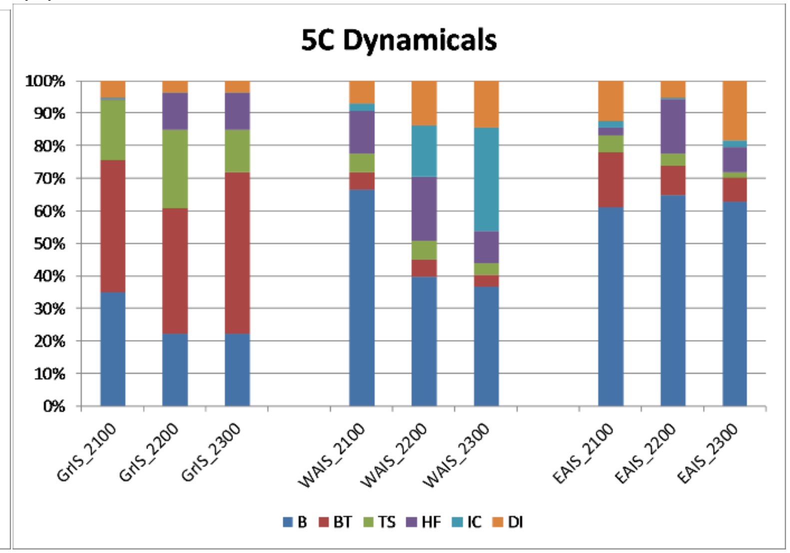


Figure S9: Descriptive rationales illustrated graphically; a) Internal Variability (IV) vs External Forcing (EF) for decadal mass trends, b) dynamical processes due to near field gravitational and vertical land motion effects (D=decrease, I=Increase, N=no change) , c) ocean circulation changes US experts only (CDW=circumpolar deep water, AMOC=Atlantic Meridional Overturning Circulation), d) polar amplification values assumed. For detailed descriptions see Supplementary Note 6.

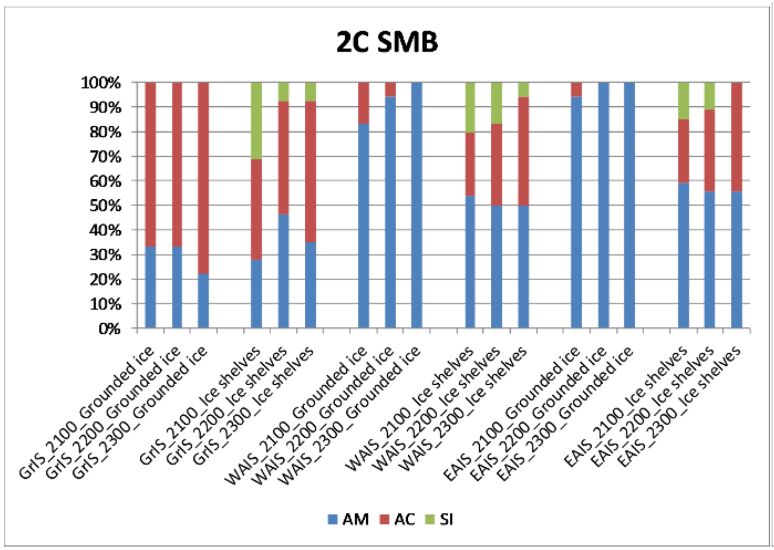
(a)


2C Dynamicals
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
GrIS_2100
GrIS_2200
GrIS_2300
WAIS_2100
WAIS_2200
WAIS_2300
EAIS_2100
EAIS_2200
EAIS_2300
B BT TS HF IC DI

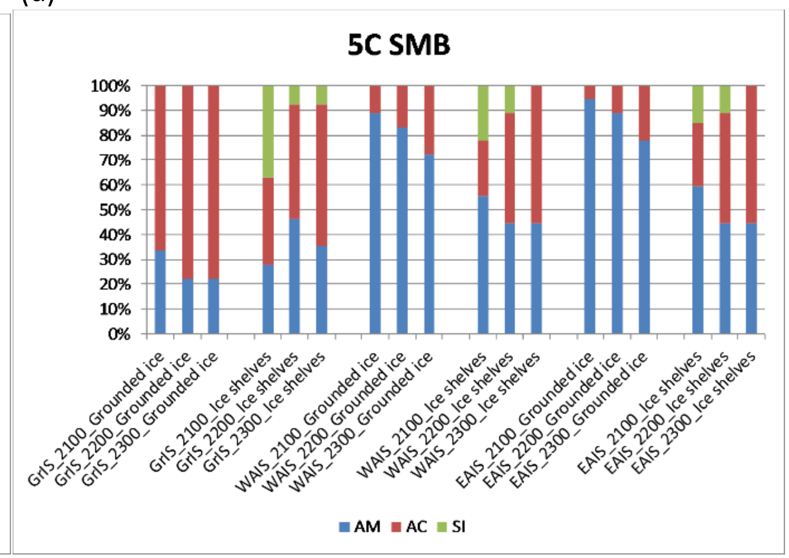
(b)


5C Dynamicals
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
GrIS_2100
GrIS_2200
GrIS_2300
WAIS_2100
WAIS_2200
WAIS_2300
EAIS_2100
EAIS_2200
EAIS_2300
B BT TS HF IC DI

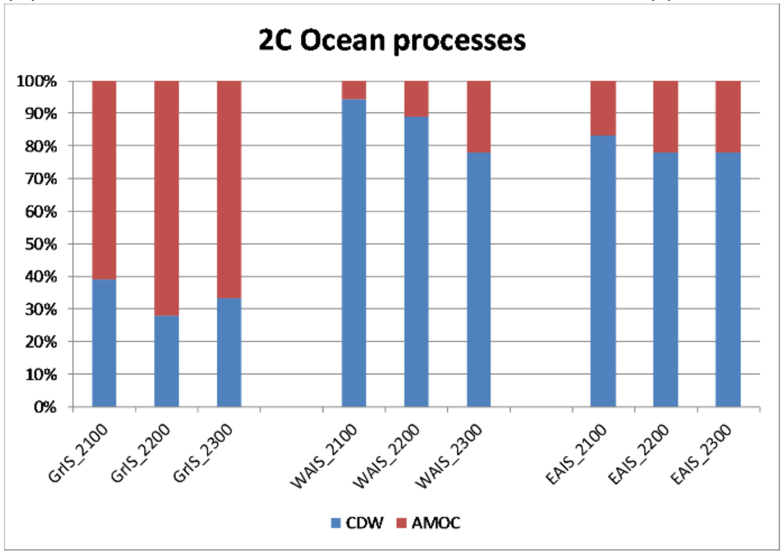
(c)


2C SMB
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
GrIS_2100_Grounded ice
GrIS_2200_Grounded ice
GrIS_2300_Grounded ice
GrIS_2100_Ice shelves
GrIS_2200_Ice shelves
GrIS_2300_Ice shelves
WAIS_2100_Grounded ice
WAIS_2200_Grounded ice
WAIS_2300_Grounded ice
WAIS_2100_Ice shelves
WAIS_2200_Ice shelves
WAIS_2300_Ice shelves
EAIS_2100_Grounded ice
EAIS_2200_Grounded ice
EAIS_2300_Grounded ice
EAIS_2100_Ice shelves
EAIS_2200_Ice shelves
EAIS_2300_Ice shelves
AM AC SI

(d)


5C SMB
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
GrIS_2100_Grounded ice
GrIS_2200_Grounded ice
GrIS_2300_Grounded ice
GrIS_2100_Ice shelves
GrIS_2200_Ice shelves
GrIS_2300_Ice shelves
WAIS_2100_Grounded ice
WAIS_2200_Grounded ice
WAIS_2300_Grounded ice
WAIS_2100_Ice shelves
WAIS_2200_Ice shelves
WAIS_2300_Ice shelves
EAIS_2100_Grounded ice
EAIS_2200_Grounded ice
EAIS_2300_Grounded ice
EAIS_2100_Ice shelves
EAIS_2200_Ice shelves
EAIS_2300_Ice shelves
AM AC SI

(e)



(f)

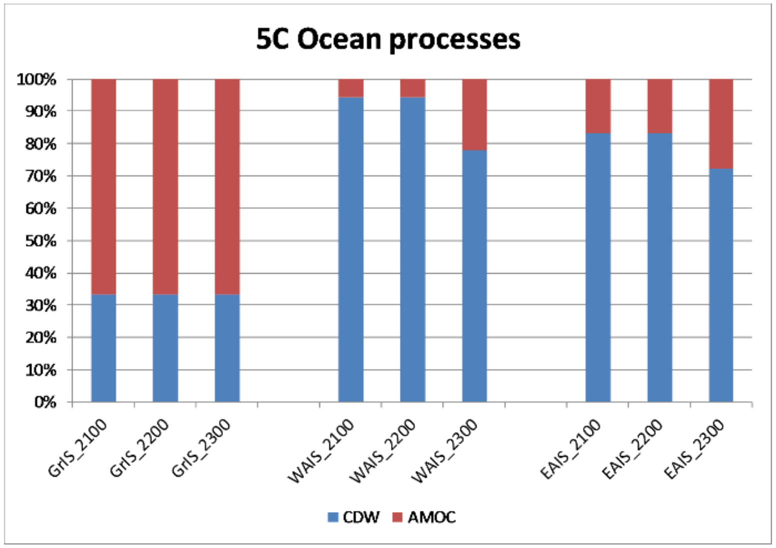


Figure S10: Rationale responses for +2° C and +5° C for EU experts only. a) processes that might influence ice dynamics for 2°C; b) as for a) but for 5°C; c) processes that might influence SMB for +2° C; d) as for c) but for +5 °C; e) ocean circulation changes that might influence ice dynamics for +2° C; f) as for e) but for +5° C. For detailed descriptions for the abbreviations and processes considered see Supplementary Note 6.

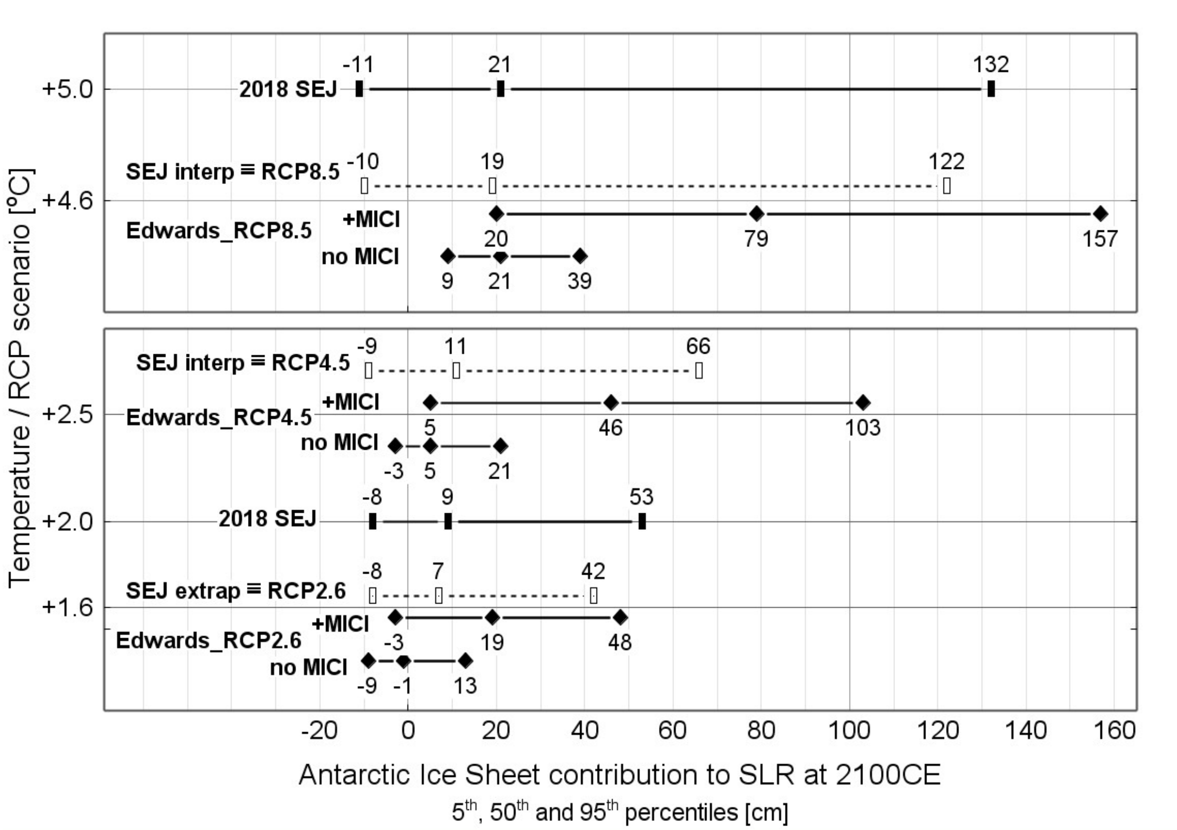


Figure S11. Comparison between the emulator results in Edwards et al (2019)(10) and this elicitation (SEJ2018) where the Marine Ice Cliff Instability (MICI) is excluded and included. The dashed lines are from a linear interpolation of SEJ2018 (at +2.0 ˚C and +5.0 ˚C, respectively) to the equivalent temperatures at 2100 for the RCP scenarios considered in (10): RCP8.5 (top panel), RCP4.5 and RCP2.6 (lower panel).

It is apparent that SEJ2018 value spreads for Antarctic Ice Sheet contribution to sea level rise in 2100CE lie above Edwards et al (2019)*(10)* no-MICI but are substantially lower than the 50th and 95th percentile MICI values obtained using the emulator in (*10*).

Supplementary note 4: Briefing note sent to experts prior to elicitation

The following briefing note was sent to all experts prior to the elicitation:

Eliciting Ice Sheets' Contribution to Sea Level Rise

Sept 28, 2017

Introduction: Probabilistic prediction for Ice Sheet contributions to Sea Level Rise
With global warming, ice sheets in Greenland and Antarctica are likely to become the primary agents of Sea Level Rise (SLR) in the coming decades and centuries. In their normally slow, march to the sea, glaciers draining the ice sheets exhibit dynamics which are highly variable from place to place, with neighboring glaciers or ice streams responding in markedly different ways to the same external forcing. Dynamic models must account for things like bedrock properties (including slipperiness and topography), ice shelf buttressing, precipitation, melt water effects on ice stiffness, grounding line migration, ocean currents, and ice cliff instability. Some of these features are directly observable, many are not.

Glaciologists focusing on individual glaciers must contend with many uncertainties when predicting future ice mechanics and dynamics out to, say 2100CE, or even 2300CE. Point predictions, whatever their pedigree, are of limited value when the uncertainties are very large. Scientists must therefore make *probabilistic predictions;* they must say, in effect "My best estimate is a *+0.2mm* contribution to SLR by 2100CE from this particular glacier, and I am 90% confident the contribution will be between -*3mm* and *+6mm*". A narrative might explain, say, "*the contribution could actually be negative (the ice sheet would actually grow) if warming and changing atmospheric and ocean circulation increased winter precipitation inland while leaving the buttressing ice shelves largely intact; a very high contribution might result if increased storminess and shifting ocean currents break up ice shelves or summer coastal precipitation causes increased calving and instability".* Capturing the narrative behind the uncertainty assessments is essential for understanding and communicating our current state of knowledge.

That is the *easy* part. Judging the cumulative future effects of the main ice sheets on sea level rise raises a host of new questions and methodological challenges, lying further outside most physical scientists' comfort zone. What might be the joint impacts of ice sheet responses on SLR if extreme conditions were encountered under global climate change?

A proof of concept
We describe a proof-of-concept demonstration for using expert judgments to constrain quantitative estimates of dependences in potentially correlated processes that affect the ice sheet[2], and indicate some preliminary trial results. We also explore the influence on these results of different ways of combining expert judgments[3].

---

[2] A talk on this subject was given at the Banff research center in 2013 by Roger Cooke and can be streamed from http://www.birs.ca/events/2013/5-day-workshops/13w5146/videos/watch/201305221037-Cooke.html

[3] A talk on performance weighting was given at the Centers for Disease Control and Prevention in Atlanta GA on May 23, 2017 by Willy Aspinall, and may be streamed at https://www.youtube.com/watch?v=FPC-h-br8i8&feature=youtu.be

A recent study extending (Bamber and Aspinall, 2013, henceforth B&A) made a first pass at assessing dependences between macro process variables relating to the Greenland, West Antarctica and East Antarctica ice sheets. Estimates of contributions to SLR were based on the B&A protocol. A typical question was

*In the case of **Greenland**, for a global mean annual Surface Average Temperature rise of 3**°C** by **2100** with respect to pre-industrial, what will be the **integrated contribution,** in **mm** to **SLR** relative to 2000-2010 of the following:*

i) accumulation
5% value: ___________ 95% value: ___________ 50%value:: ____________

ii) runoff
5% value: ___________ 95% value: ___________ 50%value:: ____________

iii) discharge
5% value: ___________ 95% value: ___________ 50%value:: ____________

Similar questions were directed to West and East Antarctica, and to different temperatures, out to 2200.

The dependence elicitation was based on exceedance probabilities, as pioneered in the 1990's by [uncertainty analyses for nuclear power plants in the US and in Europe]. Whereas the earlier nuclear work used only the 50% exceedance probabilities, our ice sheet follow-on study asked also for 95% exceedance probabilities.

Within ice sheet process dependencies to 2100CE
Greenland Ice Sheet, 2100 3°C warming

*Given **discharge** ≥ your 50% value, what is probability that **runoff** also ≥ your 50% =____*

*Given **discharge** ≥ your **95%** value, what is probability that **runoff** also ≥ your **95%** =____*

*Given **accumulation** ≥ your 50% value, what is probability that **discharge** also ≥ your 50% =____*

*Given **accumulation** ≥ your 50% value, what is probability that **runoff** also ≥ your 50% =____*

In answering questions concerning the 95% exceedances, the experts had to consider whether factors likely to produce extreme values in one variable would also produce extreme values in the other.

An extensive [procedures guide] for structured expert judgment emerging from these nuclear studies has informed many subsequent applications, including B&A. Following the Classical Model for structured expert judgment (Cooke 1991; Cooke and Goossens 2008), calibration variables from the experts' field were used by B&A to score the experts' statistical accuracy and informativeness. True values of calibration variables are known post hoc, they preferably concern near term future measurements, but can also involve unfamiliar intersections of past data or literature. An illustrative calibration variable from B&A was

There are nine main glacier/ice caps on Iceland. What was their 2009/2010 average climatic balance B_{clim}, in Kg/m^2? (please indicate gain by +value, loss by -value)

5% value: ___________ 95% value: ___________ 50%value: ____________

Based on extensive experience with the Classical Model, an equally weighted combination of experts tends to give statistically accurate assessments exhibiting wide confidence bounds (low information). The goal of the Classical Model is to demonstrate high statistical accuracy with narrow confidence bounds. This is accomplished by differentially weighting the experts so as to favor those with high statistical accuracy and high information. Recent background on the Classical Model for climate uncertainty quantification may be found here. Other recent applications are summarized here, a Wikipedia page gives some background, and an extensive study of out-of-sample validation with complete mathematical exposition in supplementary material is here.

Dependence and aggregation

The SLR contribution of, say, the Jakobshaven glacier in western Greenland up to 2100CE is a random variable; it can be described mathematically by giving a range of possible values and a probability that each value would be realized. Quantifying the uncertainty in the contribution to SLR from the Greenland and Antarctic ice sheets involves adding together hundreds of random variables. Adding random variables is not like adding ordinary numbers. In adding two random variables, say the Jakobshaven and the Petermann glaciers' contributions to SLR by 2100CE, we must consider all possible combinations of values for Jakobshaven and for Petermann, and consider the probability that these values arise together. Suppose the contribution from Jakobshaven were very large. According to the above narrative, that suggests certain influencing factors are in play; how would these influences affect the Petermann glacier, 1274 km to the north? If they would also tend to induce high contribution values for the Petermann, then this could indicate a positive dependence between the SLR contributions of the two glaciers. If, on the contrary, the drivers of elevated ice mass loss in the west of Greenland were conducive to more stable conditions in the north, then the inter-glacier dependence might be negative.

The more random variables we aggregate, the more important the effects of long range, global correlations can become, a feature which our intuitions easily under-appreciate. A neglected weak global correlation of $\rho = 0.2$ when summing *500* normal variables underestimates the confidence interval of the sum by an order of magnitude. Global correlations also amplify the correlation of aggregations. In the above example, the correlation between the sum of the first *250* variables and the sum of the second *250* variables is *0.992.* In contemplating the uncertainty in the effects of hundreds of glaciers, we must consider the overall effects of these dependencies.

Tail Dependence

The correlation coefficient represents a sort of average association between two random variables. This often yields an adequate measure of their association, but not always. The linkage between two variables may be primarily due to factors driving the extreme values, not the more mundane, central values.

For example, under normal conditions it may be that the mass loss at Jakobshaven and at Petermann vary according to local weather conditions, which are largely uncorrelated. However, very large mass loss at Jakobshaven would implicate large scale warming factors, which in turn could imply large mass loss at Petermann as well. In such cases one speaks of positive tail dependence between the two variables, here glacier processes. *Tail dependence* can be positive or negative, can affect the upper or lower tails of distributions, or both, and bears no direct relation to the ordinary correlation. Thus, two Gaussian variables are always tail independent. Given that one of them exceeds its r^{th} percentile, the probability that the second also exceeds its own r^{th} percentile tends toward the independent value of *(100-r)%* as *r* tends to *100%* regardless of the correlation, provided it is strictly between *1* and *–1*.

In other words, a very high value of one variable tends *not* to entrain a high value of the other with two Gaussian variables, but this will not be true for variables characterized by other distributions.

Results

The calculations were performed by Aspinall and Cooke defining a regular vine, using the experts' responses. Regular vines capture dependence in terms of nested bivariate and conditional bivariate distributions on the ranks of random variables, called copulae. The copulae are chosen to mimic the elicited exceedance probabilities. Our highest weighted experts evinced tail dependence between Greenland *Discharge* and Greenland *Runoff*, between West Antarctica *Discharge* and West Antarctica *Runoff*, and between West Antarctica *Discharge* and East Antarctica *Discharge*. Although the actual values of tail dependence varied between experts, they were comparable in magnitude. Other variables exhibited median dependence without tail dependence.

Table 1 presents the overall results for the case of +3˚C global warming by 2100CE, and enables us to gauge the effects of dependence and of performance weighting. "*EW*" denotes the combination based on equal weighting of all nine experts, "*PW*" denotes the optimal performance weighting in which two experts were weighted, based on statistical accuracy and informativeness[4]. "*Indep*" signifies that experts' dependence information was not used. The contribution to SLR was computed, per ice sheet, as *Runoff + Discharge – Accumulation* as if these were independent random variables. "*tail indep*" signifies that tail dependence was ignored and dependence was based only on the 50% exceedance probabilities. "*tail dep*" includes the information on tail dependence.

Table 1. "EW" denotes the combination based on equal weighting of all experts, "PW" denotes the optimal performance weighting in which the experts were weighted. "Indep" signifies that no dependence information was used. "tail indep" signifies that dependence was based only on the 50% exceedance probabilities. "tail dep" includes the information on tail dependence.

Ice sheet contribution to SLR by 2100CE with +3˚C warming [mm]						
		mean	stdev	5%	50%	95%
Expert combination method	EW indep	615	270	238	581	1120
	PW Indep	335	200	71	307	719
	PW tail indep	337	216	64	305	749
	PW tail dep	338	229	71	292	785

The largest effect is wrought by using performance-based weighting instead of expert equal weighting. The mean SLR of "EW indep" is nearly twice that of the PW combinations, and the 5- 50- and95-percentiles are substantially shifted upwards, relative to any of the alternative combinations. Focusing on the PW combinations, the effect of including dependence information is most visible in the 95th percentiles; the corresponding means are notably consistent. Including ice sheet inter-dependence,

[4] These are "linear pooling methods"; other methods have also been proposed for ice sheet uncertainty quantification, for a discussion see Bamber et al (2016).

without tail dependence, raises the 95th percentile by *+37 mm* relative to the independent case; including tail dependence raises this percentile by *+66mm* relative to the independent case.

Conclusions

Predicting the cumulative effect of ice sheets on Sea Level Rise by 2100CE involves large uncertainties. Developing science-based quantifications of these uncertainties obliges scientists to venture outside their comfort zone of deterministic model-based predictions and deal with expert subjective uncertainty assessments. Adding information on dependence and tail dependence increased the values of the upper tail *95th* percentiles of the performance weight combination. However, that increase effect was dominated by the reduction in SLR predictions produced by restricting the elicitation solution to our statistically accurate experts.

References

Bamber, J.L., and Aspinall, W.P., (2013) An expert judgement assessment of future sea 1evel rise from the ice sheets, Nature Climate Change, published online: January 6, 2013 | DOI: 10.1038/NCLIMATE1778.

Bamber, J.L., Aspinall, W.P. and Cooke, R.M. (2016) A commentary on "how to interpret expert judgment assessments of twenty-first century sea-level rise" by Hylke de Vries and Roderik SW van de Wal. Climatic Change 137(3): 321-328.
We clarify key aspects of the evaluation, by de Vries and van de Wal (2015), of our expert elicitation paper on the contributions of ice sheet melting to sea level rise due to future global temperature rise scenarios (Bamber and Aspinall 2013), and extend the conversation with further analysis of their proposed approach for combining expert uncertainty judgments.

Bedford, T.J., & Cooke, Roger M., (2002) Vines - a new graphical model for dependent Random Variables. Annals of Statistics, vol. 30, No.4, August 2002.

Colson, A. and Cooke, R.M., (2017) Cross Validation for the Classical Model of Structured Expert Judgment, Reliability Engineering and System Safety, Volume 163, July 2017, Pages 109–120 http://dx.doi.org/10.1016/j.ress.2017.02.003

Cooke, R.M., Saatchi, S. and Hagen, S. (2016) Global Correlation and Uncertainty Accounting, Dependence Modeling, vol. 4 nr 1 184-189, DOI:10.1515/demo-2016-0009 https://www.degruyter.com/view/j/demo.2016.4.issue-1/demo-2016-0009/demo-2016-0009.xml**.**

Cooke, Roger M., Goossens, L.H.J. (2008) TU Delft Expert Judgment Data Base, Special issue on expert judgment Reliability Engineering & System Safety, 93, 657-674, Available online 12 March 2007, Issue 5, May 2008.

Oppenheimer, M., Little, C.M. and Cooke, R.M. (2016) Expert Judgment and Uncertainty Quantification for Climate Change, Nature Climate Change. vol 6, May, 445-451, published online: 27 April 2016 | DOI: 10.1038/NCLIMATE2959 *Radiation Protection and Dosimetry* Special Issue, vol. 90 no 3, 2000.

Supplementary note 5: Reference values for ice-sheet processes

GrIS	Accumulation	700 Gt/yr
	Runoff	400
	Discharge	500
	Net	-200
EAIS	Accumulation	1130
	Discharge	1130
	Net	0
WAIS	Accumulation	940
	Discharge	1015
	Net	-75

Table S7: Reference values assumed for 2000-2010, where WAIS includes the Antarctic Peninsula.

In the elicitation workshops, there was extensive discussion of how to define ice sheet contributions to sea level over future periods of time in relation to the temperature rise trajectories shown in Fig. S1. It was agreed that the ice sheet contributions would be expressed as anomalies from the 2000-2010 mean mass change states, as pre-defined in Table S7. On this basis, the net baseline sea-level contributions for this period were prescribed as 0.76 mm yr^{-1} for overall SLR, and 0.56, 0.20, and 0.00 mm yr^{-1} for GrIS, WAIS, and EAIS, respectively. (The resulting joint contribution of the three ice sheets is close to an observationally-derived value of 0.79 mm yr^{-1} for the same period, which was published subsequently to the SEJ workshops (4)).

For the SLR results presented in the main text, baseline contributions -- integrated over the relevant time periods (i.e. from 2000CE to 2050CE; 2100CE; 2200CE and 2300CE) -- have been added to the elicited SLR values reported in Supplementary note 1.

50th percentile exceedance	**Probability**
Strong positive	0.8
Weak positive	0.6
Independent	0.5
Weak negative	0.4
Strong negative	0.2
95th percentile exceedance	**Probability**
Strong positive	0.8
Weak positive	0.5
Independent	0.05
Weak negative	0.03
Strong negative	0.01

Table S8: Reference probabilities assumed for estimating central and tail dependencies.

Supplementary note 6: Prompts for discussion of rationales

Some of the questions below give you options for answering that are not independent (e.g., on the second question, buttressing is not independent of hydrofracturing). In such cases, indicate the option that best captures your overall judgment. In cases where you feel more than one answer is absolutely necessary to best characterize your judgment, feel free to fit in more than one response. Where ***changes*** are referred to and a future period is specified, these are for the difference between the future period and the base period, 2000-2010.

Mass change observations and assumptions

- Are the recent ~decadal trends in mass balance largely due to internal variability of the atmosphere/ice/ocean/climate system or anthropogenic forcing, for each ice sheet overall?

Sheet	GrIS		WAIS		EAIS	
	IV	EF	IV	EF	(no trend)	(no trend)

Dynamical processes

- How will ***changes*** in near-field gravitational and vertical land motion due to past and future ice sheet unloading affect marine ice sheet instability: Decrease instability (D), Increase instability (I), or No significant change (N)?

Sheet	GrIS	WAIS	EAIS
D			
I			
N			

Among buttressing by ice shelves (B), basal traction (BT), transverse stresses (TS), hydrofracturing (HF), ice cliff instability (IC), and dissipation after iceberg formation at exit gates (DI), which one will be the most important for controlling the overall 21st, 22nd, and 23rd century discharge rate and grounding line migration for key ice streams and outlet glaciers (recognizing time variations in the role of each)? Key ice streams are those that you expect to control overall discharge for that ice sheet.

2°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
B									
BT									
TS									
HF									
IC									
DI									

5°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
B									

BT									
TS									
HF									
IC									
DI									

Surface mass balance

- Between atmospheric circulation/moisture transport ***changes*** (AM) and albedo ***changes*** (AC), which do you consider more important for determining surface mass balance of grounded ice during the 21st, 22nd, and 23rd century.

2°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
AM									
AC									

5°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
AM									
AC									

- Among ***changes*** in summer sea ice extent (SI), atmospheric circulation/moisture transport ***changes*** (AM), and albedo ***changes*** (AC), which do you consider most important for determining surface mass balance and rate of thinning of ice shelves in 21st, 22nd, and 23rd century?

2°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
SI									
AM									
AC									

5°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
SI									
AM									
AC									

Ocean processes

- Among Antarctic circumpolar current ***changes*** (ACC), ***changes*** in intrusion of circumpolar deep water onto continental shelf (CDW), and ***changes*** in AMOC (MOC), which do you consider will have the largest effect on sub-shelf basal melt rates during the 21st, 22nd, and 23rd century?

2°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
ACC									
CDW									
MOC									

5°C scenario

Sheet	GrIS			WAIS			EAIS		
	21st	22nd	23rd	21st	22nd	23rd	21st	22nd	23rd
ACC									
CDW									
MOC									

Polar Amplification

Please provide the polar amplification factor (e.g., 1.5x, 2x) or range of factors that you used in your estimates.

	2050		2100		2200		2300	
	1.5°C	2.0°C	2°C	5°C	2°C	5°C	2°C	5°C
North								
South								

Low-probability, high-consequence scenarios

Are there high-outcome scenarios above the 95% values you provided that deserve attention? If so, what are they?

Supplementary Note 7: Ambiguity relating to discharge versus sea level contribution

The questionnaire provided to experts asked for their estimate of changes in discharge (defined as the ice flux across the grounding line) that would contribute to SLR. For ice grounded below sea level, such as in large sectors of the WAIS and parts of the EAIS, the change in volume of discharge and the sea level contribution are not the same quantity. This is because it is only the volume above flotation (VAF) that contributes to SLR, while the change in discharge includes ice below flotation that will be displaced by sea water.

This issue was identified during the second SEJ workshop held in Europe and to address it we asked all experts what value they were using for discharge: total discharge or VAF (the same as sea level equivalent). Of the 22 experts, four stated they had calculated total discharge and the rest VAF. Of these four, one had a high calibration score and a strong weighting in the PW01 solutions and a correction to these discharge values for the WAIS and EAIS were considered necessary. To do this, we utilized the output of a thermomechanical ice sheet model coupled to a solid earth deformation model in a climate forced deglaciation experiment (11, 12) and calculated the ratio of total discharge to VAF. This is shown in Figure S12 alongside the gradients for the ratio for the WAIS, EAIS and AIS. For the WAIS, the ratio changes after a volume loss of about 1 m sea level equivalent (SLE), while for the EAIS it is relatively constant. For the EAIS discharge we used a constant ratio, while for the WAIS it varied as a function of the discharge anomaly. The change in gradient is only significant for the 2300 L and H and 2200 H scenarios, where WAIS discharge anomaly exceeds 1 m for this expert.

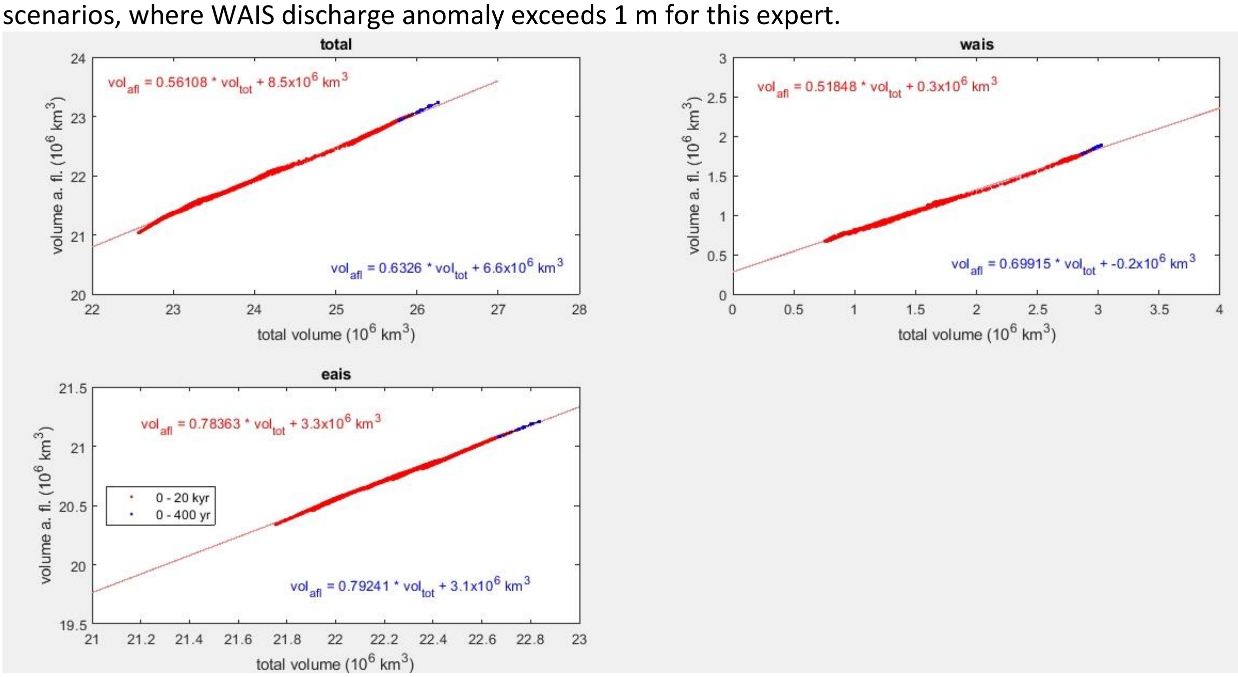


Figure S12. The ratio of volume above flotation to total ice discharge for a present-day deglaciation experiment for the Antarctic ice sheet. Fig a) AIS, b) WAIS, c) EAIS. Blue dots represent the first 400 years, red dots are for the remaining 20 Kyr.

Supplementary Note 8: Elicitation questions

The Elicitation questions are available as a pdf and Excel file at https://doi.org/10.5523/bris.23k1jbtan6sjv2huakf63cqgav

Supplementary References

1. Cooke RM (1991) *Experts in Uncertainty-Opinion and Subjective probability in science.* (Oxford University Press).
2. Colson AR & Cooke RM (2017) Cross validation for the classical model of structured expert judgment. *Reliab. Eng. Syst. Saf.* 163:109-120.
3. Colson AR & Cooke RM (2018) Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. *Review of Environmental Economics and Policy* 12(1):113-132.
4. Eggstaff JW, Mazzuchi TA, & Sarkani S (2014) The effect of the number of seed variables on the performance of Cooke's classical model. *Reliability Engineering & System Safety* 121:72-82.
5. Flandoli F, Giorgi E, Aspinall WP, & Neri A (2011) Comparison of a new expert elicitation model with the Classical Model, equal weights and single experts, using a cross-validation technique. *Reliability Engineering & System Safety* 96(10):1292-1310.
6. Cooke R (1994) Uncertainty in dispersion and deposition in accident consequence modeling assessed with performance-based expert judgmen. *Reliability Engineering & System Safety* 45(1-2):35-46.
7. Wadge G & Aspinall WP (2014) A Review of Volcanic Hazard and Risk Assessments at the Soufrière Hills Volcano, Montserrat from 1997 to 2011. *The Eruption of Soufrière Hills Volcano, Montserrat, from 2000 to 2010*, ed G. Wadge REARaBV (Geological Society, London), Vol 39.
8. Bedford T & Cooke RM (2001) Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32(1-4):245-268.
9. Bedford T & Cooke RM (2002) Vines - A new graphical model for dependent random variables. *Annals of Statistics* 30(4):1031-1068.
10. Edwards TL, *et al.* (2019) Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature* 566(7742):58-64.
11. Konrad H, *et al.* (2014) The Deformational Response of a Viscoelastic Solid Earth Model Coupled to a Thermomechanical Ice Sheet Model. *Surveys in Geophysics* 35(6):1441-1458.
12. Konrad H, Sasgen I, Pollard D, & Klemann V (2015) Potential of the solid-Earth response for limiting long-term West Antarctic Ice Sheet retreat in a warming climate. *Earth and Planetary Science Letters* 432:254-264.