This paper applies several large ensembles to evaluate to what extent the large en-sembles can overcome limitations of an original approach by Hawkins and Sutton to quantify and illustrate uncertainty in future predictions and attribute it to different uncertainties. It is very well written and very interesting. I like that it also looks at temperature targets which is a nice extension of HS. I have two minor questions that the authors could consider strengthening.
We thank the reviewer for the positive and constructive review. Our replies to the individual points raised are in blue.

The conclusions mention the issue that the SMILEs might not quite span the model range and this kept worrying me through the paper - to what extent could differences between SMILEs and CMIP5/6 be due to a different set of models? Ideally it would be nice to see a bit more discussion of this, or a sensitivity test to use a CMIP subset of runs that did SMILEs. (this might be near impossible with the original method of using only 1 run per model but that seems quite a wasteful approach to me...).
We now provide this sensitivity test for the variables and regions where the HS09 approach still works more or less robustly (Figs. R1-5; not included in the paper). The subsampled CMIP5 results show good correspondence with both the SMILEs and the full CMIP5 results, confirming that the specific SMILE models are a good representation of CMIP5 for the cases investigated here. Further, we provide an overview figure with time series for different variables and regions to demonstrate visually that SMILEs cover much of the range of CMIP5 (new Fig. S4).

Qualitatively, this conclusion is also supported by the fact that the SMILE models happen to be fairly independent of one another with regard to model structure, specifically the atmosphere and ocean components, something that might contribute to their documented relative independence when measured by temperature and precipitation projections (Knutti et al. 2013) or other climate fields (Sanderson et al. 2015).

We have added the following sentence to Section 3.1 in the paper:

- “This holds for other variables and large-scale regions subsequently investigated (Fig. S4), which is also consistent with the coincidental structural independence between the seven SMILEs (Knutti et al. 2013; Sanderson et al. 2015a).”

Also, do you think aerosol uncertainty is sampled well across SMILEs? Not doing so would have implications worth mentioning.
This is a good question, but difficult to answer given that aerosol-only simulations are rare. A literature search reveals that, among the SMILE models used here, GFDL-ESM2M and MPI-ESM are representative of weak aerosol forcing models in CMIP5, and GFDL-CM3 is representative of strong aerosol forcing models, with the other SMILE models, for which aerosol forcing estimates are available, falling in between (Forster et al. 2013; Rotstayn et al. 2015). We tentatively conclude that the SMILE models are not systematically biased in their sampling of aerosol forcing, but a more in-depth analysis might be needed to strengthen this conclusion. This is true in particular for the impact of aerosols on regional hydroclimate, which can be different from its global impact. We have added a sentence about this in the Conclusions:

- “For example, while the seven SMILEs used here cover the range of global aerosol forcing estimates in CMIP5 reasonably well (Forster et al. 2013; Rotstayn et al. 2015), their representativeness for questions of regional aerosol forcing remains to be investigated.”

The point of changing variance of variability in the future (overcoming a particular limitation of HS) is a very interesting one and it doesnt get much space in the paper other than dashed lines in figure 7. could this be brought out a bit more - is the change in variance robust across models?

While we agree that the topic of forced changes in internal variability is interesting, we see an in-depth analysis as beyond the scope of this already lengthy paper. We have, however, investigated the robustness of the forced changes in internal variability across SMILEs for the variables and regions in the paper and found that whenever there was a forced change detectable in the multi-SMILE mean (Fig. 7d-f), all 7 SMILEs agreed on the sign of change. We have added this statement to the main paper.

Minor comments:

the discussion of model imperfection l44 or so could mention climate sensitivity uncertainty which also illustrates that even if technically reducable in practice this will be a slow processes.

Climate sensitivity is implicit in model response uncertainty, but we nonetheless added an explicit mention of this point, including a citation of Roe and Baker (2007).

p6 l24: mention that I S M are variances

We have added this clarification.

Discussion l17 p21: This might also cross reference to signal to noise maximization used in optimal detection eg Hasselmann, 1979 or Allen and Tett.

We have added these references.

References

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res. D Atmos.*, **118**, 1139–1150, https://doi.org/10.1002/jgrd.50174.

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, https://doi.org/10.1002/grl.50256.

Roe, G. H., and M. B. Baker, 2007: Why is climate sensitivity so unpredictable? *Science (80-.).*, **318**, 629–632, https://doi.org/10.1126/science.1144735.

Rotstayn, L. D., M. A. Collier, D. T. Shindell, and O. Boucher, 2015: Why does aerosol forcing control historical global-mean surface temperature change in CMIP5 models? *J. Clim.*, **28**, 6608–6625, https://doi.org/10.1175/JCLI-D-14-00712.1.

Sanderson, B. M., R. Knutti, and P. Caldwell, 2015: Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J. Clim.*, **28**, 5150–5170, https://doi.org/10.1175/JCLI-D-14-00361.1.

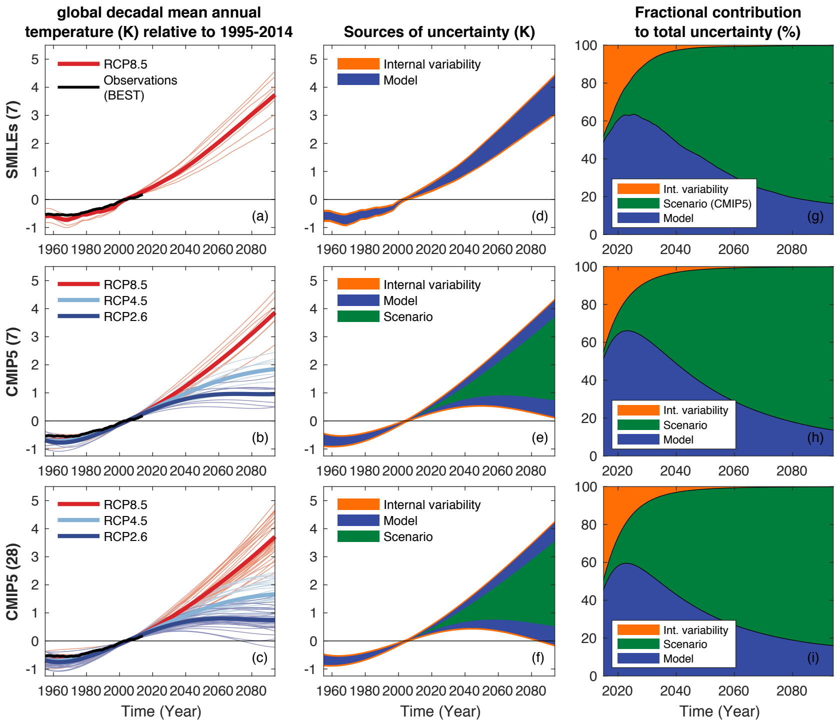

Fig. R1: Projections of global decadal mean annual temperature from (top row) the SMILEs, (middle row) the 7 SMILE models from CMIP5, processed with the HS09 approach, (bottom row) the 28 CMIP5 models. Top and bottom row are identical to the corresponding rows in Fig. 1 in the main text. The similarities of the results across the rows is an indication of the representativeness of the 7 SMILE models for CMIP5.

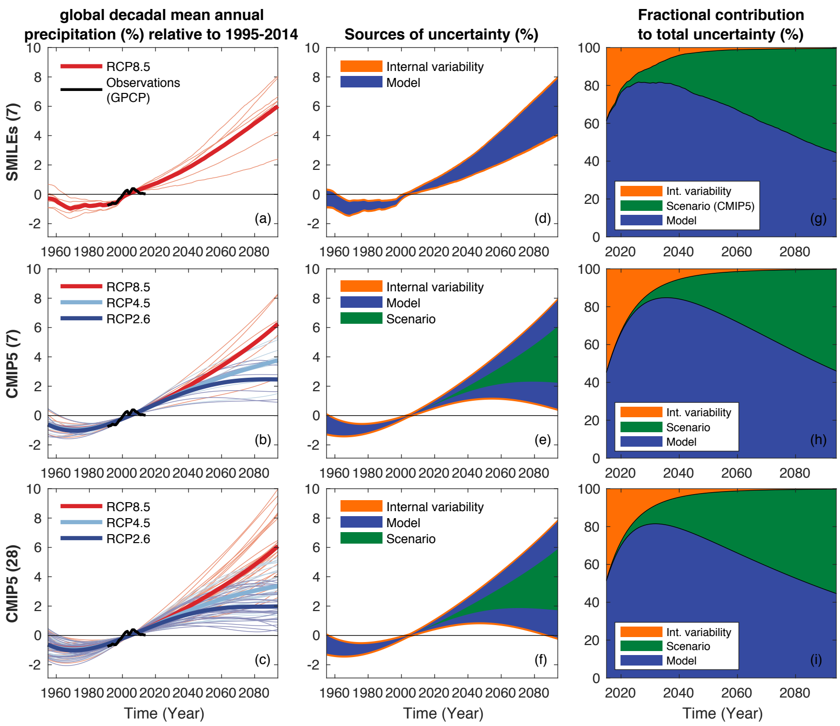

Fig. R2: Same as Fig. R1 but for global decadal mean annual precipitation.

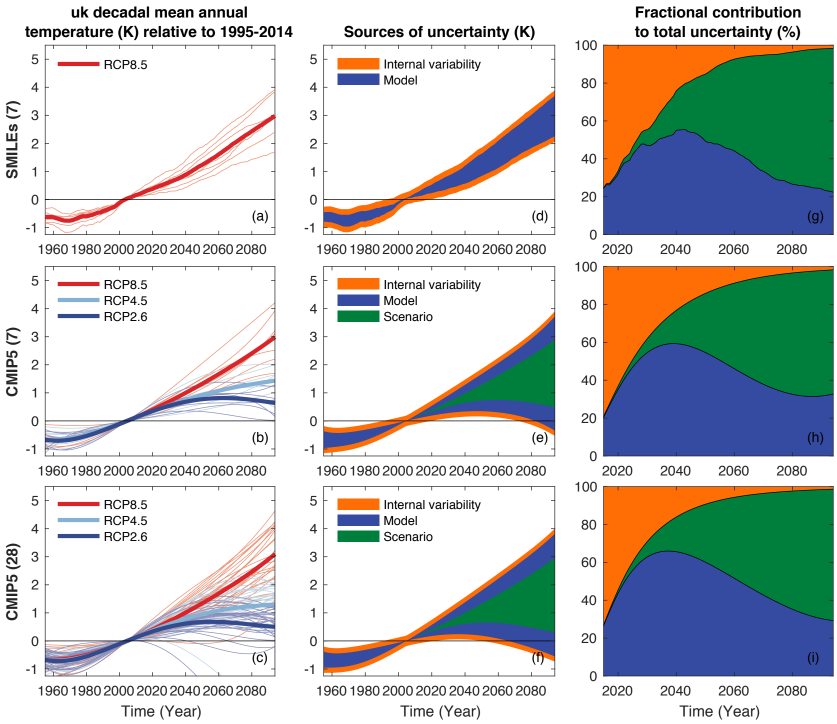

Fig. R3: Same as Fig. R1 but for British Isles decadal mean annual temperature.

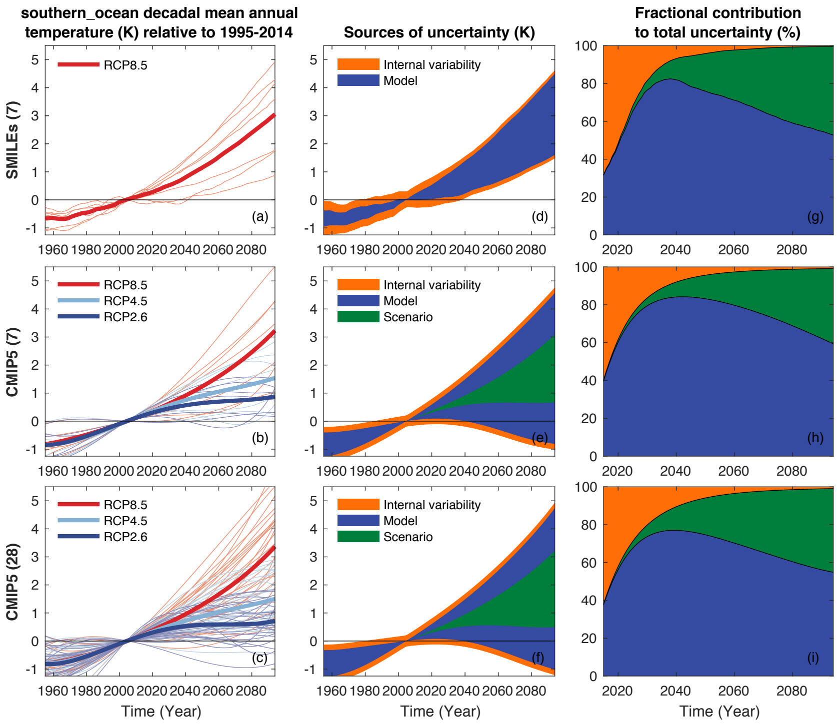

Fig. R4: Same as Fig. R1 but for Southern Ocean decadal mean annual temperature.

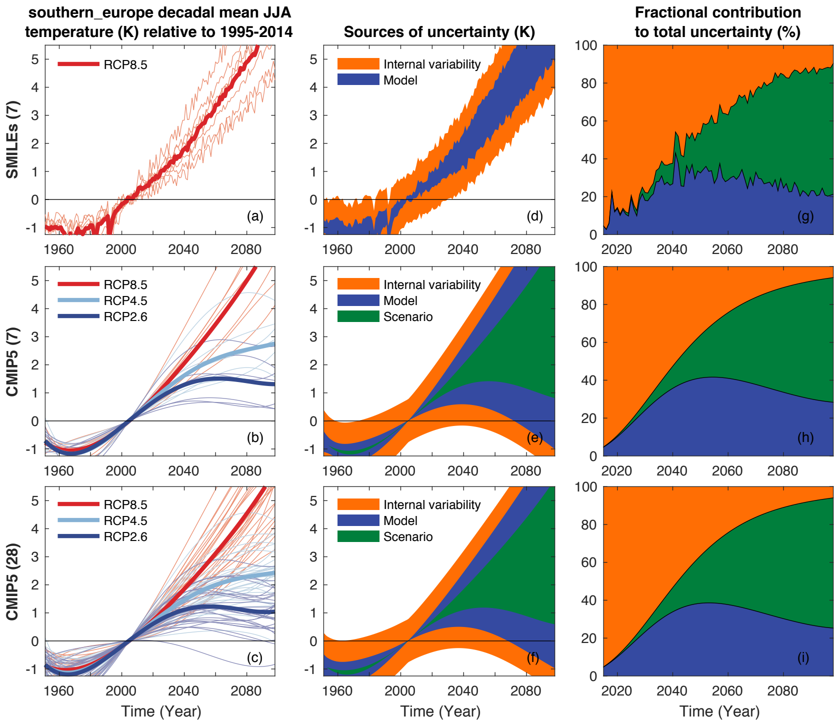

Fig. R5: Same as Fig. R1 but for Southern Europe JJA temperature.