Supplementary Information for

Partitioning climate projection uncertainty with multiple Large Ensembles and CMIP5/6

Flavio Lehner[1,2], Clara Deser[2], Nicola Maher[3], Jochem Marotzke[3], Erich Fischer[1], Lukas Brunner[1], Reto Knutti[1], Ed Hawkins[4]

[1]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

[2]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, USA

[3]Max Planck Institute for Meteorology, Hamburg, Germany

[4]National Centre for Atmospheric Science, Dept. of Meteorology, University of Reading, Reading, UK

Supplementary Table 1: CMIP5 and CMIP6 models and ensemble members used.

CMIP5 model	**Ensemble member**	**CMIP6 model**	**Ensemble member**
bcc-csm1-1-m	r1i1p1	BCC-CSM2-MR	r1i1p1f1
bcc-csm1-1	r1i1p1	CAMS-CSM1-0	r2i1p1f1
BNU-ESM	r1i1p1	CESM2	r1i1p1f1
CanESM2	r1i1p1	CESM2-WACCM	r1i1p1f1
CCSM4	r1i1p1	CNRM-CM6-1	r1i1p1f2
CESM1-CAM5	r1i1p1	CNRM-ESM2-1	r1i1p1f2
CNRM-CM5	r1i1p1	CanESM5	r10i1p1f1
CSIRO-Mk3-6-0	r1i1p1	EC-Earth3	r11i1p1f1
EC-EARTH	r8i1p1	EC-Earth3-Veg	r1i1p1f1
FGOALS-g2	r1i1p1	FGOALS-f3-L	r1i1p1f1
FIO-ESM	r1i1p1	FGOALS-g3	r1i1p1f1
GFDL-CM3	r1i1p1	GFDL-ESM4	r1i1p1f1
GFDL-ESM2G	r1i1p1	INM-CM4-8	r1i1p1f1
GFDL-ESM2M	r1i1p1	INM-CM5-0	r1i1p1f1
GISS-E2-H	r1i1p1	IPSL-CM6A-LR	r1i1p1f1
GISS-E2-R	r1i1p1	MCM-UA-1-0	r1i1p1f2
HadGEM2-AO	r1i1p1	MIROC-ES2L	r1i1p1f2
HadGEM2-ES	r1i1p1	MIROC6	r1i1p1f1
IPSL-CM5A-LR	r1i1p1	MPI-ESM1-2-HR	r1i1p1f1
IPSL-CM5A-MR	r1i1p1	MRI-ESM2-0	r1i1p1f1
MIROC5	r1i1p1	UKESM1-0-LL	r1i1p1f2
MIROC-ESM-CHEM	r1i1p1		
MIROC-ESM	r1i1p1		
MPI-ESM-LR	r1i1p1		
MPI-ESM-MR	r1i1p1		
MRI-CGCM3	r1i1p1		
NorESM1-ME	r1i1p1		
NorESM1-M	r1i1p1		

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 different time periods. In the Supplementary Information here, I is given in standard deviations for legibility, while in the main text variance is used. For a simulation from a single model (like in CMIP), I is calculated as the standard deviation of the residual of the 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 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).

In SMILEs, internal variability I can be calculated as the across-member standard deviation (I_{across}), thus 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. Comparing $I_{residual}$ and 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 are even smaller than for $I_{residual}$ (0.067 K, 0.065 K, 0.062 K). The similarity of $I_{residual}$ and I_{across} also confirm 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 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 that remains to be explored. 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).

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.

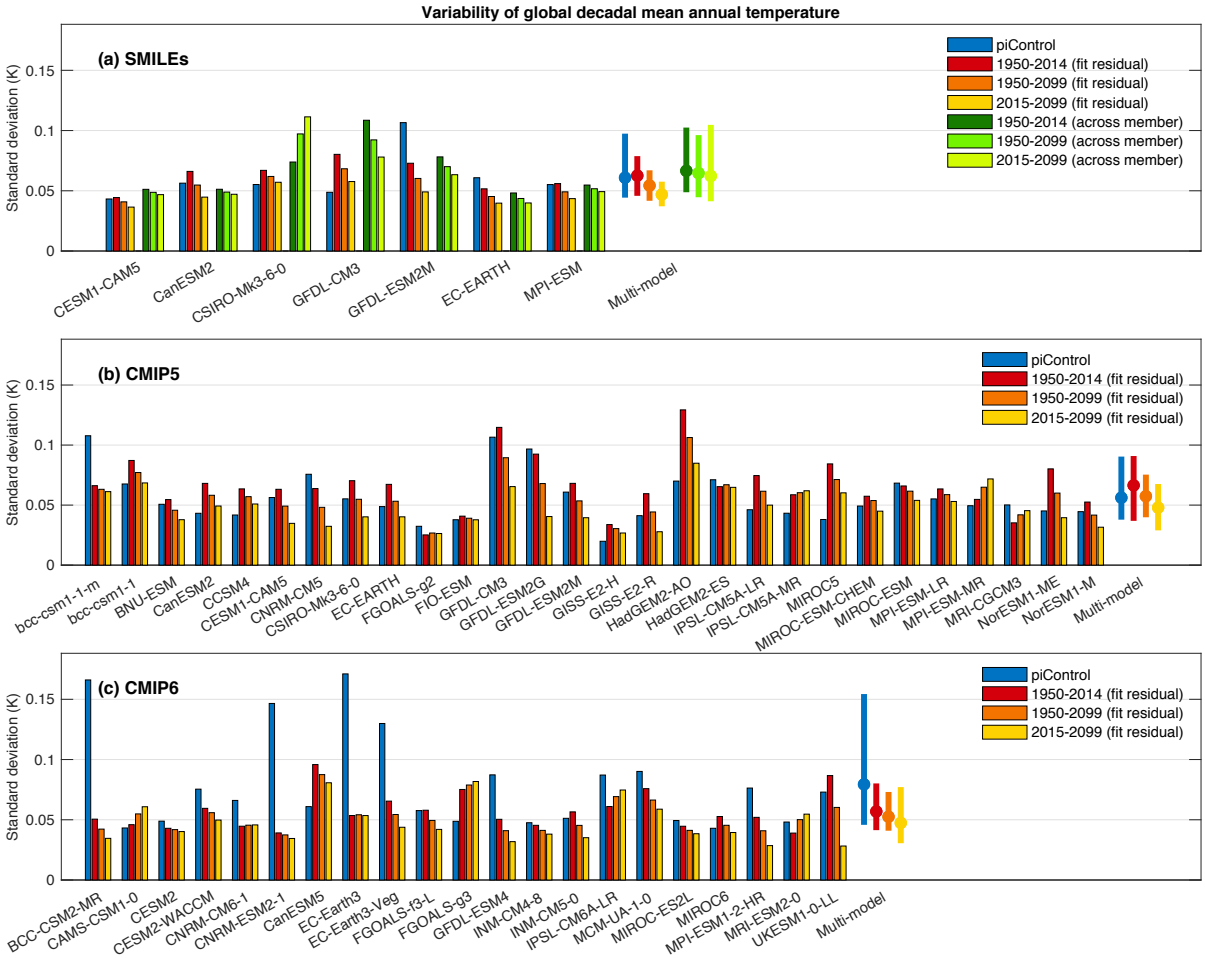

Figure S1: Standard deviation of global decadal mean annual temperature from (a) SMILEs, (b) CMIP5, and (c) CMIP6. In case of SMILEs, the average of all ensemble members is shown for each model. The multi-model mean and 10-90% range is given on the right end of the bar plots.

S2. Role of choice of scenario uncertainty

Estimating scenario uncertainty S is complicated by several factors. First, 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 assessment of future climate change. As discussed in the main text, the scenarios are also not symmetrically distributed in radiative forcing space. Second, and more tangible to explore, 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). Thus, a compromise is necessary when one wants to estimate S from the available model simulations: either use (i) 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 the one or few ensemble members available, or use (ii) a model with a SMILE for each scenario, which means the forced response for each scenario can be estimated 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 (dashed lines and flat hatching in Fig. S2a). It can be seen that S_{CMIP5} and S_{SMILEs} are very similar, suggesting that the SMILEs are a good 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 ($S_{MPI\text{-}LE}$; dotted lines and steep hatching in Fig. S2a). $S_{MPI\text{-}LE}$ 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 CMIP5 or SMILEs multi-model mean. Consequently, the trajectories of global temperature fan out slower across the different scenarios in MPI-LE than in more sensitive models, resulting in a smaller S. This is confirmed when just using data from CMIP5 (Fig. S2b), where $S_{MPI\text{-}LE}$ 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 ($S_{CanESM5\text{-}LE}$) is almost always larger than the S from CMIP6 (S_{CMIP6}), as CanESM5 constitutes a higher-sensitivity model among its CMIP6 cohort.

In summary, while any of the here presented approaches to estimate S for SMILEs is 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.

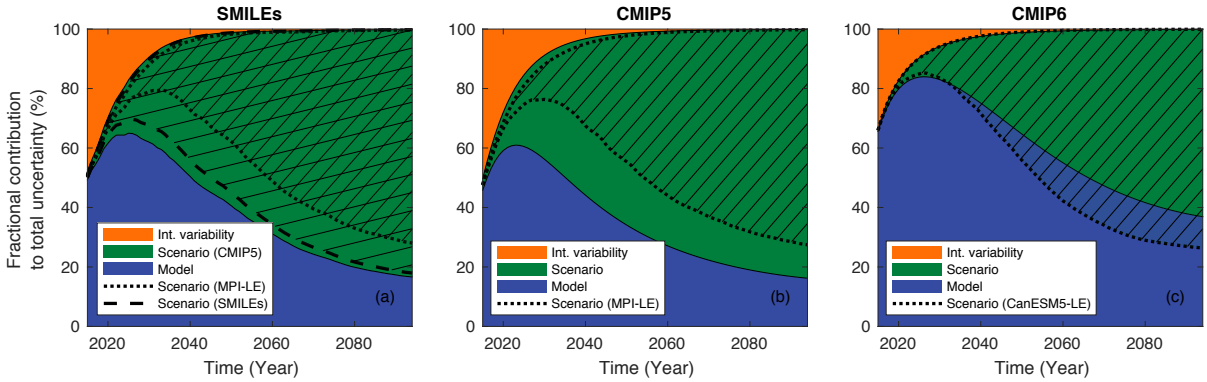

Figure S2: Fractional contribution of individual sources to total uncertainty in (a) SMILEs, (b) CMIP5, and (c) CMIP6. Scenario uncertainty for SMILEs in (a) is taken from (green shading) CMIP5, (dotted lines and steep hatching) MPI-LE, and (dashed lines and flat hatching) the models of the seven SMILEs. Scenario uncertainty in (b) is taken from (green shading) CMIP5 and (dotted line and steep hatching) MPI-LE. Scenario uncertainty in (c) is taken from (green shading) CMIP6 and (dotted line and steep hatching) CanESM5-LE.

References

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.

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.

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.