Climatic Change (2016) 137:329–332
DOI 10.1007/s10584-016-1712-3

RESPONSE

Response to commentary by J. L. Bamber, W. P. Aspinall and R. M. Cooke (2016)

Hylke de Vries[1] · Roderik S. W. van de Wal[2]

Received: 13 May 2016 / Accepted: 23 May 2016 / Published online: 7 June 2016

Abstract In a commentary paper, Bamber et al. (Nat Clim Change 3:424–427, 2016) respond to our recent assessment (De Vries and Van de Wal Clim Change 1–14, 2015) of their expert judgment based study on projections of future sea level rise due to the melting of the large ice sheets (Bamber and Aspinall Nat Clim Change 3:424–427, 2013). In this response we comment on their remarks.

We thank Bamber et al. (2016) [BAC16] for discussing our assessment (De Vries and Van de Wal 2015) [VW15] of their study (Bamber and Aspinall 2013) [BA13]. BAC16 present additional information in support of the approach taken in BA13, but also criticise our re-interpretation of the expert judgment data. Here we reflect on this criticism. We start by restating two principal reasons for publishing VW15.

1 Asking the right question

It all starts with asking the right question. A central goal of BA13 is to derive an expert-based distribution for the cumulative sea level rise arising from the ice-sheets over the 21st century. In VW15 we argue that this distribution can not be inferred straightforwardly from their raw expert data, the main reason being that the experts were not asked that question. Instead, they were questioned about the individual sea-level trend contributions in the year 2100 from the largest three ice-sheets (Greenland, West- and East Antarctica). As a consequence of this mismatch between question and goal, BA13 had to invoke a number of "post-processing" steps to allow the construction of that distribution. We regret that this

This reply refers to the comment available at: http://dx.doi.org/10.1007/s10584-016-1672-7.

✉ Hylke de Vries
hylke.de.vries@knmi.nl

[1] Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

[2] Institute for Marine and Atmospheric Research Utrecht, Utrecht, The Netherlands

point has barely received attention in BA13 and neither in their commentary BAC16. It leaves the reader of BA13 with the believe that what is shown in BA13 is a true reflection of the opinion of "the experts", whereas in practice it is not.

In VW15 we show that the required post-processing steps, made by the authors and therefore at best guided by the experts' answers, have a strong imprint on the final result. It could have been an excellent validation question if BA13 had also questioned the experts about their estimates of cumulative sea level rise in 2100. In fact, it would have been the ultimate test of their own underlying distribution reconstruction assumptions as well as providing a basic internal consistency check of the answers of the experts. The expert evaluation by Horton et al. (2014) is for this reason more straightforward to interpret as for that questionnaire the experts were asked to provide a number for the cumulative sea level. This point is also discussed in the recent paper by Oppenheimer et al. (2016).

2 There is no consensus among experts

A second key message in VW15 is that neither the general public is unanimous about sea-level rise in 2100 nor the community of climate scientists. In fact, not even those climate scientists who consider themselves sea level rise experts or are considered as such by peers or by otherwise defined metrics of their expertise-level. These experts have widely differing views on the future sea level contributions from the large ice sheets, especially the tails of the probability distribution. The origin of this lack of consensus is clear, as some of the fundamental ice-dynamical processes (e.g., calving, hydrofracturing, ice cliff stability, marine ice-sheet instability) are not fully understood. The wide range of expert viewpoints is probably the most notable characteristic of the raw data presented in the supplementary material of BA13. Yet they are not discussed in full depth in BA13.

In VW15 we argue that this lack of consensus needs to be addressed in the interpretation of post-processing of the expert answers. How can one aggregate widely differing expert viewpoints in meaningful numbers? BA13 used the Classical Model approach for that and BAC16 provide ample references to show it is a robust methodology, which we do not aim to question. BA13 and also BAC16 seem however confident that once this "hurdle" of how to weight the different expert opinions has been taken, one can thereafter safely speak about what "The experts think about sea level change in 2100". Given the huge spread in the raw data we like to bring in some nuance.

3 Discussion

In the previous sections we explained two motivations for publishing VW15. Here we give further remarks and reflect on the criticism presented in BAC16.

The "effective" number of experts Upon closer inspection of BA13 the following interesting statement can be found (p10 of their supplementary data): *"And, equally clearly, self-assessed Expertise Level scores on these seed items are completely uncorrelated with the corresponding performance-based weights determined empirically by the Classical model"*. Thus, although experts may well be very convinced they are a true expert in their field, their confidence does not relate to how well they score on the seed questions.

If we study the normalised Cooke weights in the Supplementary Note 2 of BA13, one can see that from the 13 experts, only 5 get a non-negligible weight factor (some even get weight 0.0). Therefore the initial pool of 13 experts decreases drastically after the objective validation of the experts' appropriateness to answer the question of sea level rise. BA13 do not discuss this strong reduction and apparently have no problem with the small effective size of the pool of experts. It is to be questioned where one puts the lower limit for the representativeness of "The Expert". Obviously, had only one expert 'survived' the validation (in the sense of his/her weighting factor being much larger than all the others), the paper would essentially represent the view of this one expert. We believe 5 is a small number and some discussion on this is warranted.

Reproducing pooled averages Unfortunately we were not able to reconstruct the "Perf wts" averages using the raw data and the normalised weight factors given in the Supplementary data of BA13. Either the data has been shuffled across different tables or the Cooke weighting factors differ substantially between 2010 and 2012 such that reconstruction is not possible. Note that the latter would imply that experts with high weights ("true" experts) in the 2010 study can suddenly become "non-experts" in the follow up questionnaire in 2012, or vice versa. Reproducing the "equal wts" results should be straightforward as it is a simple arithmetic mean of the contributing experts. Unfortunately, again we did not manage to reconstruct these numbers from the raw data listed in the tables (we tested the low, central and high estimates of EAIS, WAIS and GRIS). Maybe we have overlooked something, but our conclusion was that either the pooled averages are ok but the raw data is not listed correctly in the published tables, or vice versa. In VW15 we considered both possibilities.

Comments to specific remarks of BAC16 BAC16 come up with a list of criticism to VW15. We are thankful for locations where this critique is constructive such as for example in Section 1 in BAC16. VW15 has never had the purpose to come up with "the" (ultimate) solution. At first we aimed to just crosscheck the BA13 results, and asked the authors further information regarding details by direct personal communication. When this failed we proceeded with our attempts for a reconstruction and detected room for an alternative approach where at least the widely varying opinions of the experts are taken into account as is reflected in VW15.

In their Section 2 ("Consensus distribution") BAC16 provide a theoretical proof of why our methodology will result in quantiles that are essentially averages of the individual expert quantile estimates. Of course; if one follows our line of medians, then by definition 50 % of the experts will have a more conservative estimate, 50 % a more extreme view. We admit that our terminology of a "consensus distribution" is perhaps not the appropriate naming. It is a means of showing as directly as possible (and thereby necessarily as 2-dimensional plots) the impact of the experts differing opinions. We are not convinced that the long tables listed in BAC16 add anything valuable in this respect. In their Section 3 BAC16 argue that we have not explained our definition of "level of consensus" in VW15. What we had in mind is just to show the range of expert opinions at a given percentile estimate, but again we regret any misunderstanding of our terminology.

We appreciate the additional information presented in BAC16 Section 4 ("Lognormal fitting") regarding details of the log-normal distribution, especially about the way the lower and upper bounds are determined by the default intrinsic range. However, it is still unclear

to us how their software package deals with the negative values in conjunction with a log-normal distribution as the latter is principally defined for positive values only (the log of the values is normally distributed). This plays an important role in aggregating the answers for the different ice sheets to a cumulative contribution, which is at the heart of our critical remarks. We therefore completely support the remark in BAC16 that the choice of underlying distribution warrants further study.

4 Conclusion

In their conclusions BAC16 return to their primary argument that using pooled averages based on "validated" experts opinions is superior to using those based on "unvalidated" experts. They conclude with *"the goal of science-based uncertainty quantification is not served by neglecting validation"*. We cannot agree more. VW15 has been intended to open discussion on the inconclusiveness of the experts' opinions, and on the subjectivity introduced by the post-processing of their answers. We believe that BA13 has merit as a study and opened a new but narrow avenue in glaciology, but that it also highlights only one side of the coin. The side that remains mostly hidden contains a discussion on the large discrepancies between the experts, the small effective size of the sample after validation and, finally, the influences of the post-questionnaire decisions needed to come to the cumulative sea level distributions.

References

Bamber JL, Aspinall WP (2013) An expert judgment assessment of future sea level rise from the ice sheets. Nat Clim Change 3:424–427. doi:10.1038/NCLIMATE1778

Bamber JL, Aspinall WL, Cooke RM (2016) A commentary on "how to interpret expert judgment assessments of 21st century sea-level rise. Climatic Change

De Vries H, Van de Wal RSW (2015) How to interpret expert judgement assessments of 21st century sea level rise. Clim Change:1–14. doi:10.1007/s10584-015-1346-x

Horton BJ, Rahmstorf S, Engelhart SE, Kemp AC (2014) Expert assessment of sea-level rise by AD 2100 and AD 2300. Quat Sci Rev 84:1–6

Oppenheimer M, Little CM, Cooke RM (2016) Expert judgement and uncertainty quantification for climate change. Nat Clim Change 6(5):445–451