

# Expert judgement and uncertainty quantification for climate change

Michael Oppenheimer<sup>1\*</sup>, Christopher M. Little<sup>2</sup> and Roger M. Cooke<sup>3,4</sup>

**Expert judgement is an unavoidable element of the process-based numerical models used for climate change projections, and the statistical approaches used to characterize uncertainty across model ensembles. Here, we highlight the need for formalized approaches to unifying numerical modelling with expert judgement in order to facilitate characterization of uncertainty in a reproducible, consistent and transparent fashion. As an example, we use probabilistic inversion, a well-established technique used in many other applications outside of climate change, to fuse two recent analyses of twenty-first century Antarctic ice loss. Probabilistic inversion is but one of many possible approaches to formalizing the role of expert judgement, and the Antarctic ice sheet is only one possible climate-related application. We recommend indicators or signposts that characterize successful science-based uncertainty quantification.**

Managing the risks of climate change requires a consistent and comprehensive approach to quantifying uncertainty and a clear narrative to describe the process. As economist Charles Kolstad noted, such efforts are neither new nor confined to the climate arena: “Uncertainty affects many different kinds of agents in the world — including governments — and there are a whole host of instruments that have already been set up to deal with these uncertainties. We don’t need to eliminate uncertainty — uncertainty is fine as long as it’s quantified”<sup>1</sup>.

Process-based models (PBMs) often form the sole basis for uncertainty quantification of climate projections. Such models incorporate operative physics at scales that are manageable from a computational and data acquisition viewpoint. However, some climate projection uncertainties — variously termed model, structural, deep<sup>2</sup> or even wicked — take the scientific community outside its comfort zone. As we discuss below, these uncertainties cannot be tightly constrained with observations; as such, strictly speaking, PBMs cannot be validated. A variety of types of formalized expert judgement (see Box 1), some with greater rigour than others, has played an only limited role in climate-change-related assessments of various physical hazards<sup>3</sup> where deep uncertainty prevails<sup>4–9</sup>. In contrast, it has been a mainstay in other areas of risk analysis since 1975.

In the climate change arena, it has proved difficult to reconcile formalized expert judgement with PBMs, as the projections and uncertainty estimates based on each can often be substantially different. It is even more difficult to determine why they differ. Although the Intergovernmental Panel on Climate Change (IPCC) has provided guidance documents on combining multi-model climate projections and on characterizing and communicating uncertainty<sup>10–12</sup>, IPCC recently declined to provide general guidance for combining distinct lines of evidence arising from, for example, expert judgement and PBMs<sup>13</sup>. There is, however, a history of successful efforts in other areas. The nuclear sector developed techniques for melding structured expert judgement (SEJ), a type of formalization; see Box 1) with the complex suite of models used to predict the consequences of a nuclear accident<sup>14</sup>. A problem for risk analysts was that ‘domain experts’ (for example, specialists in atmospheric dispersion as opposed to whole-system modellers) working with high-fidelity

codes would not quantify uncertainty on the parameters of the less-complex physical representations, which was required for large models describing the whole system. The solution was to elicit probability distributions from these experts on observables predicted by the models and to tailor distributions on the model’s parameters so as to reproduce the elicited expert distributions. This process has come to be known as probabilistic inversion (for further details, see the Supplementary Information).

Here, we suggest that a similar strategy might allow a more seamless and consistent blending of PBMs and expert judgement, resulting in an improvement in climate uncertainty quantification. The general approach recognizes that for many aspects of the climate problem, scientists are operating in an environment where direct constraints on model behaviour are limited and the models’ predictive value is open to question. Accordingly, this method introduces constraints based on the general understanding of the physical system on which the experts draw and from which the models are derived. This converts the question ‘Where are our knowledge gaps?’ to ‘How can we shape what we do know from multiple lines of evidence into a coherent representation of probability associated with a particular physical hazard?’

We first review the limitations of PBMs in quantifying uncertainty and the pervasiveness of expert judgement in attempts to do so. We next summarize IPCC’s use of expert judgement in combination with PBMs to assess climate sensitivity and the ice sheet contribution to sea-level rise (SLR). We then go on to illustrate the utility of probabilistic inversion for fusion of PBMs and expert judgement in projecting behaviour of the Antarctic ice sheet (AIS). Finally, we discuss the generalization of this approach and propose preliminary minimal criteria or signposts to its implementation.

## Characterizing uncertainty

There is a wide gap between the temporal and spatial scales at which the governing laws of physics and chemistry can be solved in a practical sense, and what is required for climate projections (that is, solutions from local to global scale for long lead times). PBMs thus inevitably involve parameters and model structures that are not well anchored in measurements, and measurements (where available) entail their own uncertainties<sup>15</sup>.

<sup>1</sup>Department of Geosciences and Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey, 08544, USA.

<sup>2</sup>Atmospheric and Environmental Research, Inc., Lexington, Massachusetts, 02421, USA. <sup>3</sup>Resources for the Future, 1616 P St NW, Washington DC, 20036, USA. <sup>4</sup>Strathclyde Business School, University of Strathclyde, 199 Cathedral Street, Glasgow G4 0QU, UK. \*e-mail: [omichael@princeton.edu](mailto:omichael@princeton.edu)

**Box 1 | Expert, and structured expert, judgement.**

Expert judgement encompasses a wide variety of techniques ranging from a single undocumented opinion, to preference surveys, to formal elicitation with external validation<sup>55</sup>. In the nuclear safety area, Rasmussen *et al.*<sup>56</sup> formalized expert judgement by documenting all steps in the expert elicitation process for scientific review. This made visible wide spreads in expert assessments and raised questions regarding the validation and synthesis of expert judgements. A critical review endorsed the use of expert subjective probabilities<sup>57</sup> and ushered in widespread applications in nuclear risk assessment. The nuclear safety community later took on board expert judgement techniques driven by external validation<sup>14</sup>. Other blue-ribbon advisory panels have subsequently endorsed these techniques. In a seminal report<sup>58</sup>, the Committee on Risk Assessment of Hazardous Air Pollutants (National Research Council) called for quantitative uncertainty analysis as “the only way to combat the false sense of certainty which is caused by a refusal to acknowledge and (attempt to) quantify the uncertainty in risk predictions.” The US Environmental Protection Agency has advised that “the rigorous use of expert elicitation for the analyses of risks is considered to be quality science”<sup>59</sup>, and endorsed expert elicitation as “...well-suited for challenges with complex technical problems, unobtainable data, conflicting conceptual models, available experts, and sufficient financial resources”<sup>60</sup>.

Efforts to quantify uncertainties in PBMs are dominated by perturbed physics ensembles (PPEs)<sup>4,16,17</sup>. This method has been generalized in the multi-model ensemble (MME) technique, which treats the predictions of individual Atmosphere–Ocean General Circulation Models (AOGCMs) in a way similar to those of individual members of a PPE. Both are aligned with climate observations using Bayesian updating.

A simple example illustrates the assumptions in both ensemble approaches and the difficulty of using them to quantify uncertainty. Consider throwing a pair of dice that we suspect of being loaded, but with very limited information on how the loading was implemented. The dice may be loaded in the same way or different ways; we have a prior probability covering all possibilities. Suppose, however, that we can only observe repeated throws of one of the dice: what can we learn about the loading of the other, a structural uncertainty? The outcome of tossing the dice represents only two random variables. In the case of the climate, we have many, many more random variables (for example, local atmospheric and oceanic properties) of which we have repeated high-quality observations for only a few, and we do not fully understand the underlying physics (analogous to imbalances in the dice). Thus Bayesian methods have limited usefulness in the context of structural uncertainty.

In using ensemble approaches to characterize climate uncertainty, expert judgements fill this gap. In PPEs, expert judgement enters directly in the determination of prior distributions of model parameters. Current practice uses the modeller’s own uncertainty distributions (see below). Expert judgements are implicit in inference from MMEs as well<sup>18</sup>. In both PPEs and MMEs, results are sensitive to the choice of observational constraints (temperature or precipitation, or their changes or extreme behaviour; behaviour of modes of climate, etc.) and in the appropriateness and comprehensiveness of the observations employed in the likelihood function as an analogue for future change (that is, the out-of-sample issue). Most important is the common assumption that the ensemble spans the range of possible model structures.

Rougier *et al.*<sup>19</sup> approached this problem by making either the strong assumption that the truth and our models are all exchangeable,

External validation is the hallmark of science, and expert judgement techniques based on external validation are here termed structured expert judgement (SEJ). They have been used extensively in areas ranging over nuclear safety, investment banking, volcanology, public health, ecology and aeronautics/aerospace (for an overview of applications, see refs 61–76). In SEJ, experts quantify their uncertainty on potentially observable variables of interest, and on calibration variables from their field whose true values are known *post hoc*. Performance on calibration variables is used to construct performance-weighted combinations of the experts’ judgements. Among the insights emerging from studies using SEJ with calibration variables are: (i) experts’ statistical accuracy and informativeness (the ability to concentrate high probability in small regions) is very uneven, ranging from informative and statistically accurate to very overconfident<sup>61</sup>; (ii) both equal-weight and performance-based combinations of individual experts’ distributions generally result in improved statistical accuracy, and for equal weighting this improved accuracy is often purchased at the expense of very wide confidence bands; (iii) statistical accuracy and informativeness are often antagonistic — the most informative experts are also the least accurate — although many expert panels contain accurate and informative individuals; and (iv) performance weighting yields better performance, both in- and out-of-sample, than weighting schemes not based on performance<sup>61–63</sup>.

or the weaker assumption of co-exchangeability: all models are equally successful or deficient at representing reality. A ‘discrepancy’, a measure of structural error, characterizes the relation between the models and reality. The challenge is then to represent this discrepancy<sup>20–23</sup>. These methods, usually working with statistical emulators, are only useful if we assume that the structural error can be represented by a parametric distribution function. Finding this distribution creates a significant role for expert judgement in determining the magnitude and shape of the discrepancy.

Even with PBM-based approaches to uncertainty quantification, expert judgement cannot be expunged. Current efforts thus perform a useful service of circumscribing the range of possible outcomes based on variations of models and model parameters. They cannot address the question of whether these model runs bracket the truth<sup>24</sup>.

**IPCC efforts to incorporate expert judgement**

In its assessments, the IPCC has acknowledged the limitations of ensembles of PBMs, and has attempted to include expert judgement in projections of some climate system properties. However, the use of expert judgement has been application dependent, controversial and implemented differently across chapters and reports<sup>25,26</sup>.

The estimation of equilibrium climate sensitivity, described in detail in IPCC’s Fifth Assessment Report (AR5)<sup>13</sup> and elsewhere<sup>2</sup>, is perhaps the best-developed example so far of attempts to fuse different lines of evidence in the context of climate projections. AR5 incorporated instrumental, climatological, palaeo-climatological and model-based evidence as well as combinations of these lines of evidence — such as models weighted by climatology. AR5 authors then explicitly but informally applied their own expert judgement in deriving a likely (17–83%) range of 1.5–4.5 °C and also partly characterizing the tails of the sensitivity distribution. It is comforting that assessment using very different types of evidence yields good agreement in the central probability range (including studies that are not in any obvious way reliant on AOGCMs). However, individual studies disagree widely in the upper tail (95% confidence limit on sensitivity from individual studies varying from 3.5–9 °C), so expert judgement plays a critical role in any overall assessment. These

**Box 2 | Atmospheric dispersion as an example of probabilistic inversion.**

Running a model with uncertain parameters set at ‘nominal values’ or ‘best guesses’ yields deterministic predictions. Confidence in the outcomes is communicated in an accompanying narrative. Alternatively, quantitative uncertainty analysis assigns a joint distribution to uncertain model parameters, yielding distributions over model predictions. The goal is to take account of uncertainties in the model parameter values and, to some extent, even uncertainties about the model itself (for example, structural uncertainties). The question vexing earlier practitioners was how to acquire these joint distributions over model parameters. Distributions supplied by the modellers themselves were used initially, but the modellers were not always representative of the larger scientific community and this process lacked the desired transparency and verifiability. Querying domain experts about the parameters of a model to which they did not subscribe met with resistance. The solution was to query independent domain experts not involved in the model building about observable phenomena predicted by the models. After combining their uncertainty distributions over these observables, a distribution over model parameters was sought that would replicate the experts’ distributions.

Atmospheric dispersion provides a simple case where independent expert uncertainty quantification has been empirically validated and used with probabilistic inversion to constrain model parameters. It also illustrates the difference between uncertainty quantification by independent experts versus quantification by modellers. This case study is described briefly below and in more detail in the Supplementary Information.

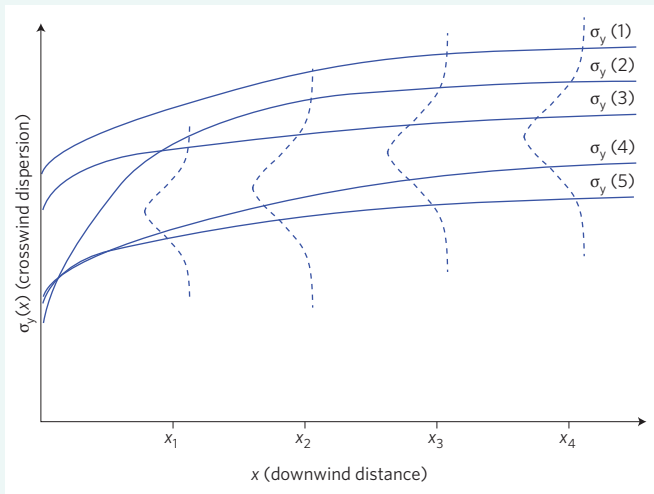
Under ideal conditions, a neutrally buoyant contaminant released in a constant wind field spreads in the crosswind and vertical directions at a rate proportional to the root of downwind distance, according to the simple Gaussian model. In reality, these ideal conditions do not apply. To address this deficiency within the context of a Gaussian model, the stability of the atmosphere and the crosswind and vertical diffusion coefficients ( $\sigma(x)$ ) can be ‘parameterized’ as functions of downwind distance  $x$  and other ambient variables. However, uncertainty quantification is challenged by the fact that independent experts are reluctant to quantify their uncertainty on the parameterized diffusion coefficients.

To address this difficulty, experts were asked to quantify their uncertainty on measurable quantities — such as crosswind dispersion at various downwind distances after a release under stipulated conditions. The expert judgements were combined into a single probability distribution using weights derived from their performance based on comparisons with measured values. Model parameters are

then forced into alignment with this distribution with probabilistic inversion, as shown schematically in the figure below.

The combined probability distributions of crosswind dispersion are shown at four distances from the source by the dashed lines. Probabilistic inversion draws a large sample of values (solid lines) from a broad distribution of diffusion coefficients. Weights are then assigned to each coefficient such that when the initial distribution is re-sampled using these weights, the dashed distributions are optimally recovered. If the probabilistic inversion problem is feasible, an optimal set of weights (in the sense of minimally departing from the starting distribution) is quickly found. This distribution typically introduces complex dependencies and is given numerically.

The confidence bands for observable quantities estimated by modellers were almost always narrower than those of the independent experts (see Supplementary Information and ref. 54). As the combined expert distributions are statistically accurate when compared to observed values, the models with parameter distributions obtained with probabilistic inversion are similarly more accurate than the model outputs.



**Combined expert assessments of uncertainty on crosswind dispersion  $\sigma_y(x)$  at four downwind distances.** Solid lines give  $\sigma_y(x)$  for five different values of model parameters A and B. Downwind distances ( $x_1$ – $x_4$ ) are indicated with dashed lines. See Supplementary Information for more details.

**Table 1 | Summary results for the nine variables involved in the probabilistic inversion.**

|          | D (WAIS) | D (EAIS) | D (TOT) | SMB (WAIS) | SMB (EAIS) | SMB (TOT) | SLR (WAIS) | SLR (EAIS) | SLR (TOT) |
|----------|----------|----------|---------|------------|------------|-----------|------------|------------|-----------|
| Mean     | 6.8      | 4.3      | 11.1    | –2.8       | –3.8       | –6.6      | 3.9        | 0.6        | 4.5       |
| Variance | 33.8     | 13.1     | 47.2    | 0.1        | 0.1        | 0.4       | 33.9       | 13.2       | 47.5      |
| Corr.    | 0.8      | 0.5      | 1.0     | 0.1        | 0.1        | 0.1       | 0.8        | 0.5        | 1.0       |

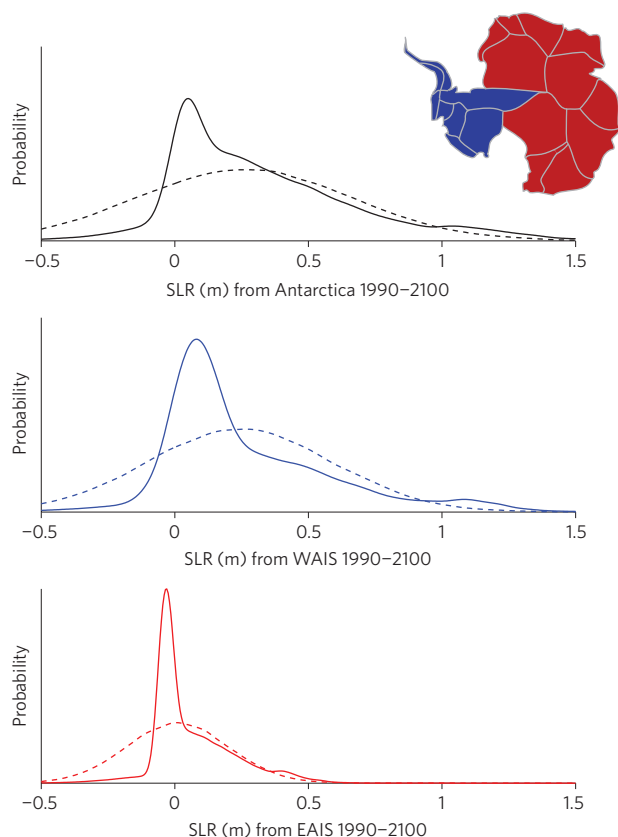
D, discharge (mass flux across Antarctic grounding lines); SMB, surface mass balance; SLR, D + SMB (positive values indicate a net contribution to sea level). The third row is the correlation coefficient for each variable and the total sea-level contribution from Antarctica. All values are in  $\text{mm yr}^{-1}$  for the year 2100.

low-probability outcomes may be the key to risk management<sup>3,13,27,28</sup>. Despite this experience, AR5 authors eschewed inclusion of formalized expert judgement in evaluating climate sensitivity “because the experts base their opinion on the same studies as we assess”<sup>213</sup>.

SLR projections require the use of alternatives to PBMs, as process studies and model intercomparisons (for example, Pattyn *et al.*<sup>29</sup>) are just beginning to illuminate the physical and numerical requirements

for the adequate representation of decadal-to-century timescale mass changes of ice sheets. The current generation of continental-scale models does not represent many of these processes<sup>30,31</sup> and exhibits considerable spread even when forced with identical boundary conditions and climate forcing<sup>29,32,33</sup>.

The incorporation of lines of evidence not tied to PBMs to address these uncertainties — for example, SEJ<sup>34</sup> and semi-empirical



**Figure 1 | The effect of probabilistic inversion on SLR probability distributions. a–c.** Comparison of the probability distributions of cumulative sea level contribution (1990–2100, in metres) for the AIS (a) WAIS (b) and EAIS (c) both before (dashed lines) and after (solid lines) the probabilistic inversion.

methods<sup>35</sup> — has varied widely across IPCC reports and has proven difficult and controversial<sup>36,37</sup>. Although non-PBM-based projections of ice sheet mass balance were discussed in the AR5, there was no quantitative incorporation of estimates of SLR from the SEJ of Bamber and Aspinall<sup>38</sup> reported in 2013 (hereafter referred to as BA13), noting solely that expert estimates from marine-based sectors of the AIS have a wide spread. Instead, AR5 authors used the projections of Little *et al.*<sup>39</sup> (hereafter referred to as L13) for Antarctic discharge, in which subjective judgement was imposed earlier in the process (on drainage-basin-specific ice discharge growth rates and their covariance, rather than sea-level contribution). Another layer of expert judgement was then imposed in the AR5, *ad hoc*, by converting the L13 5–95 percentile range into the AR5 17–83rd percentile range. This re-categorization of the uncertainty range recognized the possibility of more rapid mass loss than taken into account in L13, and implied that the possibility of such an outcome (for example, rapid collapse of part of the ice sheet) lay outside the 17–83rd ‘likely’ range, but provided no further quantification of its likelihood.

Subsequent effort to fuse PBMs and SEJ across the complete range of probabilities has proven difficult. Kopp *et al.*<sup>40</sup> fused the tail structure from BA13 with that from AR5 to produce a more complete estimate of the ice sheet contribution. However, this was done *ad hoc* without using a generalizable approach; consequently, the central range may not be internally consistent with the appended tail. Jevrejeva *et al.*<sup>41</sup> simply substituted the BA13 projections of ice loss for those generated by AR5.

IPCC’s meta-assessment of climate sensitivity is fairly transparent about the reasoning behind its judgements. However, its method

of arriving at judgements is unclear: did the judgements arise from deliberations among two experts? Ten experts? Was the membership of the group consistent or did it vary over time? Were certain views eliminated from consideration for lack of consensus, for example, what opinions were held about the right-hand tail? Did all experts access the same information? Perhaps most importantly, do we understand the conditions under which such techniques have been successful and when they have failed? The lack of such clarity about the process is especially troubling. In addition, the issue of independence of lines of evidence arises in any process of statistical combination, formal (for example, Bayesian updating used in many of the studies above) or informal (such as IPCC’s), which one supposes to be internally consistent. This problem only worsens for situations such as SLR projection, dependent as it is on ice sheet behaviour where the lines of evidence available are fewer and each is less compelling than for climate sensitivity, and the whole distribution is likely to be heavily dependent on expert judgement.

### Merging climate projections with probabilistic inversion

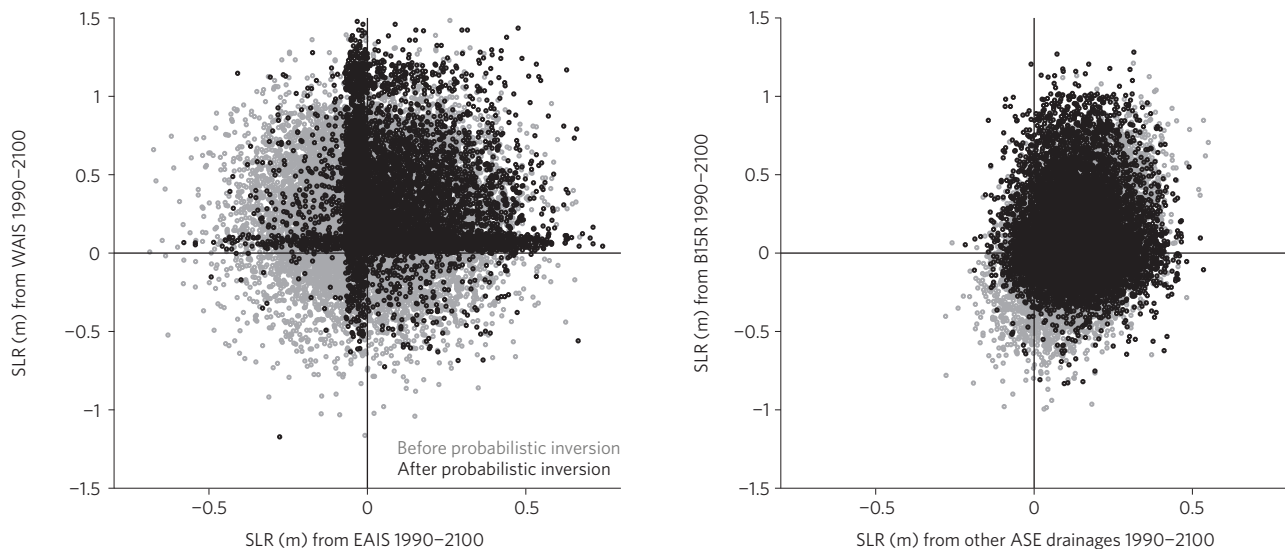
If the same quantities are predicted by a PBM and SEJ, then can we constrain the parameter space of the PBM using the uncertainty distribution derived from SEJ? If so, we will have developed an uncertainty quantification on PBM parameters anchored in evidence that is largely independent of the model (independence may not be perfectly achievable due to the diffusion of expert knowledge) and that itself is subject to empirical validation. In this context, SEJ applies to quantities with which experts have some familiarity (for example, SLR). It does not apply to parameters whose meaning depends on models to which the experts may not subscribe, for example, friction coefficients at the base of an ice sheet.

Probabilistic inversion, the operation of inverting a function or a model at a distribution<sup>42,43</sup> (for a more detailed and precise definition, see the Supplementary Information), is one way of doing this. Box 2 describes a simple example (atmospheric dispersion) and graphically demonstrates the probabilistic inversion process; the remainder of this section demonstrates the use of this technique in a deeply uncertain climate problem: twenty-first-century ice loss from Antarctica.

BA13 (ref. 38) and L13 (ref. 39) are two recent attempts to quantify ice sheet mass loss. In an application of SEJ, BA13 elicited 5th, 50th and 95th percentile mass balance projections for the East Antarctic ice sheet (EAIS) and the West Antarctic ice sheet (WAIS). In contrast, in L13, prior probability distributions were assigned to the growth rates of: (i) ice discharge ( $D$ , mass flux across Antarctic grounding lines); and (ii) surface mass balance (SMB; roughly equivalent to snowfall accumulation) for 19 separate drainage basins based on the authors’ assessment of regional constraints on climate and ice dynamics. Monte Carlo sampling was used to derive probability distributions of mass balance changes over different sectors of the ice sheet.

Here, probabilistic inversion is used to invert the year 2100 mass balance projections elicited in the SEJ of BA13 onto the model of L13. Using the iterative proportional fitting (IPF) algorithm described and implemented in the Supplementary Information, each of the Monte-Carlo samples ( $n = 100,000$ ) from L13 is assigned a weight resulting in year 2100 5th/50th/95th percentile values of  $D$ , SMB and SLR for the EAIS, WAIS and entire AIS identical to those of BA13. These weights are used to construct joint distributions on  $D$  and SMB growth rates, their covariance and the mass balance baseline at the basin scale (in contrast to L13, here we focus only on larger aggregations of drainage basins).

Importantly, our analysis requires a widening of the ‘base case’ prior assumptions in L13 to accommodate the SLR projections derived from SEJ. L13 accounted for the possibility of alternative priors using several sensitivity tests. Here, we expand the range of linear growth rates in Antarctic marine-based basins, especially



**Figure 2 | Scatter plots of 10,000 randomly selected samples from before (grey) and after (black) the probabilistic inversion. a,b**, Samples are identified by their cumulative EAIS and WAIS contribution (**a**) and their cumulative Amundsen Sea Embayment (ASE) and B15R contribution (**b**).

the combined discharge of the Thwaites, Smith and Kohler glaciers (B15R) and Pine Island Glacier, to reflect a broader set of prior assumptions on discharge growth rates ( $5.8\% \text{ yr}^{-1} \pm 5.8\% \text{ yr}^{-1}$  (mean  $\pm$  s.d.) for Pine Island Glacier and  $3.8\% \text{ yr}^{-1} \pm 5.8\% \text{ yr}^{-1}$  for B15R). The mean discharge growth rate of all other marine-based basins (as defined in L13) is increased by 50% of the 1974–2008 historical linear rate of discharge increase in Pine Island Glacier. These adjustments permit significantly higher rates of discharge growth from the Amundsen Sea sector, consistent (in spirit) with recent process-based modelling<sup>31,44</sup> and observations<sup>45</sup>.

With these adjustments, the SLR distribution of year 2100 from the L13 model matches the quantities elicited in BA13 (Table 1). Probabilistic inversion changes the year 2100 5th/50th/95th percentile SLR projections by  $+5.8/-2.3/+3.4 \text{ mm yr}^{-1}$  for WAIS and  $+4.6/-0.4/+1.1 \text{ mm yr}^{-1}$  for EAIS. The largest change is in the low (left-hand) tail, where the approximately normal distributions of L13 are in stark disagreement with the positively skewed results of BA13. In this analysis, the total SLR from Antarctica is largely determined by WAIS discharge (correlation coefficient 0.8); however, the contribution of EAIS is non-negligible (0.5).

With the assumption of linear growth rates, we can present results in terms of the cumulative sea-level contribution (Fig. 1). As noted by de Vries *et al.*<sup>46</sup>, a super linear rate of growth would satisfy the BA13 rates with a smaller cumulative SLR contribution (for example, see the quadratic assumptions in refs 47–49). We leave exploration of this issue for future work, noting that elicitation could also be designed to target the cumulative contribution.

The 95th percentile values for the cumulative 1990–2100 SLR from WAIS, EAIS and AIS (100, 38 and 104 cm, respectively) are close to that of BA13 (which also assumed linear growth); differences arise due to a different baseline, the assumed inter-basin correlations, and the projection period. Probabilistic inversion transforms the normal distributions closer to a lognormal distribution; the low tail is cut off, there is more probability mass in the centre and the high tail is shifted to match the 95th percentile. However, unlike other analyses<sup>38,42,43</sup>, the form of the distribution is not specified; it is a function of the elicited quantiles, which are not bound to a specific form.

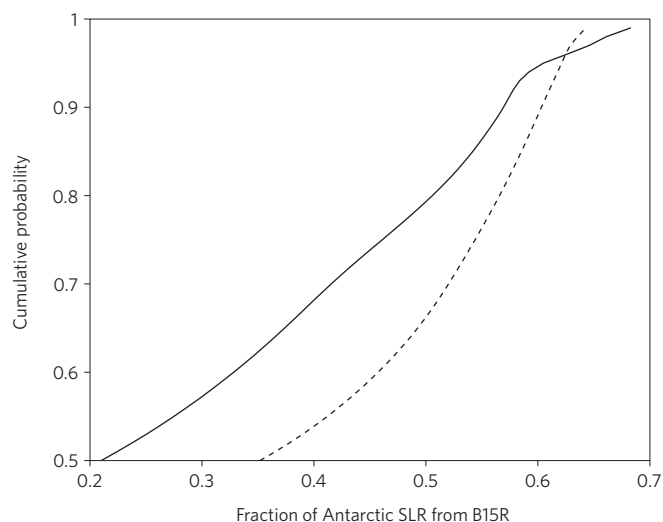
At the largest scale, the inverted distributions must satisfy the dual constraints of SLR from the WAIS and EAIS. Figure 2a indicates that there are two favoured ‘pathways’ of ice loss implied by the elicitation and the L13 model: one in which both ice sheets behave

relatively independently and one in which both the EAIS and WAIS exhibit high SLR. Simulations where either the WAIS or EAIS, or both, generate substantial sea-level fall are strongly down-weighted by the inversion. Similar behaviour is shown at a smaller scale in the Amundsen Sea (Fig. 2b), especially in B15R; moderate to high SLR samples are up-weighted. In this analysis, the future behaviour of B15R is shown to control an increasing part of Antarctica’s sea-level contribution as lower probability outcomes are considered (Fig. 3). Probabilistic inversion diminishes the contribution of B15R (consistent with the overall WAIS contribution shown in Fig. 1) over most of the probability density function, but up-weights its contribution above the 92nd percentile to satisfy the elicited high-end WAIS contribution.

### A path forward

For this Perspective, we applied a particular method — probabilistic inversion — aimed at integrating expert judgement with other lines of evidence. In contrast to efforts to fuse projections ‘after the fact’, the probability distribution for Antarctic ice loss shown in Fig. 1 is internally consistent. Inverted prior distributions are generated at a finer spatial scale than would be possible from SEJ alone, and these might be updated with regional PBMs and observational datasets<sup>39</sup>. Alternatively, they may be updated with a subsequent round of probabilistic inversion based on further observations, or with a different set of experts. Designing PBMs and SEJ conjointly would offer the prospect of much richer integration.

This approach has its limits. A probabilistic inverse does not always exist; that is to say, it may not be possible to recover the combined expert uncertainty distributions on observables from a distribution over the PBM’s parameters. In the atmospheric dispersion example (Box 2 and Supplementary Information), the inversion was quite successful<sup>50</sup>. Suppose, however, that the dotted densities for  $\sigma_y(x)$  were decreasing as  $x$  increased, that is, a plume that becomes less diffuse as it moves downwind; in almost all conditions, this behaviour is physically impossible. If the experts had given such distributions, the probabilistic inversion would certainly fail. Alternatively, experts might favour a plume growth rate that cannot be captured by a simple form; or, as in the ice sheet analysis in the previous section, distributions on the discharge growth rate parameters (or assumptions about their linearity) that are quite different than that envisaged by the modeller. If inversion is not feasible then the departure between what we would like and what we can get with



**Figure 3 | Fraction of total Antarctic sea level contribution originating from B15R, before (dashed line) and after (solid line) the probabilistic inversion.**

probabilistic inversion is assessed. The analyst must judge whether the departure is acceptable or if not, whether other models or other experts should be used. Regardless, we claim that the explicit process of comparing expert judgement and PBMs using probabilistic inversion is invaluable and can certainly be extended beyond the ice sheet problem.

Although uncertainty cannot be reduced without acquiring new knowledge, we believe that there is a reserve of such knowledge in expert judgement that can be carefully elicited and implemented so as to improve characterization of uncertainty. Probabilistic inversion would be most usefully applied to problems where there is at least a modest body of observational information, but PBMs are performing poorly. However, for some problems, the availability of geophysical observables is insufficient to usefully constrain the models. For others, non-observational lines of evidence can serve as input for SEJ, for example, experimental understanding of physical processes.

Stepping back from probabilistic inversion to the general problem of uncertainty quantification, we end by suggesting a few signposts pointing towards an informative approach. First, uncertainty quantification should have a component that is model independent. All models are idealizations and so all models are wrong. An uncertainty quantification that is conditional on the truth of a model or model form is insufficient. Second, the method should be widely applicable in a transparent and consistent manner. As already discussed, several approaches to uncertainty quantification have been proposed in the climate context but fall short in their generalizability or clarity. Third, the outcomes should be falsifiable. Scientific theories can never be strictly verified, but to be scientific they must be falsifiable<sup>51</sup>. Whether theories succumb to crucial experiments or expire under a ‘degenerating problem shift’<sup>52</sup>, the principle of falsifiability remains a point of departure. With regard to uncertainty quantification, falsification must be understood probabilistically. The point of predicting the future is that we should not be too surprised when it arrives. Comparing new observations with the probability assigned to them by our uncertainty quantification gauges that degree of surprise. With this in mind, outcomes should also be subject to arduous tests. Being falsifiable is necessary but not sufficient. As a scientific claim, uncertainty quantification must withstand serious attempts at falsification. Surviving arduous tests is sometimes called confirmation or validation, not to be confused with verification<sup>53</sup>. Updating a prior distribution does not constitute validation. Bayesian updating is the correct way to learn, based

on a likelihood and prior distribution, but it does not mean that the result of the learning is valid. Validation ensues when posterior ‘prediction intervals’ are shown to capture out-of-sample (for example, future) observations with requisite relative frequencies. This is the case for the coefficients in the Gaussian plume model (Supplementary Information and ref. 54). Time will tell whether the uncertainty quantification for ice sheets presented here survives.

Received 3 September 2015; accepted 12 February 2016; published online 27 April 2016

## References

1. *The Economic and Financial Risks of a Changing Climate: Insights from Leading Experts Workshop Report* (AAAS, 2014).
2. Draper, D. Assessment and propagation of model uncertainty. *J. R. Statist. Soc. B* **57**, 45–97 (1995).
3. Oppenheimer, M. *et al.* in *Climate Change: Impacts, Adaptation, and Vulnerability*. (eds Field, C. B. *et al.*) 1039–1099 (IPCC, Cambridge Univ. Press, 2014).
4. Frigg, R., Smith, L. A. & Stainforth, D. A. The myopia of imperfect climate models: the case of UKCP09. *Philos. Sci.* **80**, 886–897 (2013).
5. Knutti, R. & Hegerl, G. C. The equilibrium sensitivity of the Earth’s temperature to radiation changes. *Nature Geosci.* **1**, 735–743 (2008).
6. Morgan, M. G. & Keith, D. W. Subjective judgements by climate experts. *Environ. Sci. Technol.* **29**, 468–476 (1995).
7. Zickfeld, K., Morgan, M. G., Frame, D. J. & Keith, D. W. Expert judgements about transient climate response to alternative future trajectories of radiative forcing. *Proc. Natl Acad. Sci. USA* **107**, 12451–12456 (2010).
8. Church, J. A. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) (IPCC, Cambridge Univ. Press, Cambridge, 2013).
9. Morgan, G. M. *et al.* *Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decisions* Synthesis and Assessment Product 5.2. (US Climate Change Science Program, 2009).
10. Mastrandrea, M. D. *et al.* *Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties* (IPCC, 2010).
11. Knutti, R. *et al.* in *Meeting Report of the Intergovernmental Panel on Climate Change: Expert Meeting on Assessing and Combining Multi Model Climate* (eds Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. & Midgley, P. M.) 1–13 (IPCC, 2010).
12. Moss, R. H. & Schneider, S. H. in *Guidance papers on the cross cutting issues of the Third Assessment Report of the IPCC* (eds R. Pachauri, T. Taniguchi and K. Tanaka) 33–51 (IPCC, 2000).
13. Collins, *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) Ch. 13 (Cambridge Univ. press, 2013).
14. Goossens, L. H. J. & Kelly, G. N. Radiation protection dosimetry expert judgement and accident consequence. *J. Uncertainty Anal.* **90**, 295–301 (2000).
15. Mason, D. & Knutti, R. Predictor screening, calibration, and observational constraints in climate model ensembles: an illustration using climate sensitivity. *J. Clim.* **26**, 887–898 (2013).
16. Murphy, J. M. *et al.* A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Phil. Trans. Royal Soc. A* **365**, 1993–2028 (2007).
17. Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G. A. Challenges in combining projections from multiple climate models. *J. Clim.* **23**, 2739–2758 (2010).
18. Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A* **365**, 2053–2075 (2007).
19. Rougier, J. C., Goldstein, M. & House, L. Second-order exchangeability analysis for multi-model ensembles. *J. Am. Statist. Assoc.* **108**, 852–863 (2013).
20. Goldstein, M. & Rougier, J. Reified Bayesian modelling and inference for physical systems. *J. Statist. Plan. Infer.* **139**, 1221–1239 (2009).
21. Murphy, J. M. *et al.* UK Climate Projections Science Report: Climate Change Projections (Met Office, 2009).
22. Sexton, D. M. H., Murphy, J. M., Collins, M. & Webb, M. J. Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Clim. Dynam.* **38**, 2513–2542 (2012).
23. Sexton, D. M. H. & Murphy, J. M. Multivariate probabilistic projections using imperfect climate models part II: robustness of methodological choices and consequences for climate sensitivity. *Clim. Dynam.* **38**, 2543–2558 (2012).
24. Parker, W. S. Ensemble modeling, uncertainty and robust predictions. *WIREs Clim. Change* **4**, 213–223 (2013).
25. O’Reilly, J., Oreskes, N. & Oppenheimer, M. The rapid disintegration of predictions: climate science, bureaucratic institutions, and the West Antarctic ice sheet. *Social Stud. Sci.* **42**, 709–731 (2012).

26. van der Sluijs, J. *et al.* Anchoring devices in science for policy: the case of consensus around climate sensitivity. *Social Stud. Sci.* **28**, 291–323 (1998).
27. Houser, T. *et al.* *American Climate Prospectus: Economic Risks in the United States* (Rhodium Group, 2014).
28. Hinkel, J. *et al.* Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl Acad. Sci. USA* **111**, 3292–3297 (2014).
29. Pattyn, F. *et al.* Grounding-line migration in plan-view marine ice-sheet models: results of the ice2sea MISIP3d intercomparison. *J. Glaciol.* **59**, 410–422 (2013).
30. Durand, G. & Pattyn, F. Reducing uncertainties in projections of Antarctic ice mass loss. *Cryos. Discuss.* **9**, 2625–2654 (2015).
31. Pollard, D., DeConto, R. M. & Alley, R. B. Potential Antarctic ice sheet retreat driven by hydrofracturing and ice cliff failure. *Earth Planet. Sci. Lett.* **412**, 112–121 (2015).
32. Bindshadler, R. A. *et al.* Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project). *J. Glaciol.* **59**, 195–220 (2013).
33. Nowicki, S. *et al.* Insights into spatial sensitivities of ice mass response to environmental change from the searise ice sheet modeling project ii: Greenland. *J. Geophys. Res. Earth Surf.* **118**, 1025–1044 (2013).
34. Vaughan, D. & Spouge, J. Risk estimation of collapse of the West Antarctic ice sheet. *Climatic Change* **52**, 65–91 (2002).
35. Rahmstorf, S., Perrette, M. & Vermeer, M. Testing the robustness of semi-empirical sea level projections. *Clim. Dynam.* **39**, 861–875 (2012).
36. Oppenheimer, M., O'Neill, B. & Webster, M. Negative learning. *Climatic Change* **89**, 155–172 (2008).
37. Oppenheimer, M. *et al.* The limits of consensus. *Science* **317**, 1505–1506 (2007).
38. Bamber, J. L. & Aspinall, W. P. An expert judgement assessment of future sea level rise from the ice sheets. *Nature Clim. Change* **3**, 424–427 (2013).
39. Little, C. M., Oppenheimer, M. & Urban, N. M. Upper bounds on twenty-first-century Antarctic ice loss assessed using a probabilistic framework. *Nature Clim. Change* **3**, 654–659 (2013).
40. Kopp, R. E. *et al.* Probabilistic 21st and 22nd century sea-level projections at a global network of tide gauge sites. *Earth's Future* **2**, 383–406 (2014).
41. Jevrejeva, S., Grinsted, A. & Moore, J. C. Upper limit for sea level projections by 2100. *Environ. Res. Lett.* **9**, 104008 (2014).
42. Kraan, B. C. P. & Bedford, T. J. Probabilistic inversion of expert judgements in the quantification of model uncertainty. *Manag. Sci.* **51**, 995–1006 (2005).
43. Du, C., Kurowicka, D. & Cooke, R. M. Techniques for generic probabilistic inversion. *Comp. Stat. Data Anal.* **50**, 1164–1187 (2006).
44. Joughin I., Smith, B. & Medley, B. Marine ice sheet collapse potentially under way for Thwaites Glacier Basin, West Antarctica. *Science* **344**, 735–738 (2014).
45. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011. *Geophys. Res. Lett.* **41**, 3502–3509 (2014).
46. de Vries, H. & van de Wal, R. S. W. How to interpret expert judgement assessments of 21st century sea-level rise. *Climatic Change* **130**, 87–100 (2015).
47. Horton, B. P., Rahmstorf, S., Engelhart, S. E. & Kemp, A. C. Expert assessment of sea-level rise by AD 2100 and AD 2300. *Quatern. Sci. Rev.* **84**, 1–6 (2014).
48. *Global sea level rise scenarios for the United States National Climate Assessment* (Climate Program Office, 2012).
49. Little, C. M., Urban, N. M. & Oppenheimer, M. Probabilistic framework for assessing the ice sheet contribution to sea level change. *Proc. Natl Acad. Sci. USA* **110**, 3264–3269 (2013).
50. Jones, J. A. *et al.* *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Uncertainty from the Atmospheric Dispersion and Deposition Module* EUR 18822EN (European Commission, 2001).
51. Popper, K. R. *The Logic of Scientific Discovery* (Hutchinson, 1959).
52. Lakatos, I. *The Methodology of Scientific Research Programmes* Philos. Papers Vol. 1 (Cambridge Univ. Press, 1978).
53. Oreskes, N., Shrader-Frechete, K. & Belitz, K. Verification, validation and confirmation of numerical models in the Earth sciences. *Science* **26**, 641–646 (1994).
54. Cooke, R. M. Uncertainty in dispersion and deposition in accident consequence modelling assessed with performance-based expert judgement. *Rel. Eng. Syst. Saf.* **45**, 35–46 (1994).
55. Aspinall, W. P. & Cooke, R. M. in *Risk and Uncertainty Assessment in Natural Hazards* (eds Hill, L., Rougier, J. C. & Sparks R. S. J.) 64–99 (Cambridge University Press, 2013).
56. Rasmussen, N. C. *et al.* *Reactor Safety Study: An Assessment of Accident Risks in US Commercial Nuclear Power Plants* WASH-1400 (NUREG75/014) (US Nuclear Regulatory Commission, 1975).
57. Lewis, H. *et al.* *Risk Assessment Review Group Report to the US Nuclear Regulatory Commission* NUREG/CR-04000 (Chemical Rubber Company, 1979).
58. *Science and Judgement in Risk Assessment* (The National Academies, 1994).
59. *Guidelines for Carcinogen risk assessment* EPA/630/P-03/001F (US Environmental Protection Agency, 2005).
60. *Expert Elicitation Task Force White Paper* (US Environmental Protection Agency, 2011).
61. Cooke, R. M. & Goossens, L. H. J. TU Delft Expert judgment data base, special issue on expert judgement. *Rel. Eng. Syst. Saf.* **93**, 657–674 (2008).
62. Aspinall, W. P. A route to more tractable expert advice. *Nature* **463**, 294–95 (2010).
63. Eggstaff, J. W., Mazzuchi, T. A. & Sarkani, S. The effect of the number of seed variables on the performance of Cooke's classical model. *Rel. Eng. Syst. Saf.* **121**, 72–82, (2014).
64. Aspinall, W. P. in *Statistics in Volcanology*. (eds Mader, H. M., Coles, S. G., Connor, C. B. & Connor, L. J.) 15–30 (Geological Society, 2006).
65. Cooke, R. M. *et al.* A probabilistic characterization of the relationship between fine particulate matter and mortality: elicitation of European experts. *Environ. Sci. Technol.* **41**, 6598–6605 (2007).
66. Tuomisto, J. T., Wilson, A., Cooke, R. M., Tainio, M. & Evans J. S. Mortality in Kuwait due to PM from oil fires after the Gulf War: combining expert elicitation assessments. *Epidemiol.* **16**, S74–S75 (2005).
67. Evans J. S., Wilson A., Tuomisto J. T., Tainio M. & Cooke R. M. What risk assessment can tell us about the mortality impacts of the Kuwaiti oil fires. *Epidemiol.* **16**, S137–S138 (2005).
68. Cooke, R. M. *et al.* Out-of-sample validation for structured expert judgement of Asian carp establishment in Lake Erie. *Integr. Environ. Assess. Manag.* **10**, 522–528 (2014).
69. Hoffmann, S. *et al.* Research synthesis methods in an age of globalized risks: lessons from the global burden of foodborne disease expert elicitation. *Risk Analysis* **36**, 191–202 (2015).
70. Koch, B. J. *et al.* Suburban watershed nitrogen retention: estimating the effectiveness of storm water management structures. *Elementa: Sci. Anthropocene* **3**, 1–18 (2015).
71. Wittmann, M. E., Cooke, R. M., Rothlisberger, J. D. & Lodge, D. M. Using structured expert judgement to assess invasive species prevention: Asian carp and the Mississippi — Great Lakes hydrologic connection. *Environ. Sci. Technol.* **48**, 2150–2156 (2014).
72. Wittmann, M. E. *et al.* Structured expert judgement to forecast species invasions: bighead and silver carp in Lake Erie. *Cons. Biol.* **29**, 187–197 (2014).
73. Tyshenko, M. G. *et al.* 2010 expert elicitation for the judgement of prion disease risk uncertainties using the classical model and EXCALIBUR. *J. Toxicol. Environ. Health A.* **74**, 261–285 (2011).
74. Gerstenberger, M. C., McVerry, G. H., Rhoades, D. A. & Stirling, M. W. Seismic hazard modeling for the recovery of Christchurch, New Zealand. *Earthquake Spectra* **30**, 17–29 (2014).
75. Christophersen, A., Nicol, A. & Gerstenberger, M. C. *The Feasibility of Using Seed Questions for Weighting Expert Opinion in CCS Risk Assessment* CO2CRC Report RPT11–2868 (Cooperative Research Centre for Greenhouse Gas Technologies, 2011).
76. Gerstenberger, M. C. *et al.* in *11th International Conference on Greenhouse Gas Control Technologies* (eds Dixon, T. & Yamaji, K.) 2775–2782 (Elsevier, 2013).

## Acknowledgements

We thank J. Hall (Oxford University, UK), K. Keller (Pennsylvania State University, USA), R. Kopp (Rutgers University, USA), J. Rougier (University of Bristol, UK), D. Sexton (Met Office Hadley Centre, UK) and C. Tebaldi (National Center for Atmospheric Research, USA) for either helpful comments on an earlier draft of the manuscript, or useful discussions of the issues raised, or both.

## Author contributions

M.O., C.M.L. and R.M.C. designed the research, conducted analysis of data and results, and contributed to writing, editing and revision. C.M.L. and R.M.C. performed statistical modelling.

## Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.O.

## Competing financial interests

The authors declare no competing financial interests.