

Supplement of Earth Syst. Dynam., 11, 491–508, 2020  
<https://doi.org/10.5194/esd-11-491-2020-supplement>  
© Author(s) 2020. This work is distributed under  
the Creative Commons Attribution 4.0 License.



*Supplement of*

## **Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6**

**Flavio Lehner et al.**

*Correspondence to:* Flavio Lehner (flehner@ucar.edu)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

**Table S1:** CMIP5 and CMIP6 models and ensemble members used.

<b>CMIP5 model</b>	<b>Ensemble member</b>	<b>CMIP6 model</b>	<b>Ensemble member</b>	<b>CMIP6 DOI</b>
bcc-csm1-1-m	rlilp1	BCC-CSM2-MR	rlilp1f1	10.22033/ESGF/CMIP6.2948
bcc-csm1-1	rlilp1	CAMS-CSM1-0	r2ilp1f1	10.22033/ESGF/CMIP6.9754
BNU-ESM	rlilp1	CESM2	rlilp1f1	10.22033/ESGF/CMIP6.7627
CanESM2	rlilp1	CESM2-WACCM	rlilp1f1	10.22033/ESGF/CMIP6.10071
CCSM4	rlilp1	CNRM-CM6-1	rlilp1f2	10.22033/ESGF/CMIP6.4066
CESM1-CAM5	rlilp1	CNRM-ESM2-1	rlilp1f2	10.22033/ESGF/CMIP6.4068
CNRM-CM5	rlilp1	CanESM5	r10ilp1f1	10.22033/ESGF/CMIP6.3610
CSIRO-Mk3-6-0	rlilp1	EC-Earth3	rlilp1f1	n/a
EC-EARTH	r8ilp1	EC-Earth3-Veg	rlilp1f1	10.22033/ESGF/CMIP6.4706
FGOALS-g2	rlilp1	FGOALS-f3-L	rlilp1f1	10.22033/ESGF/CMIP6.3355
FIO-ESM	rlilp1	FGOALS-g3	rlilp1f1	10.22033/ESGF/CMIP6.3356
GFDL-CM3	rlilp1	GFDL-ESM4	rlilp1f1	10.22033/ESGF/CMIP6.8597
GFDL-ESM2G	rlilp1	INM-CM4-8	rlilp1f1	10.22033/ESGF/CMIP6.5069
GFDL-ESM2M	rlilp1	INM-CM5-0	rlilp1f1	10.22033/ESGF/CMIP6.5070
GISS-E2-H	rlilp1	IPSL-CM6A-LR	rlilp1f1	10.22033/ESGF/CMIP6.5195
GISS-E2-R	rlilp1	MCM-UA-1-0	rlilp1f2	10.22033/ESGF/CMIP6.8888
HadGEM2-AO	rlilp1	MIROC-ES2L	rlilp1f2	10.22033/ESGF/CMIP6.5602
HadGEM2-ES	rlilp1	MIROC6	rlilp1f1	10.22033/ESGF/CMIP6.5603
IPSL-CM5A-LR	rlilp1	MPI-ESM1-2-HR	rlilp1f1	10.22033/ESGF/CMIP6.6594
IPSL-CM5A-MR	rlilp1	MRI-ESM2-0	rlilp1f1	10.22033/ESGF/CMIP6.6842
MIROC5	rlilp1	UKESM1-0-LL	rlilp1f2	10.22033/ESGF/CMIP6.6113
MIROC-ESM-CHEM	rlilp1			
MIROC-ESM	rlilp1			
MPI-ESM-LR	rlilp1			
MPI-ESM-MR	rlilp1			
MRI-CGCM3	rlilp1			
NorESM1-ME	rlilp1			
NorESM1-M	rlilp1			

## 15 **S1. Time period over which to estimate internal variability**

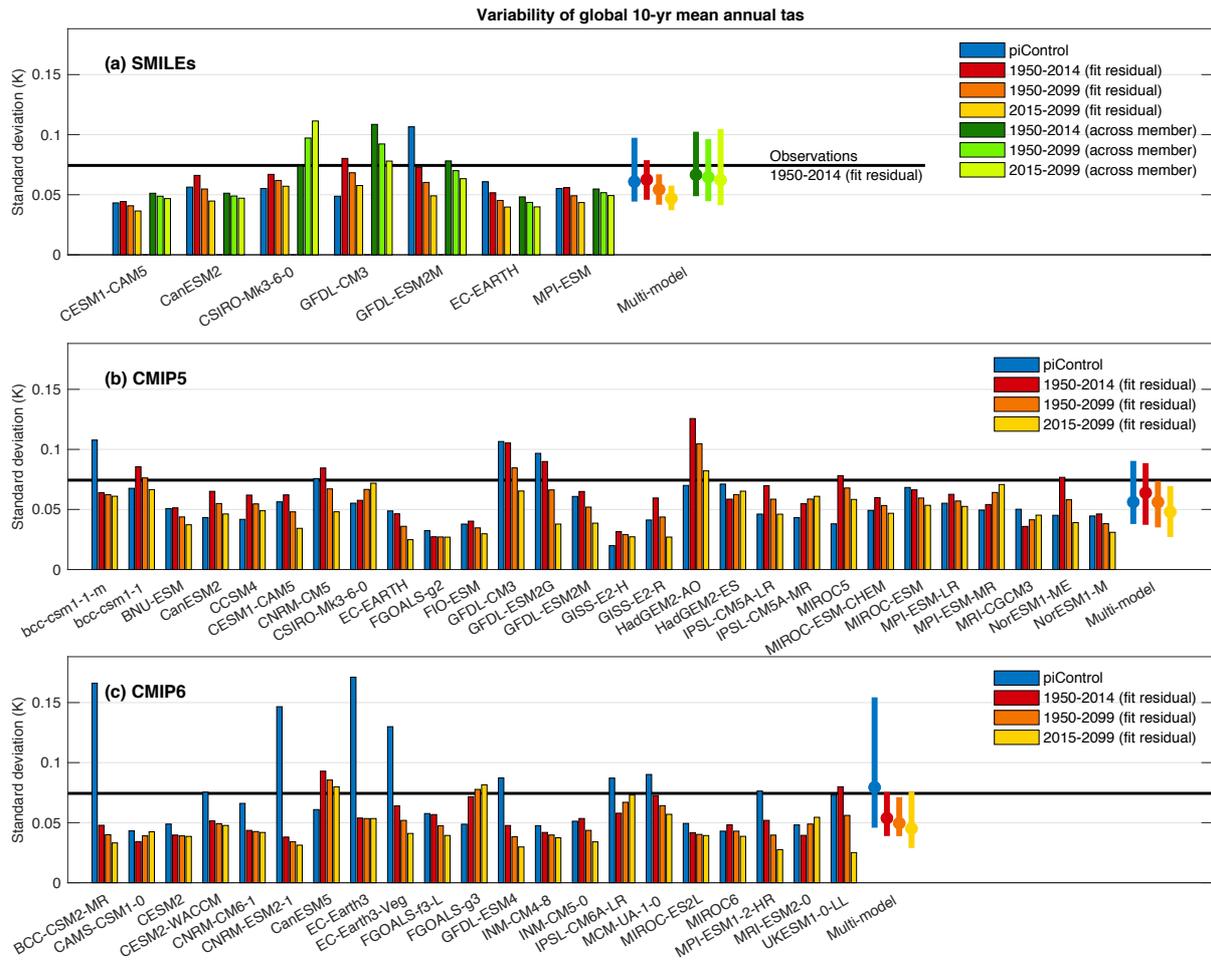
Explosive volcanic eruptions can significantly affect global and regional climate in subsequent years (Swingedouw et al. 2015; Lehner et al. 2016). If internal variability  $I$  of a quantity is calculated over time and including volcanic eruptions (for example over 1950-2014), it might be larger than when calculated without volcanic eruptions (for example over 2015-2099). Here, we quantify this potential effect by calculating  $I$  over  
20 different time periods. Note that here in the Supplementary Information,  $I$  is given in standard deviations for legibility, while in the main text  $I$  is given as variance for mathematical reasons. For a single simulation from a single model (like in CMIP),  $I$  is calculated as the standard deviation of the residual of the Hawkins and Sutton (2009; HS09) approach applied to historical and future simulations ( $I_{residual}$ ). For global decadal mean annual temperature in the SMILEs, the multi-model mean  $I_{residual}$  over 1950-2014 (0.063 K) is indeed larger than over  
25 1950-2099 (0.054 K), which is again larger than 2015-2099 (0.047 K; Fig. S1a). However, the differences are small and fall within the range of  $I_{residual}$  across models (Fig. S1a). Very similar results are found for CMIP5 (0.066 K, 0.057 K, 0.048 K) and CMIP6 (0.057 K, 0.052 K, 0.047 K), except the range of  $I_{residual}$  across models is even larger than in SMILEs (Fig. S1b-c).

30 In SMILEs, internal variability  $I$  can also be calculated as the across-member standard deviation ( $I_{across}$ ), such that  $I$  at any point in time might be expected to be independent from (or at least less affected by) volcanic eruptions, as all members experience the impact of the eruption simultaneously. Investigating  $I_{across}$  shows that this is not necessarily the case, with  $I_{across}$  for 1950-2014 being largest and  $I_{across}$  for 2015-2099 being smallest (Fig. S1a), however, the differences between time periods are even smaller than for  $I_{residual}$  (0.067 K, 0.065 K,  
35 0.062 K). The general similarity of  $I_{residual}$  and  $I_{across}$  also confirms again that the HS09 approach for separating forced response and internal variability works well for global temperature.

Finally, the variability from ‘piControl’ simulations  $I_{control}$  is shown. In this case, variability is calculated over the last 252 years of each model’s piControl simulation (a common length among models) after linearly detrending  
40 and applying a 10-year running mean.  $I_{control}$  is generally comparable to  $I_{residual}$  (and  $I_{across}$  in case of the SMILEs), except for a few models (e.g., GFDL-ESM2M, bcc-csm1-1m, BCC-CSM2-MR, CNRM-ESM2-1, EC-Earth3, EC-Earth3-Veg, GFDL-ESM4) which show large unforced decadal variability in piControl, the reasons for which remains to be investigated. The clustering of such high-variability piControl simulations in CMIP6 yields a multi-model mean  $I_{control}$  that is substantially higher in CMIP6 than in CMIP5 or SMILEs (Fig. S1c).

45

In summary, there exists a sensitivity to the choice of period over which variability is estimated, but it is of secondary importance compared to differences in variability magnitude between models. In the main text, we use 1950-2099 as the time period to estimate internal variability ( $I_{residual}$ ) for CMIP5 and CMIP6.



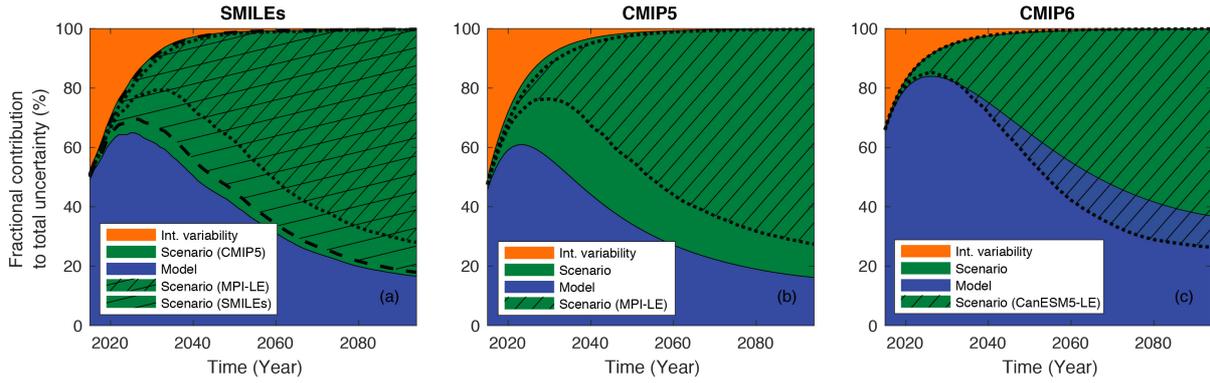
50

**Figure S1:** Standard deviation of global decadal mean annual temperature from (a) SMILEs, (b) CMIP5, and (c) CMIP6. In case of SMILEs, the average of all ensemble members is shown for each model. The multi-model mean and 10-90% range is given on the right end of the bar plots. Observations are from Rohde et al. (2013).

## 55 **S2. Role of choice of scenario uncertainty**

Estimating scenario uncertainty  $S$  is complicated by several factors: (1) The scenarios that climate model are run with represent only a subsample of available scenarios (Riahi et al. 2017). Although the representative scenarios chosen in CMIP5 and CMIP6 span a large range of possible future radiative forcing pathways, this subjective choice will always limit the CMIP archives to be “an ensemble of opportunity” rather than a true probabilistic  
60 assessment of future climate change. The scenarios are also not symmetrically distributed in radiative forcing space. (2) Not all modelling centers ran each of the chosen scenarios. Even rarer is the case where a modelling center ran a SMILE for each scenario (e.g., MPI-LE with CMIP5 scenarios and CanESM5 with CMIP6 scenarios). (3) Different methods to calculate scenario uncertainty exist. Here we explore points (2) and (3) in more detail.

65 Regarding (2), a compromise is necessary when one wants to estimate  $S$  from the available model simulations: either (i) use a consistent set of multiple models which ran at least one simulation per scenario, which means the forced response in any given model needs to be estimated via a statistical fit to one or few ensemble members available, or (ii) use a model that provides a SMILE for each scenario, which means the forced response for each  
70 scenario can be estimated more robustly, but the resulting  $S$  is model-specific. Here, we explore these two approaches at the example of global decadal mean temperature. In the main paper, we use  $S$  from CMIP5 ( $S_{CMIP5}$ ) for the uncertainty breakdown with SMILEs, using one simulation per CMIP5 model and scenario (green shading in Fig. S2a). We can also subselect the CMIP5 archive to just use the seven models that we have SMILEs for (see Table 1 in main paper) to calculate  $S$  ( $S_{SMILEs}$ ), but still just using one simulation per model and scenario (hatching in Fig. S2a). It can be seen that  $S_{CMIP5}$  and  $S_{SMILEs}$  are very similar, suggesting that the SMILEs are a good  
75 representation of CMIP5. Then, we use the MPI-LE (Maher et al. 2019), which has 100 ensemble members for each of the CMIP5 scenarios RCP2.6, RCP4.5 and RCP8.5, to estimate  $S$  ( $S_{MPI-LE}$ ; cross hatching in Fig. S2a).  $S_{MPI-LE}$  results in a smaller contribution from  $S$  to the total uncertainty. This is due to the relatively lower transient climate response of MPI compared to the multi-model mean of CMIP5 or SMILEs. Consequently, the trajectories of global temperature fan out slower across the different scenarios in MPI-LE than in more sensitive models,  
80 resulting in a smaller  $S$ . This is confirmed when just using data from CMIP5 (Fig. S2b), where  $S_{MPI-LE}$  is also smaller than  $S_{CMIP5}$ . The same exercise can be repeated for CMIP6, where the CanESM5-LE provides 50 ensemble members for each of the CMIP6 scenarios SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 (Swart et al. 2019). In this case, however, the  $S$  from CanESM5-LE ( $S_{CanESM5-LE}$ ) is almost always larger than the  $S$  from CMIP6 ( $S_{CMIP6}$ ), as CanESM5 constitutes a higher-sensitivity model among its CMIP6 cohort. While any of the approaches to  
85 estimate  $S$  for SMILEs are imperfect, we chose to use  $S_{CMIP5}$  in the main text due to it representing the expected true  $S_{SMILEs}$  well. It also facilitates a clean comparison of SMILEs with CMIP5 with regards to the other sources of uncertainty (internal variability and model uncertainty), as  $S$  is kept consistent between SMILEs and CMIP5.



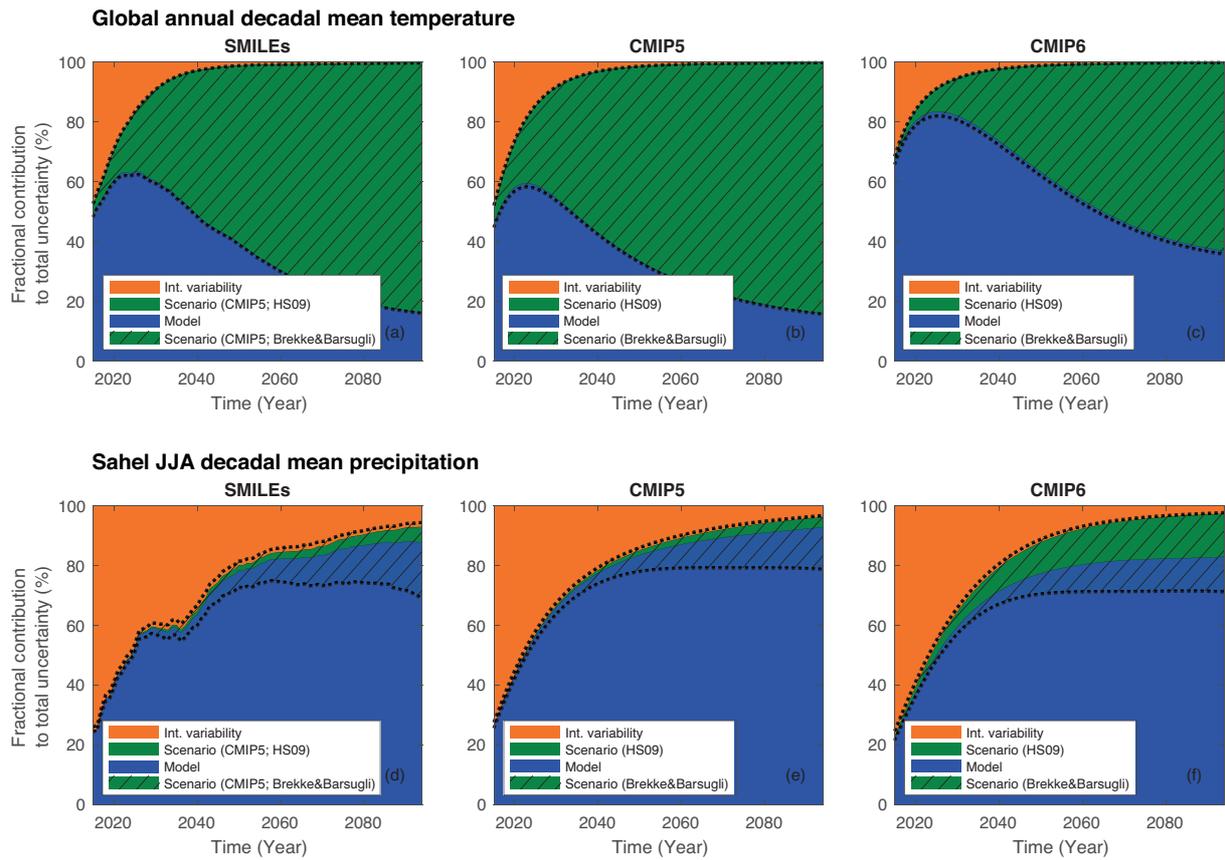
90 **Figure S2:** Fractional contribution of individual sources to total uncertainty for global annual decadal mean temperature in (a) SMILEs, (b) CMIP5, and (c) CMIP6. Scenario uncertainty for SMILEs in (a) is taken from (green shading) CMIP5, (cross hatching) MPI-LE, and (hatching) the models of the seven SMILEs. Scenario uncertainty in (b) is taken from (green shading) CMIP5 and (hatching) MPI-LE. Scenario uncertainty in (c) is taken from (green shading) CMIP6 and (hatching) CanESM5-LE.

95

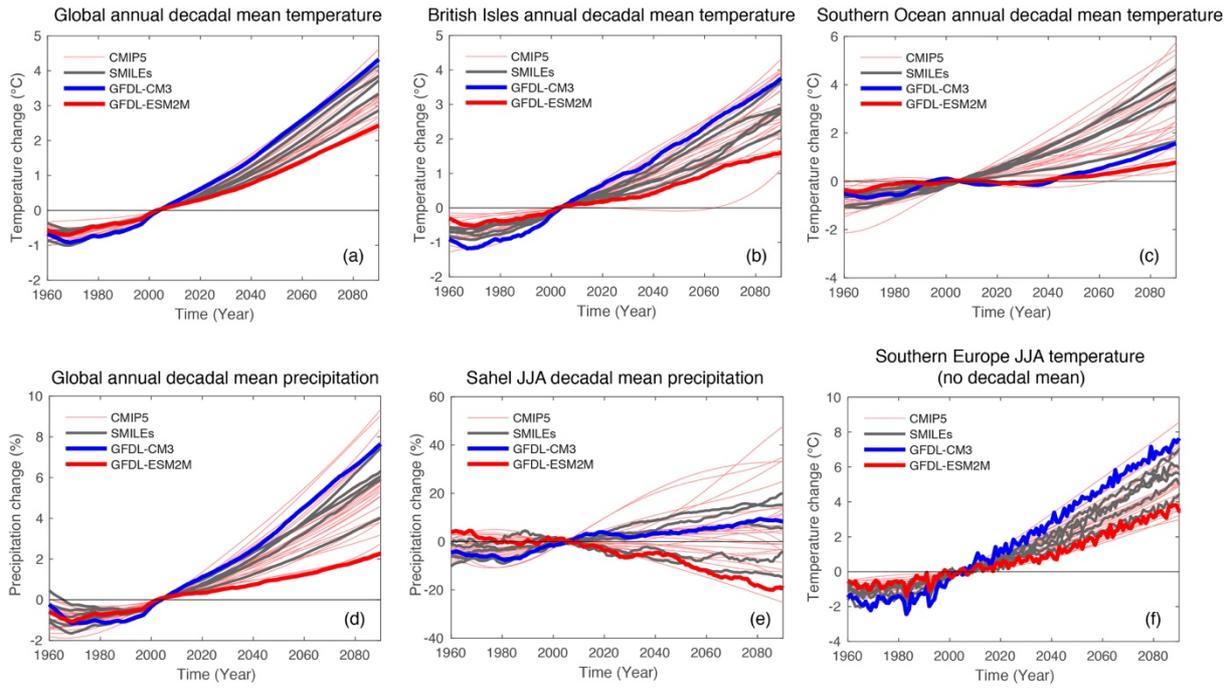
Regarding (3), it has been argued that it is difficult to estimate  $S$  correctly with the HS09 approach when the multi-model mean signal is weak and the sign of change uncertain (Brekke and Barsugli 2013; BB13). In HS09,  $S$  is calculated as the variance across multi-model means from different scenarios ( $S_{HS09}$ ). Thus, if models disagree on the sign of change, the multi-model means from the different scenarios can all be close to zero and close to each other –  $S$  would be small. However,  $S$  might actually be larger in any individual model. Thus, an alternative approach is to first calculate the variance across scenarios separately for each model (call this  $S_m$ ) and then average all  $S_m$  to obtain  $S$  ( $S_{BB13}$ ). If there indeed is a consistent scenario dependence across models that is masked in  $S_{HS09}$ ,  $S_{BB13}$  would be larger than  $S_{HS09}$ . We test this at examples of global temperature and Sahel precipitation (Fig. S3).

100 For global temperature, models show a strong multi-model mean change and agree well on the sign of change, so  $S_{HS09}$  and  $S_{BB13}$  are almost identical (Fig. 3a-c). For Sahel precipitation, on the other hand, which shows a weak multi-model mean change and lack of model agreement (see also Fig. S4e), this is not the case and  $S_{BB13}$  is larger than  $S_{HS09}$  (Fig. 3d-f). Still, we choose  $S_{HS09}$  as the approach in the main paper due to consistency with HS09 and because the main focus of the paper is not on  $S$  but on the separation of internal variability and model uncertainty.

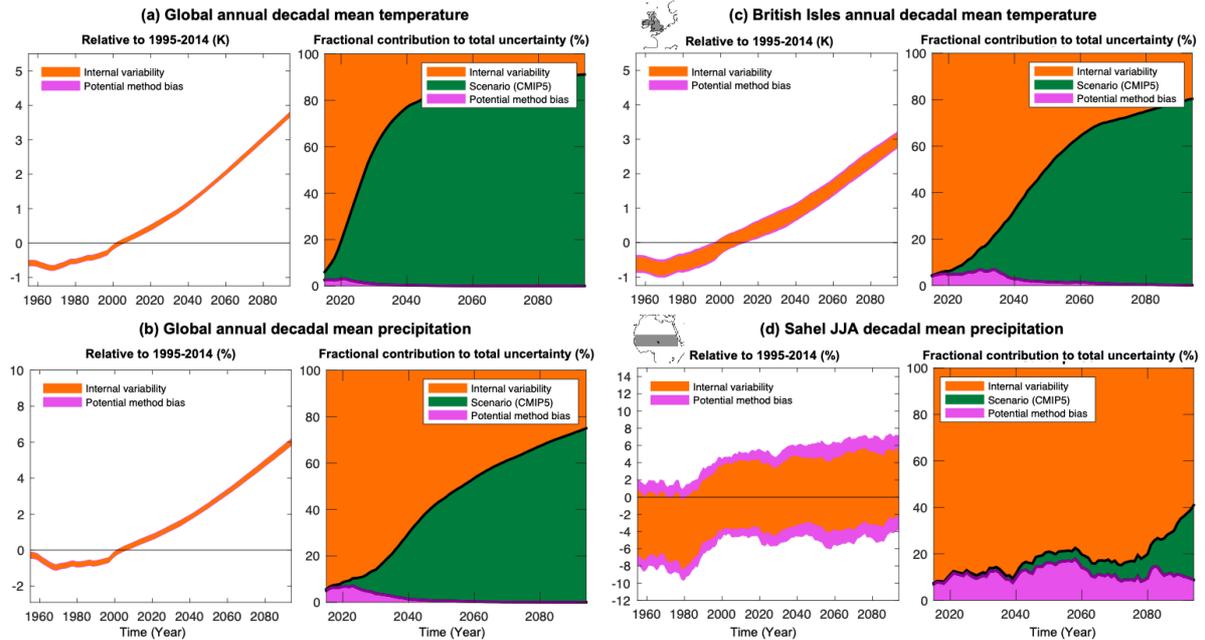
110



**Figure S3:** Fractional contribution of individual sources to total uncertainty for global annual decadal mean temperature in (a) SMILEs, (b) CMIP5, and (c) CMIP6. Scenario uncertainty is calculated as in (green shading) Hawkins and Sutton (2009), HS09, or as in (hatching) Brekke and Barsugli (2013). (d-f) same as in (a-c), but for Sahel JJA decadal mean precipitation.



**Figure S4:** Estimates of the forced response under RCP8.5 for (a-f) different regions and variables used in the main paper. For each CMIP5 model, the forced response is estimated as the 4<sup>th</sup> order polynomial, while for the SMILEs it is estimated as the ensemble mean. CMIP5 models are marked with thin red lines and SMILEs are marked with thick gray lines. The two GFDL models that are part of SMILEs are marked with thick blue and red lines. Overall, the SMILEs cover much of the range of CMIP5 models and the two GFDL models behave rather differently, even though they share components.



125 **Figure S5:** Decadal mean projections from SMILEs and fractional contribution to total uncertainty (using scenario  
 130 uncertainty from CMIP5) for (a) global mean annual temperature, (b) global mean annual precipitation, (c) British Isles  
 annual temperature, and (d) Sahel June-August precipitation. The pink color indicates the potential method bias and is  
 calculated the same way as model uncertainty in the HS09 approach, except instead of different models and the 4<sup>th</sup> order  
 polynomial, we use different ensemble mean estimates from the same SMILE. Specifically, we randomly select 16 members  
 from the largest SMILE (MPI) to mimic the ensemble size of the smallest SMILE (EC-EARTH) and calculate the ensemble  
 mean. We do this 100 times and calculate the variance across these ensemble means to be the potential method bias. Thus if  
 the SMILE ensemble mean method were perfect, the bias would be zero. This bias here is also non-zero but substantially  
 smaller than with the HS09 approach (see Fig. 5 in main text).

## References

- 135 Brekke, L. D., and J. J. Barsugli, 2013: Uncertainties in Projections of Future Changes in Extremes. *Extremes in a Changing Climate*, A. Aghakouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, Eds., Springer, 309–346.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.*, **90**, 1095–1107, <https://doi.org/10.1175/2009BAMS2607.1>.
- 140 Lehner, F., A. P. Schurer, G. C. Hegerl, C. Deser, and T. L. Frölicher, 2016: The importance of ENSO phase during volcanic eruptions for detection and attribution. *Geophysical Research Letters*.
- Maher, N., and Coauthors, 2019: The Max Planck Institute Grand Ensemble – enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst.*, <https://doi.org/10.1029/2019MS001639>.
- Riahi, K., and Coauthors, 2017: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Environ. Chang.*, **42**, 153–168, <https://doi.org/10.1016/j.gloenvcha.2016.05.009>.
- 145 Rohde, R., and Coauthors, 2013: Berkeley Earth temperature averaging process. *Geoinformatic Geostatistics An Overv.*, **1**, 1–13.
- Swart, N. C., and Coauthors, 2019: The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev. Discuss.*, **5**, 1–68, <https://doi.org/10.5194/gmd-2019-177>.
- 150 Swingedouw, D., P. Ortega, J. Mignot, E. Guilyardi, V. Masson-Delmotte, P. G. Butler, M. Khodri, and R. Séférian, 2015: Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions. *Nat. Commun.*, **6**, <https://doi.org/10.1038/ncomms7545>.