Supplementary Information for

Importance of the Pre-Industrial Baseline in Determining the Likelihood of Exceeding the Paris Limits

Andrew P. Schurer[1], Michael E. Mann[2], Ed Hawkins[3], Simon F. B. Tett[1], Gabriele C. Hegerl[1]

1. School of GeoSciences, University of Edinburgh, Crew Building, Alexander Crum Brown Road, Edinburgh, EH9 3FF, United Kingdom

2. Dept. of Meteorology and Atmospheric Science & Earth and Environmental Systems Institute, Pennsylvania State University, State College, PA

3. NCAS-Climate, Dept. of Meteorology, University of Reading, Reading, RG6 6BB, United Kingdom

This supplementary information file contains

- **Details of the model simulations used over the last millennium and the strength of radiative forcings.**
- **Details of the models used and the effect of applying weighting to calculate projected temperatures over the 21st century**
- **Figures showing the sensitivity to different methodological choices.**

Calculating pre-industrial temperatures using GHG only simulations

We use simulations forced with well-mixed greenhouse gases only, from three different models: HadCM3[1], CESM1-CAM5[2], CSIRO-Mk3L-1.2[3,4]. The respective transient climate sensitivities of these models are: 2.0K, 2.2K and 1.6K. The surface temperature response to this forcing is shown in figure S1 and was calculated as a blend of SATs and SSTs following Cowtan et al 2015[5]. This shows that by the end of the 19th century a non-negligible amount of warming has already taken place. The mean of the 3 model's ensemble means gives a warming of 0.13°C by the period 1851-1900 relative to a pre-industrial (1401-1800) baseline.

Model	# of Ensemble members
CESM1-CAM5	3
CSIRO-Mk3L-1-2	3*
HadCM3	4

Table S1 – Models used to determine response to GHGs. Asterisked value for CSIRO-Mk3L-1-2 is because GHG response in is calculated as the mean of three ensemble members with GHG and ozone forcing subtracted from 3 ensemble members with only ozone forcing

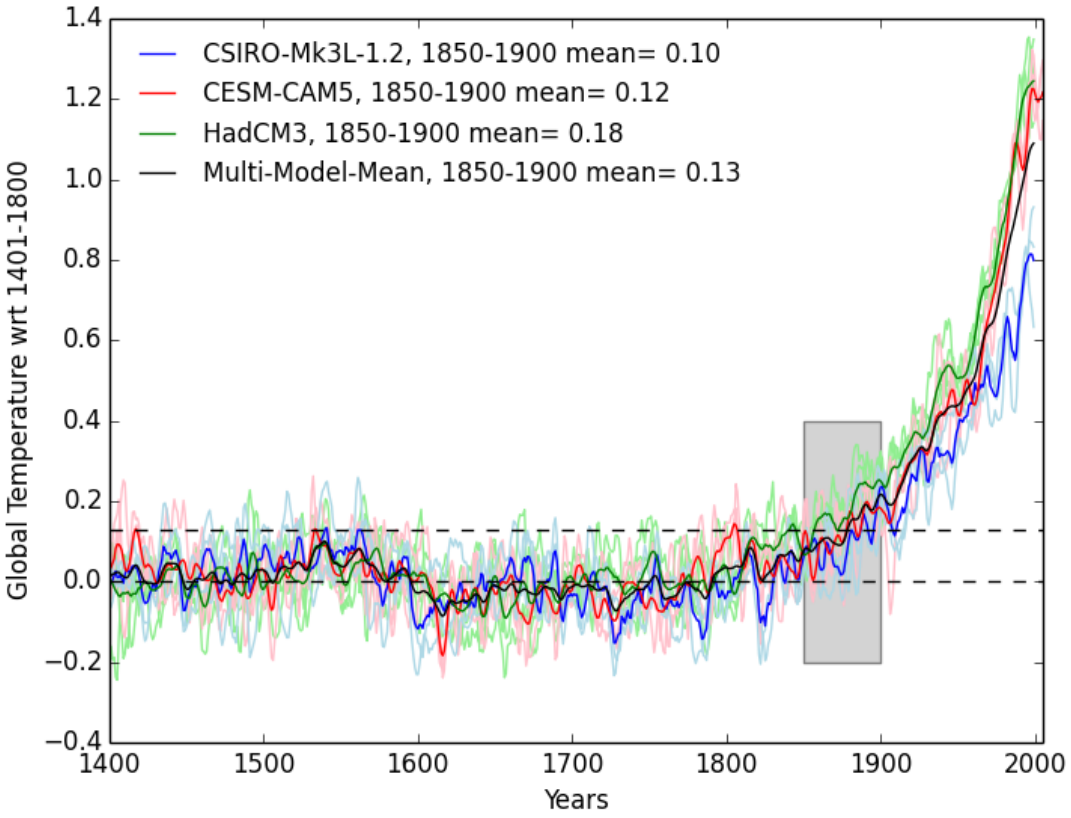

Fig S1- Temperature response to Greenhouse Gas forcing Global mean temperature for three models, smoothed by a 5-year running mean, details given in table S1. Bold coloured lines model means, light coloured lines individual ensemble members. Black line multi-model mean. 1851-1900 highlighted by grey box. Mean for this period shown in legend for the different models and multi-model mean for this period is highlighted by a horizontal dashed line.

Calculating pre-industrial temperatures using Volcanic eruptions only simulations

We use simulations forced with stratospheric volcanic aerosols only, from three different models: HadCM3, CESM1-CAM5, CSIRO-Mk3L-1.2. The surface temperature response to this forcing is shown in figure S2 and was calculated as a blend of SATs and SSTs following Cowtan et al 2015. The mean of the 3 model's ensemble means gives a cooling of 0.02°C by the period 1851-1900 relative to a pre-industrial (1401-1800) baseline.

Model	# of Ensemble members
CESM1-CAM5	5
CSIRO-Mk3L-1-2	3*
HadCM3	4

Table S2 – Models used to determine response to volcanic forcing. Asterisked value for CSIRO-Mk3L-1-2 is because volcanic response in is calculated as the mean of three ensemble members with volcanic, solar, GHG and ozone forcing subtracted from 3 ensemble members with only solar, GHG and ozone forcing

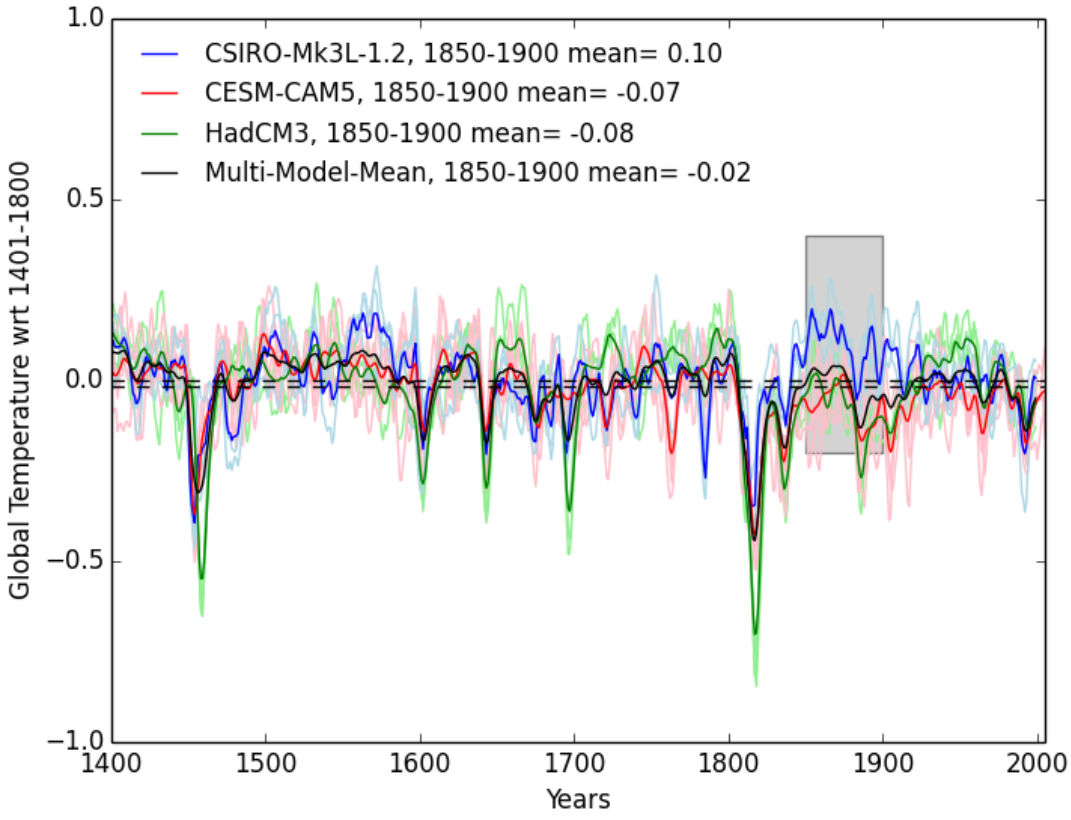

Fig S2- Temperature response to volcanic forcing Global mean temperature for three models, smoothed by a 5-year running mean, details given in table S2. Bold coloured lines model means, light coloured lines individual ensemble members. Black line multi-model mean. 1851-1900 highlighted by grey box. Mean for this period shown in legend for the different models and multi-model mean for this period is highlighted by a horizontal dashed line.

Calculating pre-industrial temperatures using Solar forcings only simulations

We use simulations forced with solar forcings only, from three different models: HadCM3, CESM1-CAM5, CSIRO-Mk3L-1.2. The surface temperature response to this forcing is shown in figure S3 and was calculated as a blend of SATs and SSTs following Cowtan et al 2015. The mean of the 3 model's ensemble means gives a cooling of 0.01°C by the period 1851-1900 relative to a pre-industrial (1401-1800) baseline.

Model	# of Ensemble members
CESM1-CAM5	4
CSIRO-Mk3L-1-2	3*
HadCM3	4

Table S3 – Models used to determine response to SOLAR. Asterisked value for CSIRO-Mk3L-1-2 is because SOLAR response in is calculated as the mean of three ensemble members with solar, GHG and ozone forcing subtracted from 3 ensemble members with only GHG and ozone forcing

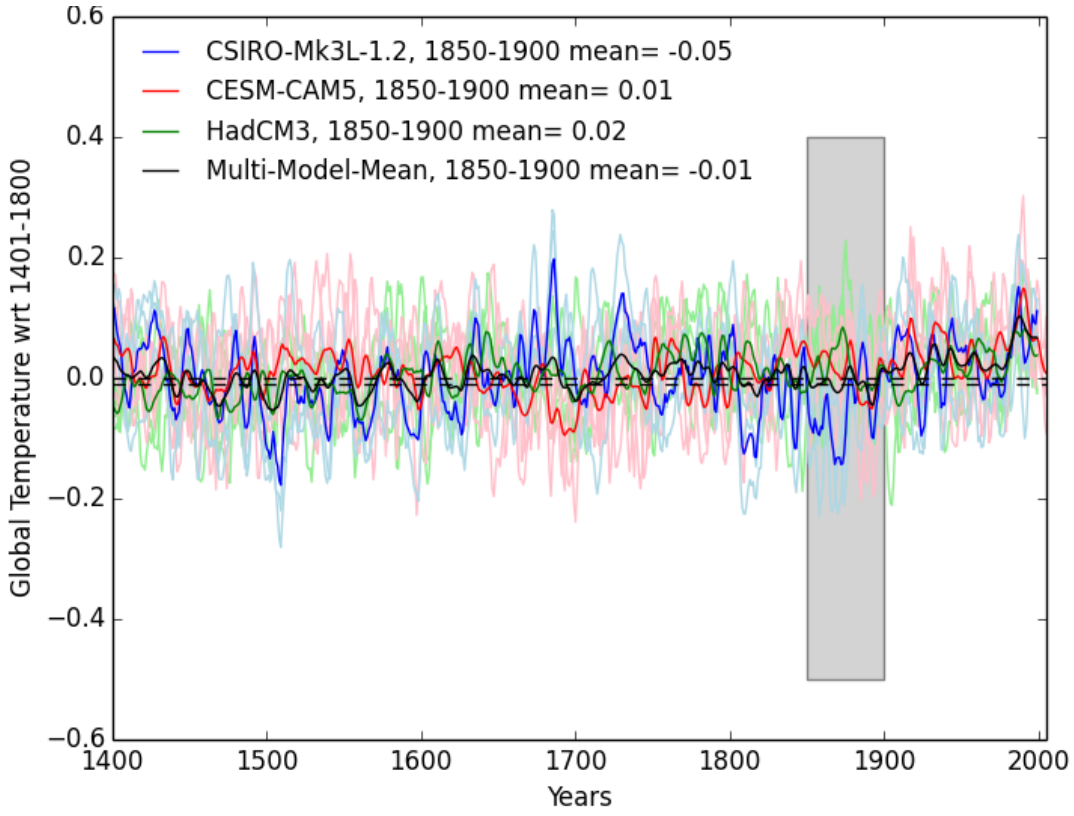

Fig S3- Temperature response to solar forcing Global mean temperature for three models, smoothed by a 5-year running mean, details given in table S3. Bold coloured lines model means, light coloured lines individual ensemble members. Black line multi-model mean. 1851-1900 highlighted by grey box. Mean for this period shown in legend for the different models and multi-model mean for this period is highlighted by a horizontal dashed line

Calculating pre-industrial temperatures using ALL forced simulations

We use simulations forced with all known forcings from seven different models (see table S4 for details). The surface temperature response to this forcing is shown in figure S4 and was calculated as a blend of SATs and SSTs following Cowtan et al 2015[5]. The mean of the 7 model's ensemble means gives a warming of 0.09°C by the period 1850-1900 relative to a pre-industrial (1401-1800) baseline.

Model	# of Ensemble members
bcc-csm1-1	1
CCSM4	1
CESM1-CAM5	10
CSIRO-Mk3L-1-2	3
GISS-E2-R	3
HadCM3	4
MPI-ESM-P	1

Table S4 –Models simulations covering the last 600 years which are used to determine response to All forcings.

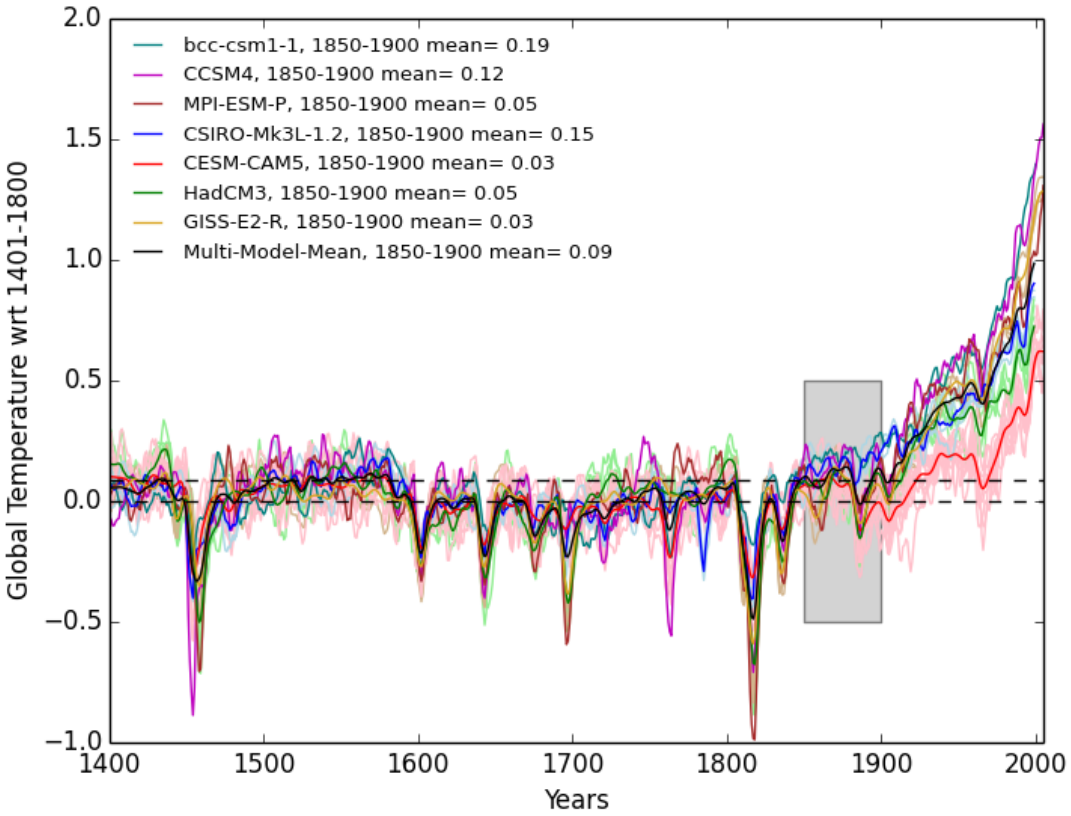

Fig S4- Temperature response to ALL forcings Global mean temperature for seven models, smoothed by a 5-year running mean, details given in table S4. Bold coloured lines model means, light coloured lines individual ensemble members. Black line multi-model mean. 1851-1900 highlighted by grey box. Mean for this period shown in legend for the different models and multi-model mean for this period is highlighted by a horizontal dashed line.

Calculating probability distributions of post 1850-1900 warming (PRE)

As described in the method section probability distributions are calculated for mean temperature difference from 1850-1900 for all 100-year periods from 1401-1800 for all available models. We calculate PDFs from these model results using kernel density estimation (see fig S5). The all forced PDF and GHG pdf are plotted in fig 3 of the main paper. The PDFs were constructed from all possible model simulations listed in tables S2 to S4. Models providing multiple ensemble members were weighted down so that each model contributed equally to the final PDF.

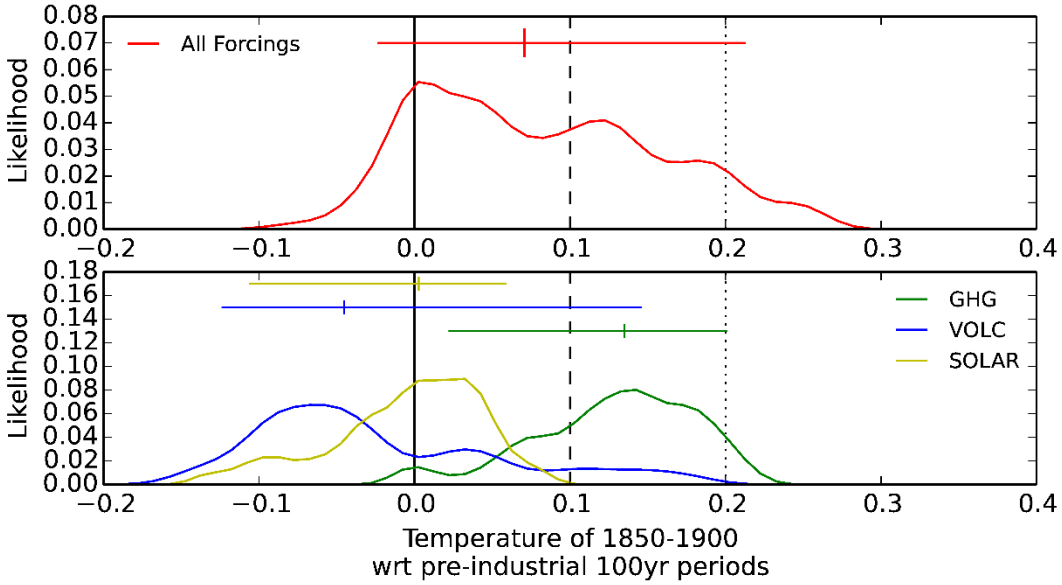

Figure S5. Probability density functions for mean temperature difference before 1850-1900. Mean temperature difference from 1850-1900 for all 100-year periods from 1401-1800 for all available models in different model experiments. Horizontal lines show 5-95% range, cross median value.

Calculating pre-industrial temperatures for periods before 1400

We use all simulations, which cover the period 850-2000. This includes all 7 of the all-forced models listed in Table S4, although the HadCM3 model only has a single ensemble member covering this period and the CESM1-CAM5 and CSIRO-Mk3L-1-2 (tables S1,2,3) single forced simulations (the HadCM3 simulations begin in 1400). For the GISS model, we remove initial drift from the simulations by fitting a second-order polynomial to the control simulation as described in Schurer et al[6]

The surface temperature response to these forcings is shown in figure S6 and was calculated as a blend of SATs and SSTs following Cowtan et al 2015[5]. Figure S6 shows that for some early periods in the last millennium temperatures were probably warmer than in the late 19th century due to a combination of solar and volcanic forcing.

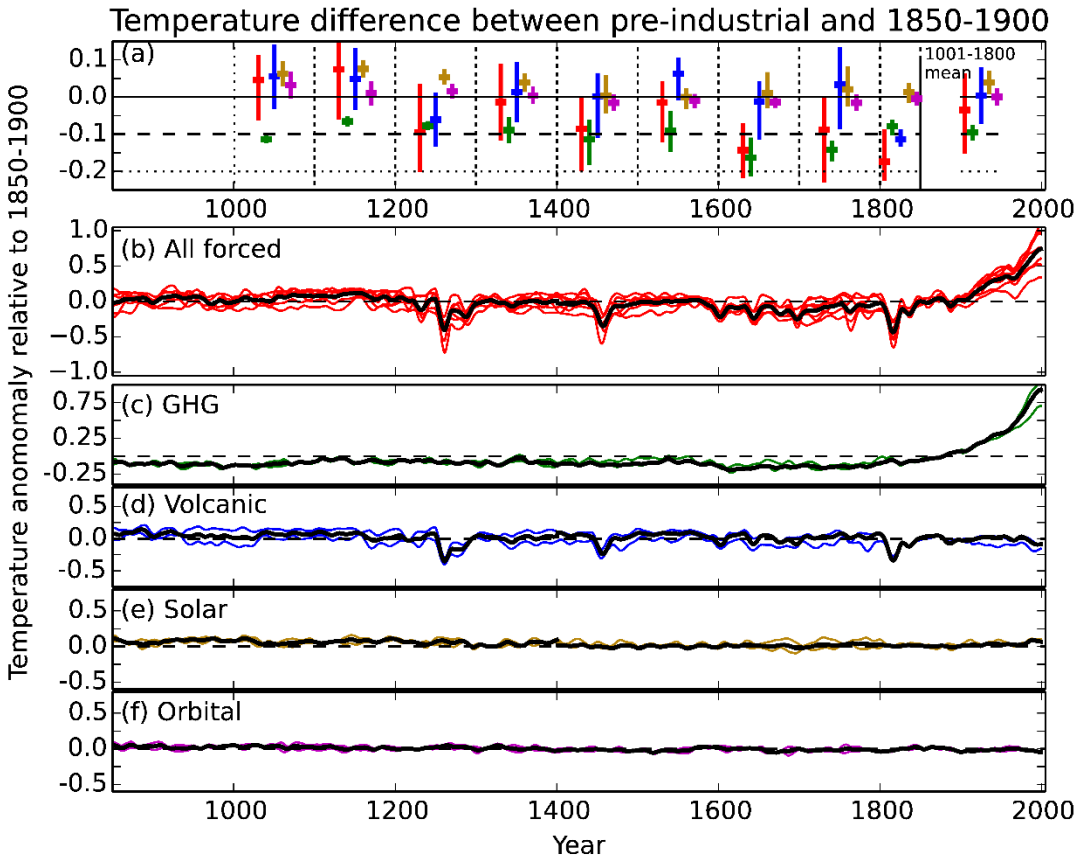

Fig S6. Model simulated difference in global mean temperature between different pre-industrial periods and 1850-1900. a) Range of ensemble means for different models, and for different forcing combinations (model mean: cross, model range: bar; red: All forcings combined; green: greenhouse gas forcing only, blue: volcanic forcing only, brown: solar forcing only, purple: orbital forcing only). Differences refer to the mean of the period enclosed by the dotted lines; except on far right where they are means for the full period 1001-1800 (relative to 1850 to 1900). b)-f) Model means for different forcing combinations, colours ensemble means for individual models, black line – mean over all models. This figure is the same as Fig 2 main paper but covers a longer time period and includes less models.

Levels of natural forcings over different time periods

In order to determine if a time period has a "typical" level of natural forcings it is useful to consider long-term means of volcanic and solar activity. Of particular interest is whether 1850-1900 can be considered typical or has unusual activity.

Volcanic Activity

One of the most important forcings[1], volcanic activity, varies considerably throughout the last millennium. The Crowley reconstruction of aerosol optical depth[7] (see fig S7) suggests that the 50-year period 1850-1900 has fairly typical activity with respect to other periods over the last millennium.

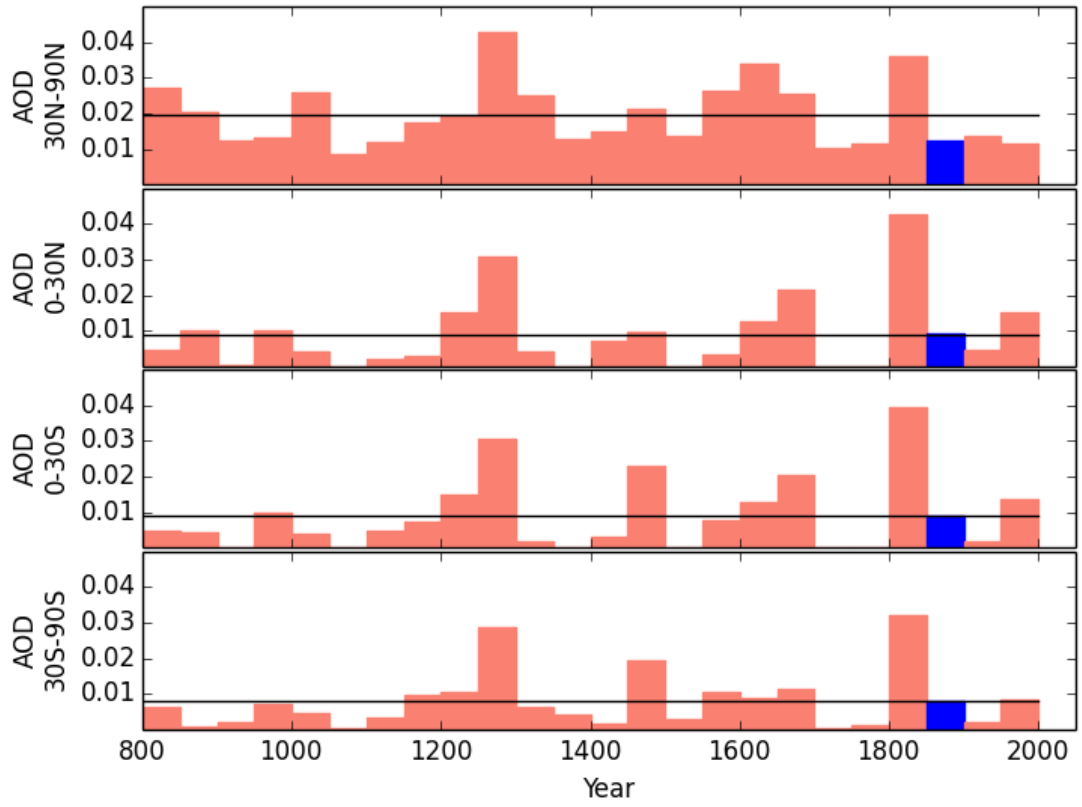

Fig S7 – 50-year means of volcanic activity. Aerosol optical depth (AOD) estimates are taken from Crowley and Unterman 2012[7] and are plotted in 4 different latitudinal bands. The period 1850-1900 is highlighted in blue, the mean activity across all time periods is shown by the black horizontal line.

Solar Activity

Total stellar irradiance estimates indicate higher solar activity in the first half of the last millennium, it then decreases before a rise to a maximum at the end of the 20th century (see figure S8). Within the context of the last millennium as a whole the value in 1850-1900 appears typical and is close to the mean.

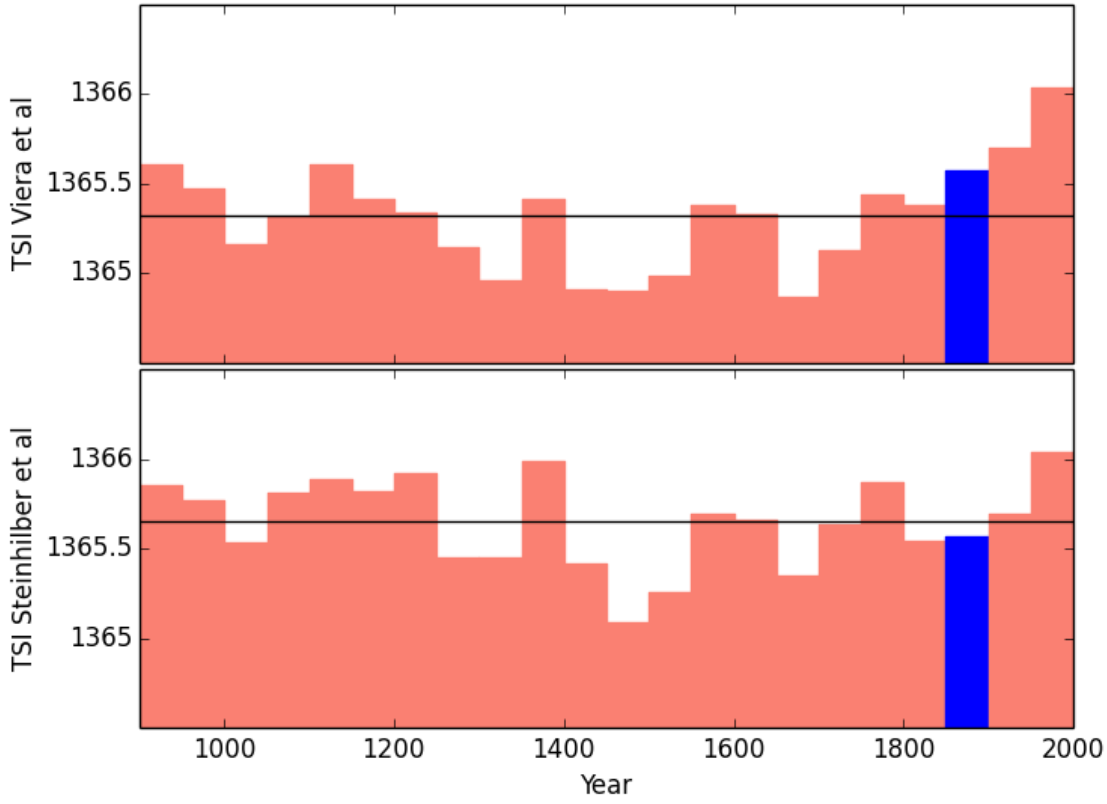

Fig. S8 – 50-year mean solar activity- Total stellar irradiance (TSI) estimates are taken from Viera et al[8] (top) and Steinhilber et al[9] (bottom), where both timeseries have been spliced onto the Wang et al[10] TSI reconstruction after 1850, following the recommendation of Schmidt et al[11]. The period 1850-1900 is highlighted in blue, the mean activity across all time periods is shown by the black horizontal line.

Effect of pre-industrial Baseline on 1.5° and 2°C stabilisation likelihoods

Based on climate simulations we have determined that a true pre-industrial baseline is likely to be cooler than the 1850-1900 mean by potentially up to 0.2°C (figure S5). We therefore investigate the effect a baseline that is 0.1° and 0.2°C colder than the 1850-1900 mean has on the projected temperature.

The effect of this additional warming since pre-industrial is shown in figure S9.

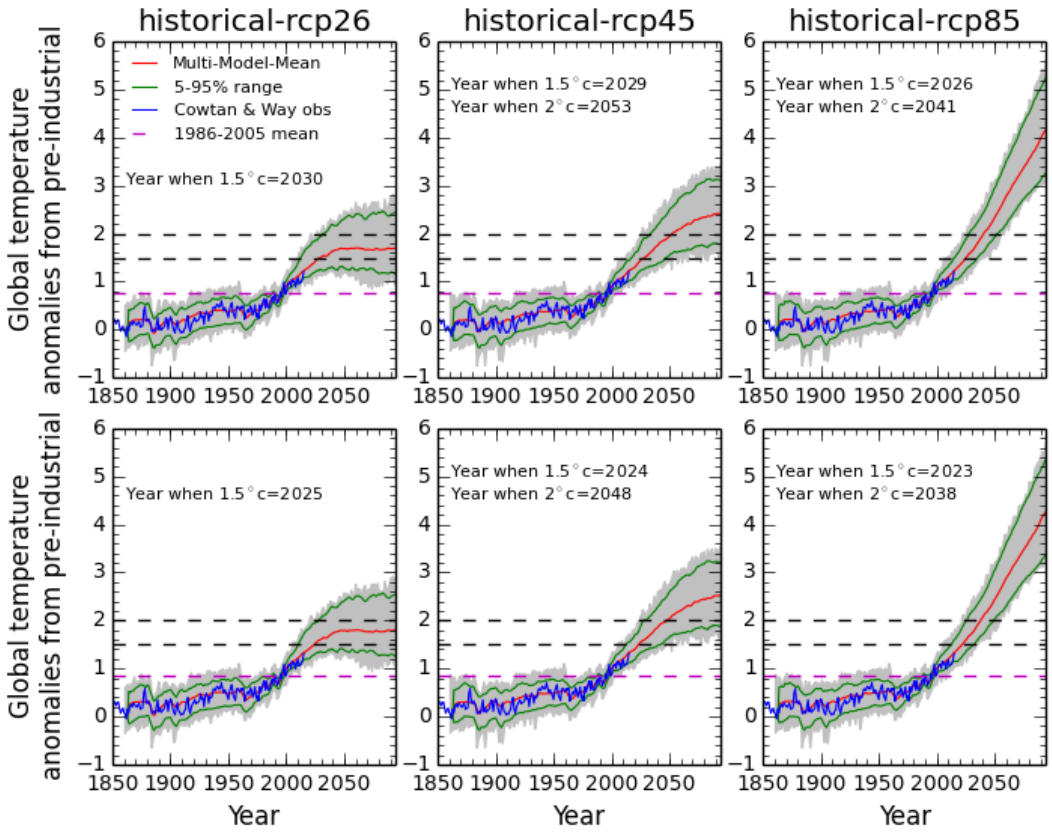

Fig S9– Historical and future projections for global mean temperature accounting for pre-instrumental warming. Identical to figure 1 except using a different baseline for temperature anomalies. In the top panels 0.1°of additional warming is assumed to have occurred by 1850-1900 and in the bottom panels an additional 0.2° of warming is assumed.

The effect of weighting on historic and Future projections

We use projections following three RCP scenarios (RCP 2.6, RCP 4.5 and RCP 8.5). Modelled surface temperature values are calculated from a blend of SATs and SSTs following Cowtan et al 2015[5] for total global coverage. Details of which models have been used are described in table S5.

Number	Model	# of ens. members	Colour in fig S10
1	ACCESS1-0	1	Green
2	ACCESS1-3	1	Green
3	bcc-csm1-1-m	1	Blue
4	bcc-csm1-1	1	Blue
5	BNU-ESM	1	Green
6	CanESM2	5	Blue
7	CCSM4	6	Green
8	CESM1-BGC	1	Blue
9	CESM1-CAM5	3	Blue
10	CMCC-CM	1	Green
11	CMCC-CMS	1	Green
12	CNRM-CM5	1	Blue
13	CSIRO-Mk3-6-0	10	Green
14	CSIRO-Mk3L-1-2	3	Green
15	EC-EARTH	5	Blue
16	FIO-ESM	3	Green
17	GFDL-CM3	1	Blue
18	GFDL-ESM2G	1	Blue
19	GFDL-ESM2M	1	Blue
20	GISS-E2-H-CC	1	Green
21	GISS-E2-H	16	Green
22	GISS-E2-R-CC	1	Green
23	GISS-E2-R	17	Green
24	HadGEM2-AO	1	Blue
25	HadGEM2-CC	1	Blue
26	HadGEM2-ES	4	Blue
27	inmcm4	1	Green
28	IPSL-CM5A-LR	4	Blue
29	IPSL-CM5A-MR	1	Blue
30	IPSL-CM5B-LR	1	Blue
31	MIROC5	3	Green
32	MIROC-ESM-CHEM	1	Green
33	MIROC-ESM	1	Green
34	MPI-ESM-LR	3	Blue
35	MPI-ESM-MR	3	Blue
36	MRI-CGCM3	1	Green
37	NorESM1-ME	1	Blue
38	NorESM1-M	1	Blue

Table S5 – The model simulations included in our analysis. The number in the first column corresponds to the x-axis fig S10.

We have calculated the effect of weighting the model based on how well they match the observational record. To improve the model-data intercomparison we carry out the weighting analysis using HadCRUT4[12] and model data masked to have identical observational coverage, where the model data is a blend of SATs and SSTs following the "HadCRUT method" described in *Cowtan et al 2015*[5]. We use a metric based on the global mean decadal residual, with 14 values between 1865-2005, which is calculated from the difference between observations, Y and each model X. To calculate a likelihood, we assume that the residuals follow a multivariate normal distribution, making the assumption that "errors" are Gaussian:

$$f = \frac{1}{\sqrt{2\pi|\sum|}} \exp\left(\frac{-1}{2}(\mathbf{Y}-\mathbf{X})^{T}\sum^{-1}(\mathbf{Y}-\mathbf{X})\right) \quad (3)$$

Where the covariance matrix, $\sum$ includes the observational uncertainty and internal variability:

$$\sum = \sum_{Obs} + 2\sum_{IV} \quad (4)$$

The observational covariance, $\sum_{Obs}$ is calculated from the 100 available possible realisations of observed temperature (it should be noted that this will likely be an underestimate as these realisations do not sample all the uncertainty types in HadCRUT4). The internal variability covariance, $\sum_{IV}$, is calculated from piControl simulations, the factor of 2 is included to account for unforced variability in both the observations and models. Equation 3 is used to calculate a likelihood for each model simulation and an average likelihood is calculated for each model. Some models have contributed far more simulations than others. So as not to bias the analysis towards particular models, for each model, we divide its likelihood by the number of ensemble members available for that model. To calculate a weight for a particular model we divide its likelihood by the sum of likelihoods across all models. The calculated weights are shown in figure S10.

To determine the importance of the weighting to our conclusions we repeat the same analysis as given in the main article but weighting each projection by its likelihood. The main results as shown in the main paper are based on an unweighted mean across many different models, figures S11, S12 and S13 show results where every model weighted according to its agreement with observed temperature. The consequence of weighting the models can be seen when comparing these figures to those in the main paper.

We find that the main effect of the weighting is to reduce the uncertainty bounds for the future projections, in particular the lower bound of the projected temperature is raised by discarding some models that warm comparatively less. This is most noticeable in scenarios RCP4.5 and RCP 8.5. We find that the choice of whether to use weighting or not does not change the overall conclusions of the paper. For example, the likelihood of exceeding 1.5°C in the RCP 2.6 scenario is little changed if weighting is not applied. Although the reduction of the uncertainty range due to weighting does change several of the exceedance likelihoods. For example, following RCP 4.5, the chance of exceeding 2.0°C increases from 80% to 91% if weighting is used (in the case where PRE is taken to be 0°C), due to the raising of the lower bounds of the temperature projections.

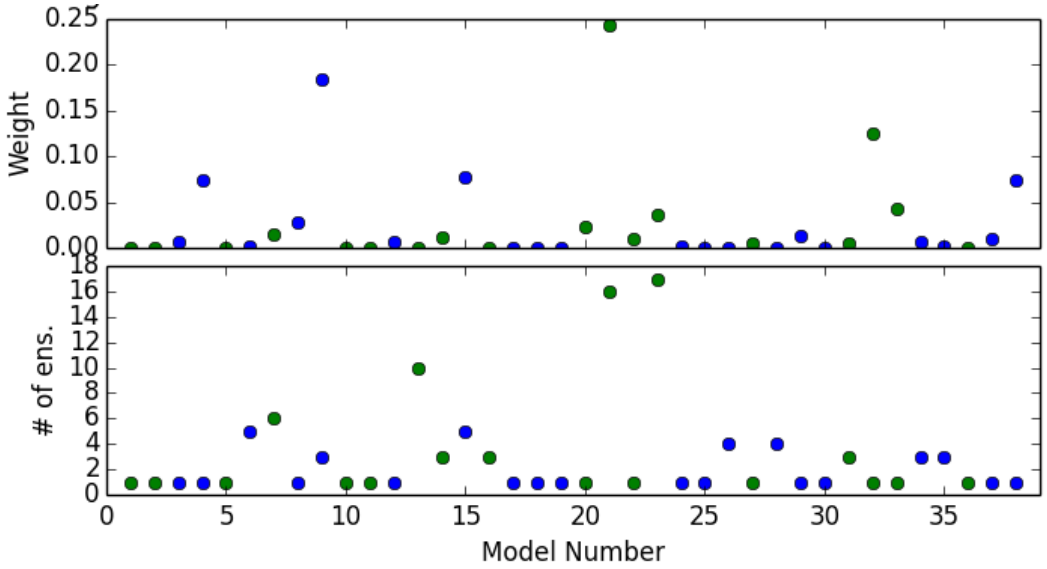

Fig S10. Model Weightings Top – Weighting calculated for model. Bottom - Number of ensemble members. Names of the models are given in table S5, the alternate colours are included for comprehension purposes only, and group models with similar names.

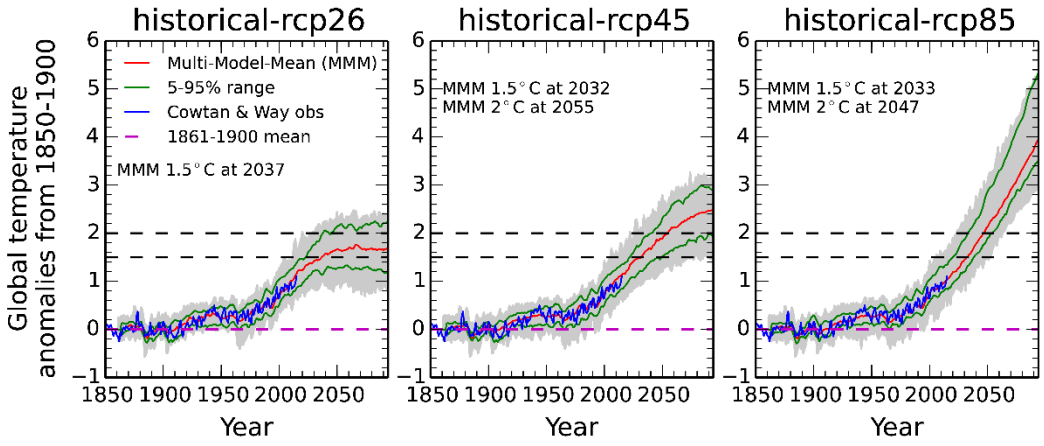

Fig S11 – Historical and future projections for global mean temperature (weighted). Identical to figure 1 main paper except that the 5-95% and 50% distributions are calculated from a weighted model distribution.

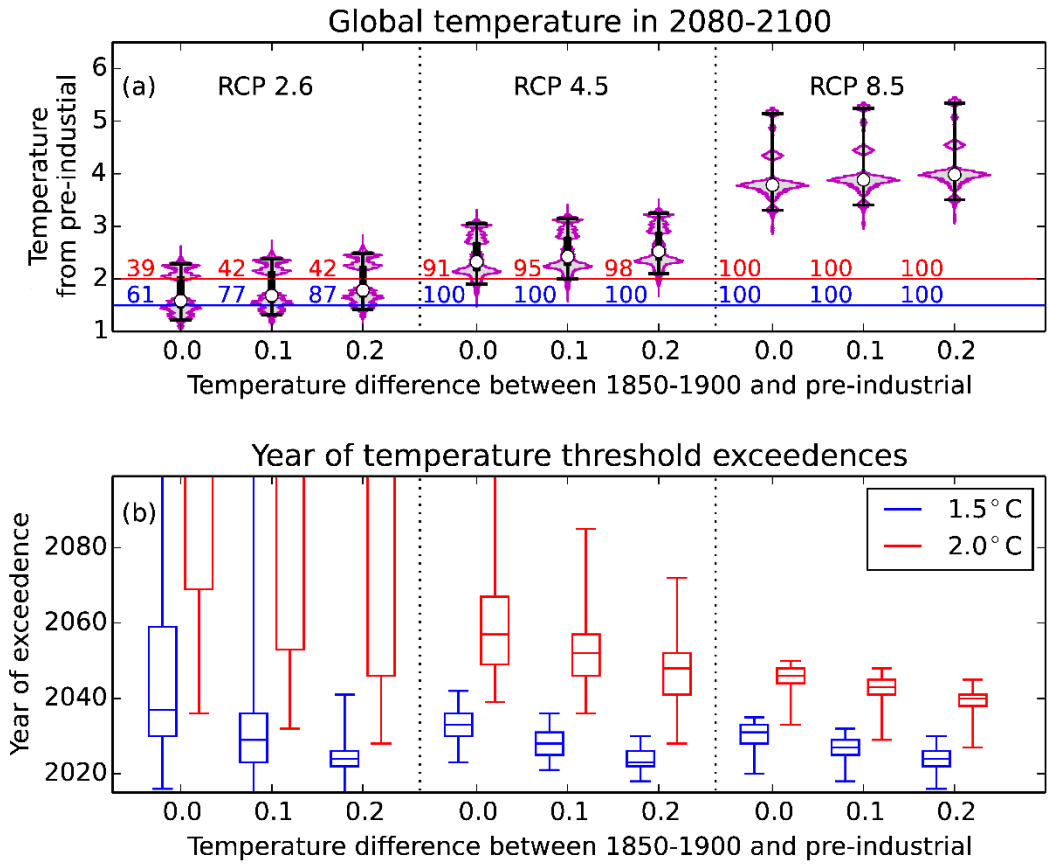

Fig S12 – Probability distributions for mean temperatures and times of threshold exceedance (weighted). Identical to fig. 4 in the main paper except that all distributions are calculated from weighted models.

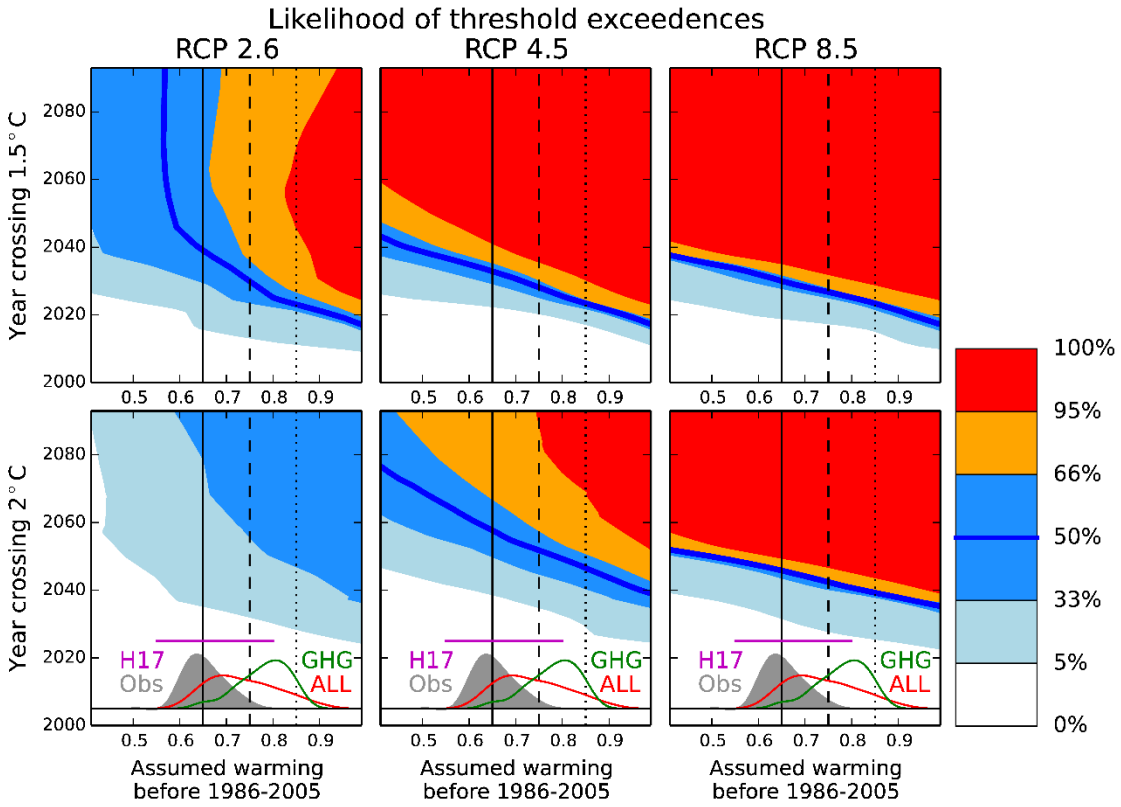

Fig S13 – Likelihood of exceeding temperature thresholds and its dependence on assumed pre-industrial baseline (weighted). Identical to figure 3 main paper except that the distributions are calculated from weighted models.

Sensitivity of the results to details of the analysis

In the main analysis anomalies were calculated from 1986-2005 and an amount of warming was added to account for the temperature change before this period (see Eqn. 1 in the method section). Figures S14, S15 and S16 show results where instead we follow Eqn. 2 (method section) and calculate anomalies from 1861-1900.

The differences between the methods is clearly shown in a comparison of figures 1, 3, 4 (main paper) with fig S14-S16. Our main conclusions are unaffected by these two different methodological choices. The only scenario which can keep global temperatures below 1.5°C is RCP2.6 and the likelihood is greatly affected by the definition of the pre-industrial baseline (PRE in eqs 1&2) regardless which method is used as is the timing of exceedance of both thresholds. The methodological choice does not greatly affect the mean warming or most likely time of threshold exceedance. It does however, effect the uncertainty ranges – with taking anomalies from 1861-1900, on the whole, increasing uncertainty.

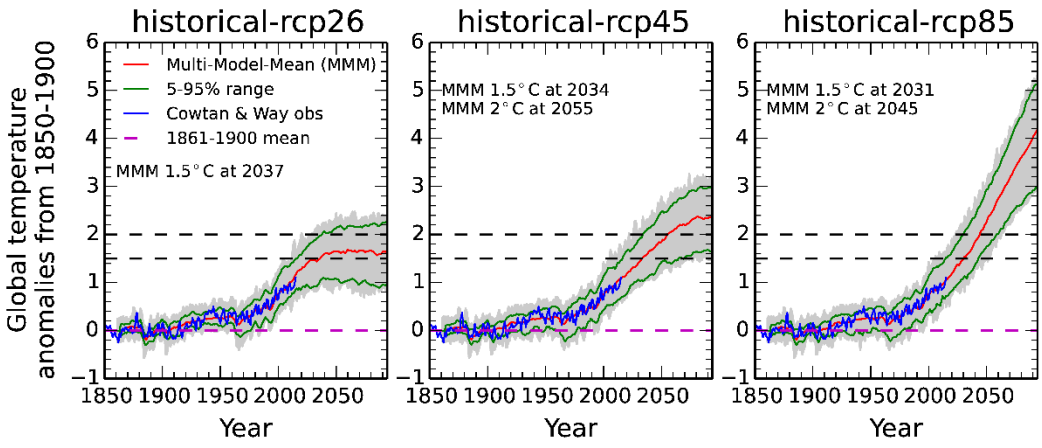

Fig S14 – Historical and future projections for global mean temperature (anomalies from 1861-1900). Identical to figure 1 main paper except that the model simulations have anomalies from 1861-1900 instead of 1986-2005.

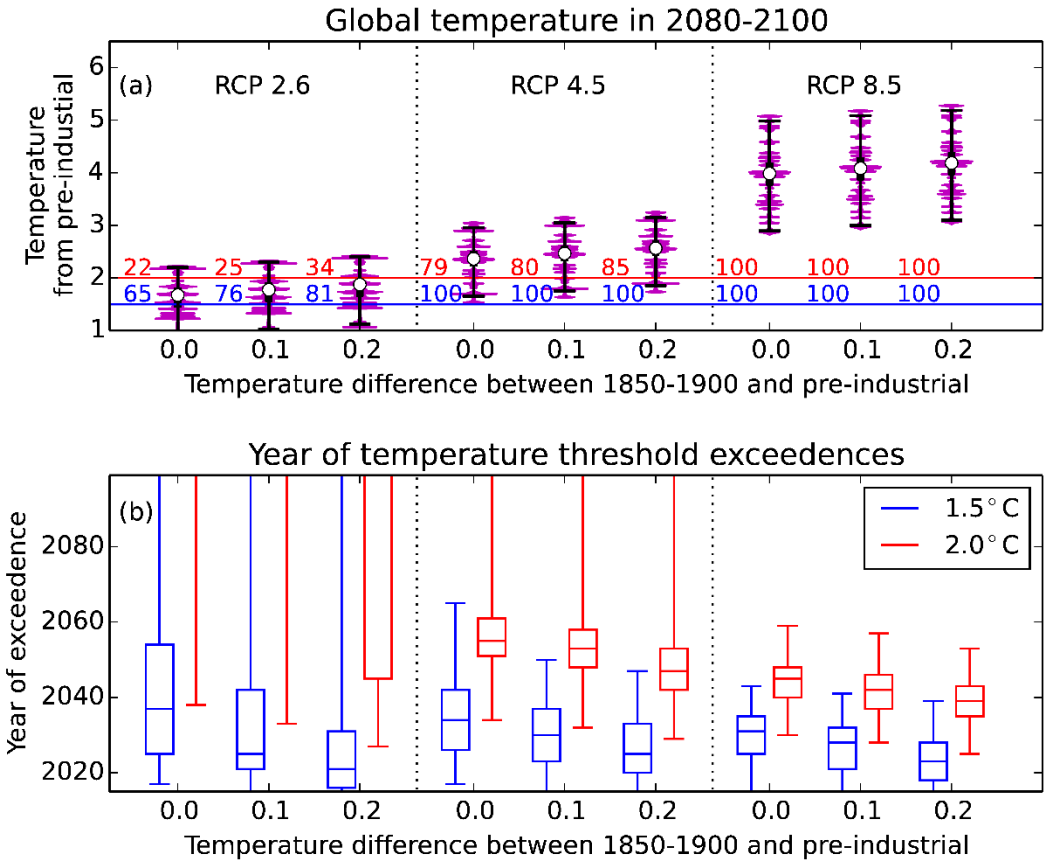

Fig S15 – Probability distributions for mean temperatures and times of threshold exceedance (anomalies from 1861-1900). Identical to figure 4 main paper except that all distributions are calculated from models with anomalies from 1861-1900

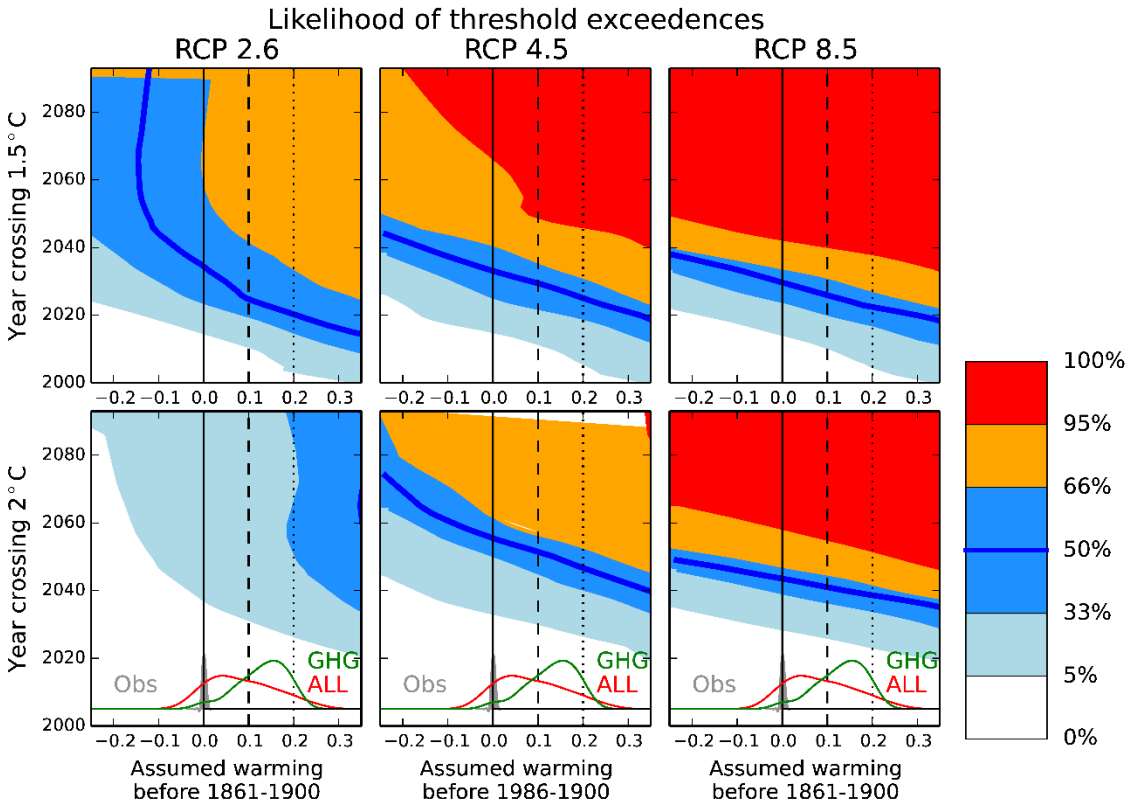

Fig S16 – Likelihood of exceeding temperature thresholds and its dependence on assumed pre-industrial baseline (anomalies from 1861-1900). Identical to figure 3 main paper except that the distributions are calculated from models which have anomalies from 1861-1900.

The use of global mean SAT instead of a blended mean of SSTs and SATs

Fig S17 show results which replicate the same analysis as in the fig S14 but using global SATs instead of a blended mean of SSTs and SATs. They show that if SATs are used projected temperature increases are larger than if blended temperatures are used. For mean temperature by 2080-2099 (if PRE in Eqn. 2 is zero), when using blended temperature this is 1.67°C, 2.26°C, 4.00°C for RCP2.6, RCP4.5, RCP8.5 respectively. While if SATs only are used this increases to 1.73°C, 2.55°C, 4.39°C for RCP2.6, RCP4.5, RCP8.5 respectively

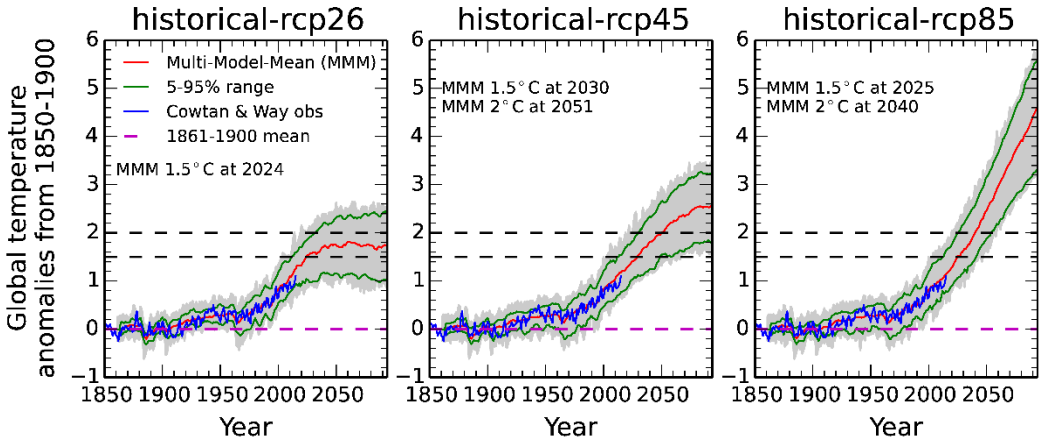

Fig S17– Historical and future projections for global mean surface air temperature. Identical to figure 1 except global mean SATs are plotted instead of a blend of SATs and SSTs and anomalies are taken from 1861-1900.

Effect of the use of a smoothing filter in fig 3

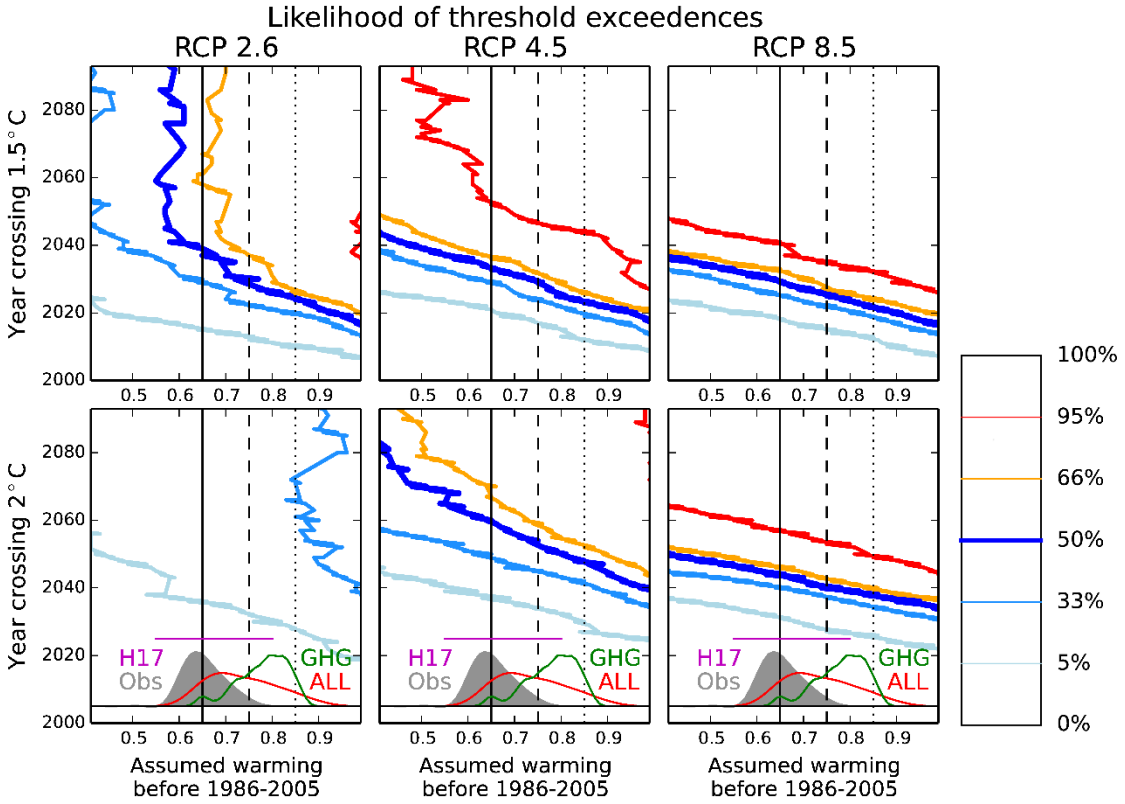

Fig S18 – Probability of exceeding temperature thresholds – Probabilities are shown by the coloured lines which indicate the location of the 5, 33, 50, 66 and 95 percentiles. Vertical lines indicate assumed warming of 0°, 0.1° and 0.2°. This is identical to figure 3 (main paper) except the percentile lines are not filtered and are not shaded.

References

1. Schurer, A. P., Tett, S. F. B. & Hegerl, G. C. Small influence of solar variability on climate over the past millennium. *Nat. Geosci.* **7,** 104–108 (2013).

2. Otto-Bliesner, B. L. *et al.* Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model. *Bull. Am. Meteorol. Soc.* **97,** 735–754 (2016).

3. Phipps, S. J. *et al.* The CSIRO Mk3L climate system model version 1.0 – Part 1: Description and evaluation. *Geosci. Model Dev.* **4,** 483–509 (2011).

4. Phipps, S. J. *et al.* The CSIRO Mk3L climate system model version 1.0 – Part 2: Response to external forcings. *Geosci. Model Dev.* **5,** 649–682 (2012).

5. Cowtan, K. *et al.* Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.* **42,** 6526–6534 (2015).

6. Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B. & Phipps, S. J. Separating Forced from Chaotic Climate Variability over the Past Millennium. *J. Clim.* **26,** 6954–6973 (2013).

7. Crowley, T. J. & Unterman, M. B. Technical details concerning development of a 1200-yr proxy index for global volcanism. *Earth Syst. Sci. Data Discuss.* **5,** 1–28 (2012).

8. Vieira, L. E. A., Solanki, S. K., Krivova, N. A. & Usoskin, I. Evolution of the solar irradiance during the Holocene. (2011). doi:10.1051/0004-6361/201015843

9. Steinhilber, F., Beer, J. & Fröhlich, C. Total solar irradiance during the Holocene. *Geophys. Res. Lett.* **36,** L19704 (2009).

10. Wang, Y. -M. *et al.* Modeling the Sun's Magnetic Field and Irradiance since 1713. *Astrophys. J.* **625,** 522–538 (2005).

11. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.* **4,** 33–45 (2011).

12. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res. Atmos.* **117,** n/a-n/a (2012).