# Supporting Information to "Internal variability in European summer temperatures at 1.5°C and 2°C of global warming"

**Laura Suarez-Gutierrez[1,2], Chao Li [1], Wolfgang A Müller [1], and Jochem Marotzke [1]**

[1] Max-Planck-Institut für Meteorologie, Hamburg, Germany.
[2] International Max Planck Research School on Earth System Modelling, Hamburg, Germany.

E-mail: `laura.suarez@mpimet.mpg.de`

## Evaluation of the MPI-ESM Grand Ensemble

In this section we illustrate the ability of the MPI-ESM Grand Ensemble to capture the observed variability in anomalies of global mean surface temperature (GMST) and European summer monthly mean temperature (EuST). GMST observations occur generally across the whole ensemble width and within the ensemble spread for the majority of the record (Fig. S1a). The histogram of the rank that the observed GMST data would have as a member of the ensemble presents a pattern that is reasonably flat and continuous (Fig. S1b). For EuST observations, we find that they occur rarely outside of the ensemble limits, however they tend to fall with higher frequency in the upper half of the ensemble, and with lower frequency towards the ensemble minimum (Fig. S1c and S1d). This tendency may indicate an overestimation of the frequency and amplitude of low European summer temperatures in the Grand Ensemble during some periods. However, the observational record may not be sufficiently long to allow for a complete characterisation of the large multi-decadal variability in European summer temperatures.

The probability distribution functions of GMSTs (Fig. S1e) and EuSTs (Fig. S1f) for the reference period of 1981–2010 again present good agreement between the simulated and observed estimates in both the shape and the amplitude of the distributions. Overall, our evaluation indicates that the MPI-ESM Grand Ensemble offers an adequate representation of the estimated internal variability in spatially averaged GMST and EuST observations.

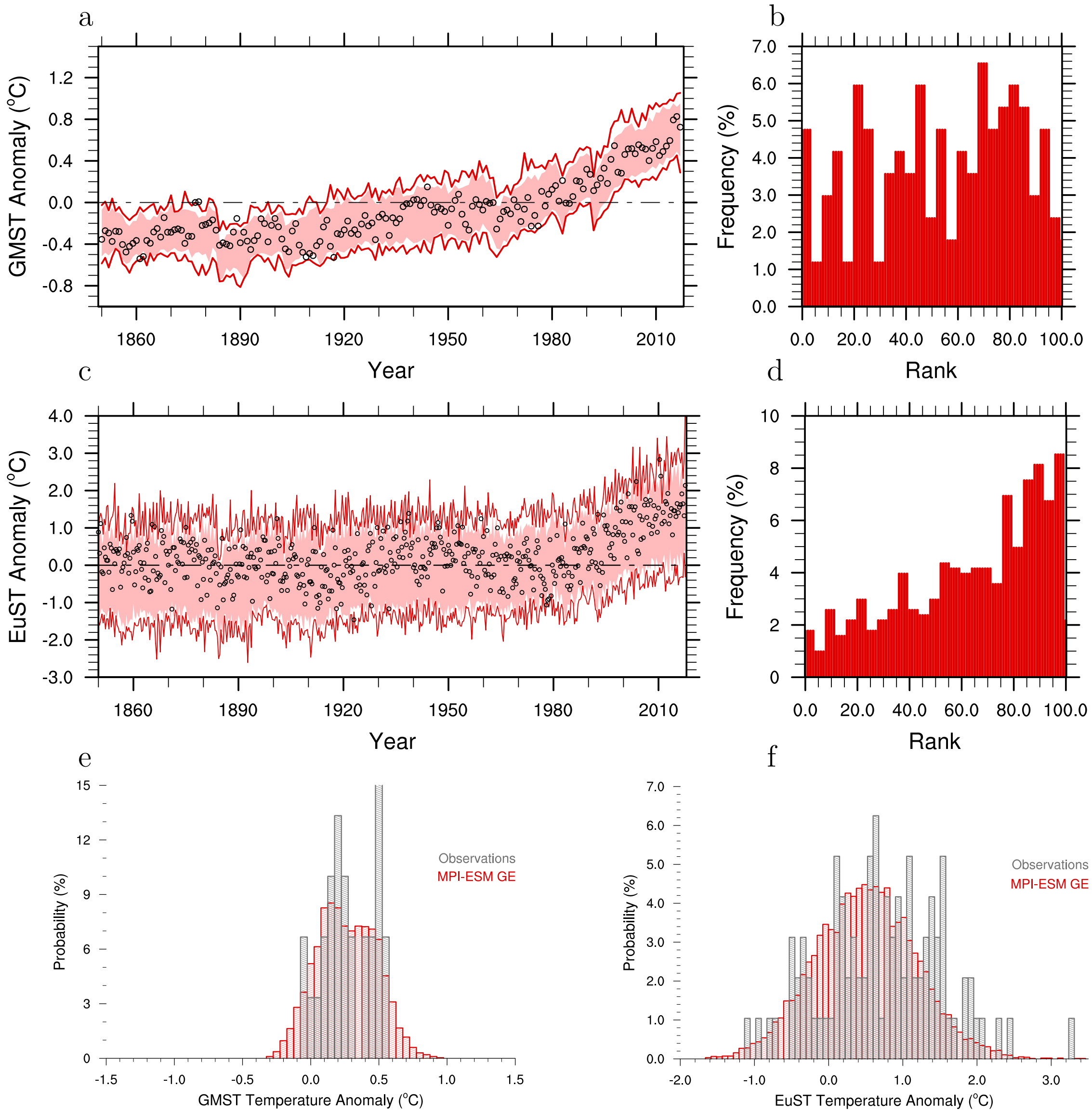


**Figure S1. GMST and EuST in the Grand Ensemble vs. observations. (a)** Time series of GMST anomalies simulated by the MPI-ESM Grand Ensemble (GE) (red) compared to HadCRUT4 observations (black markers). **(b)** Rank histogram for the HadCRUT4 GMST observations as a member of the Grand Ensemble for the period of 1850–2016. The frequency is normalized to unity and translated to percentage. **(c)** Time series of EuST anomalies simulated by the MPI-ESM GE (red) compared to CRUTEM4 observations (black markers), as in (a). **(d)** Rank histogram for the CRUTEM4 EuST observations as a member of the Grand Ensemble for the period of 1850–2017. The frequency is normalized to unity and translated to percentage. **(e)** Probability distribution of GMST anomalies simulated by the MPI-ESM GE (red) compared to HadCRUT4 observations (gray) for the period of 1981–2010. **(f)** Probability distribution of EuST anomalies simulated by the MPI-ESM GE (red) compared to CRUTEM4 observations (gray) for the period of 1981–2010. Simulations are historical runs for the period 1850–2005 and RCP4.5 for the period 2006–2017. Anomalies are calculated with respect to climatological levels defined by the the period of 1961–1990. Simulated data are subsampled to grid cells where observations are available.

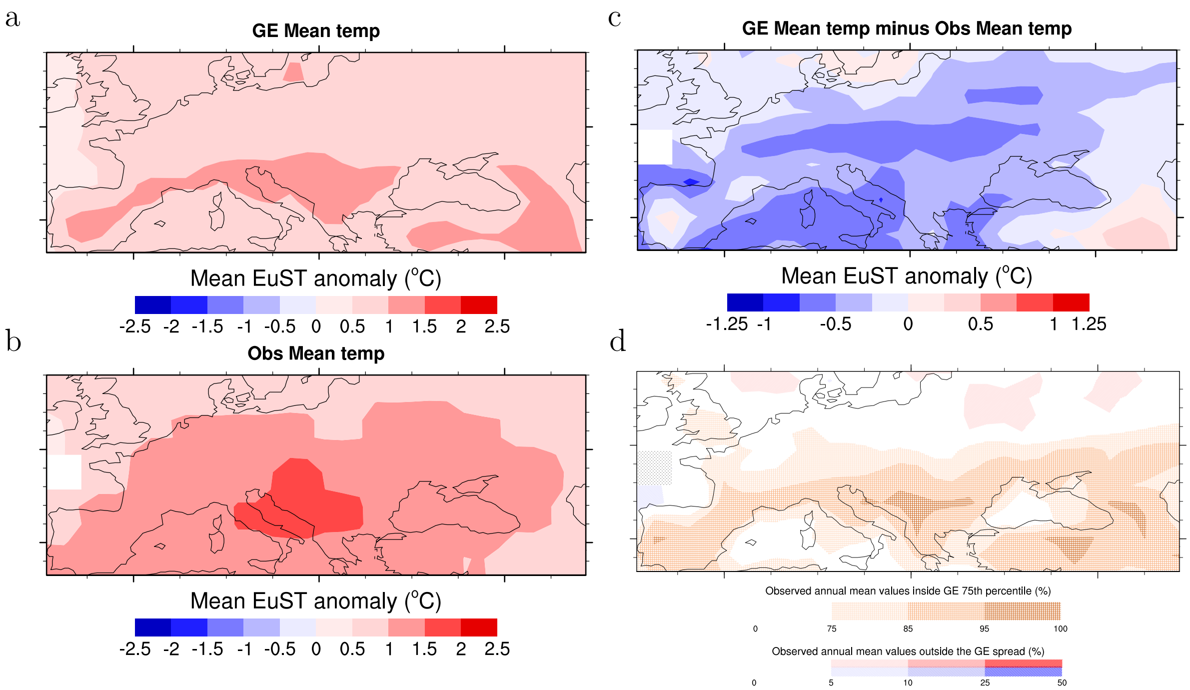


**Figure S2. EuST in the Grand Ensemble vs. observations. (a)** Ensemble mean EuST anomaly averaged over the period 1990–2017. **(b)** CRUTEM4 mean EuST anomaly averaged over the period 1990–2017. **(c)** Difference between Ensemble mean EuST anomaly minus CRUTEM4 mean EuST anomaly over the period 1990–2017. **(d)** Evaluation of the EuST variability simulated by the GE compared to CRUTEM4 observed variability for the whole observational record (1850–2017). Orange shading shows the regions where the observed estimated variability is smaller than the simulated variability, by representing the frequency with which observations occur within the 75th ensemble percentile. Red shading represents regions where the observed estimated variability is larger than the simulated variability, by representing the frequency with which observations are larger than the maximum EuST anomaly simulated by the ensemble at that time step. Blue shading also represents regions where observed estimated variability is larger than the simulated variability, by representing the frequency with which observations are lower than the ensemble minimum. Simulations are historical runs for the period 1850–2005 and RCP4.5 for the period 2006–2017. Anomalies are calculated with respect to the climatological level defined by the the period of 1961–1990.

Figure S2 illustrates the ability of the MPI-ESM Grand Ensemble to simulate observed summer monthly mean temperatures on average and estimated summer temperature temperature variability over Europe. For average temperatures in our current climate, represented by mean EuSTs over the period of 1990–2017, the ensemble mean temperatures are slightly lower than the observed EuSTs, with the largest differences around 0.5°C (Fig. S2c). We evaluate the ability of the ensemble to capture the estimated observed variability by calculating with which frequency observations are either not randomly distributed across the whole ensemble spread or occur outside the ensemble limits during the observational record (Fig. S2d). The orange shading represents how often observational estimates occur within the $75^{\mathrm{th}}$ percentile of the ensemble spread at each grid cell. In case the observed variability is correctly simulated by the ensemble, this frequency would equal to around 75% of the observed estimates. We find values of above 85% over Souther Europe, indicating that simulated temperature variability is slightly larger than the observed estimate in these areas. In case this analysis is performed for the frequency of observational estimates within the ensemble's 50$^{\text{th}}$ percentile, we find good agreement between observed and simulated variability estimates, with around 50% frequency for the whole domain. We also find that observational estimates occur mostly within the ensemble limits, indicating that summer temperature variability is not underestimated over Europe.

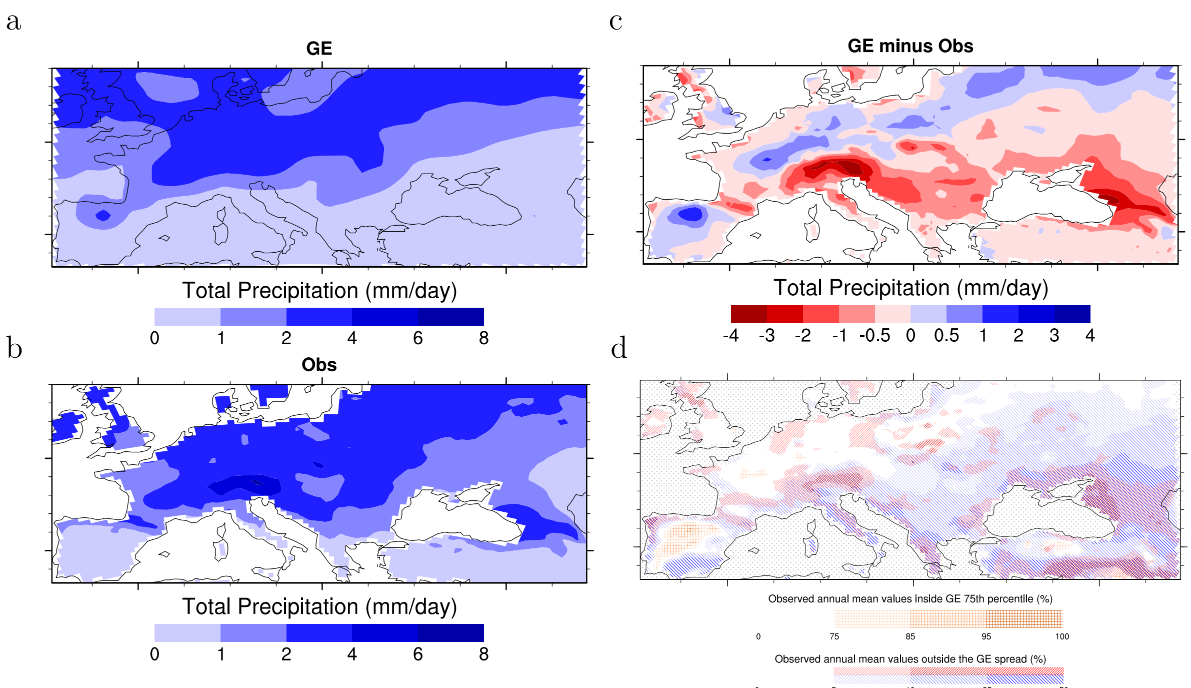


**Figure S3. Total precipitation in the Grand Ensemble vs. observations. (a)** Ensemble mean total precipitation averaged over the period 1990–2017. **(b)** E-OBS mean total precipitation averaged over the period 1990–2017. **(c)** Difference between Ensemble mean total precipitation minus E-OBS mean total precipitation over the period 1990–2017. **(d)** Evaluation of the variability in precipitation anomalies simulated by the GE compared to E-OBS observed variability for the whole observational record (1950–2017). Orange shading shows the regions where the observed estimated variability is smaller than the simulated variability, by representing the frequency with which the observed anomalies occur within the 75th ensemble percentile. Red shading represents regions where the observed estimated variability is larger than the simulated variability, by representing the frequency with which observed anomalies are larger than the maximum precipitation anomaly simulated by the ensemble at that time step. Blue shading also represents regions where observed estimated variability is larger than the simulated variability, by representing the frequency with which observed anomalies are lower than the ensemble minimum. Simulations are historical runs for the period 1950–2005 and RCP4.5 for the period 2006–2017. Anomalies are calculated with respect to the climatological levels defined by the the period of 1961–1990.

We perform a similar analysis for total precipitation over Europe (Fig. S3). In this case we find that the ensemble mean precipitation for the period 1990–2017 is around 1 mm/day larger than the observed average over the northern part of the domain and part of the Iberian Peninsula, while being around 1 to 4 mm/day lower than the observed average in Southern Europe, particularly over alpine regions (Fig. S3c). We also find that the ensemble tends to simulate precipitation variability that is smaller that the estimated observed variability over Southern Europe (Fig. S3d). The frequency of extremely wet and, particularly, extremely dry summer months is around 25% too low in the ensemble simulations in comparison with observations, represented by the frequency with which observed estimates lay below (represented in blue) or above (represented in red) the ensemble limits. This bias in summer precipitation variability occurs over the same area where the bias in EuST variability occurs. These results appear to agree with our findings in Fig. S1d and Fig. S2d, that indicate some overestimation of the frequency of colder than average summer months in the ensemble simulations that may be partially caused by biases in precipitation variability.

## Supplementary figures

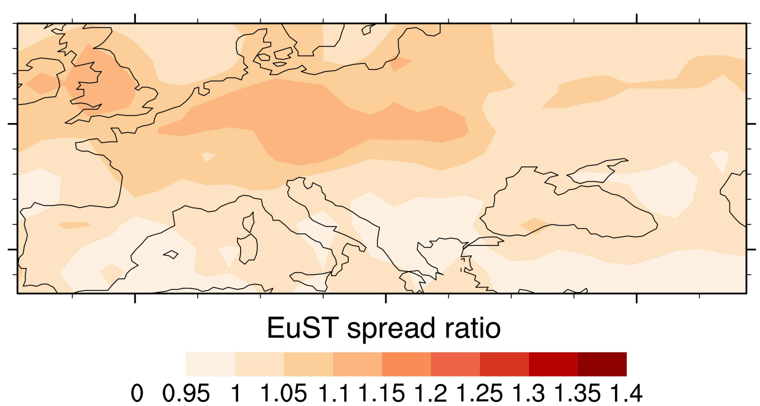


**Figure S4. Variability change in European summer monthly mean temperatures.** Ratio of the spread of EuST anomalies at 2°C of global warming over EuST spread at 1.5°C of global warming above pre-industrial conditions, simulated by the MPI-ESM Grand ensemble.

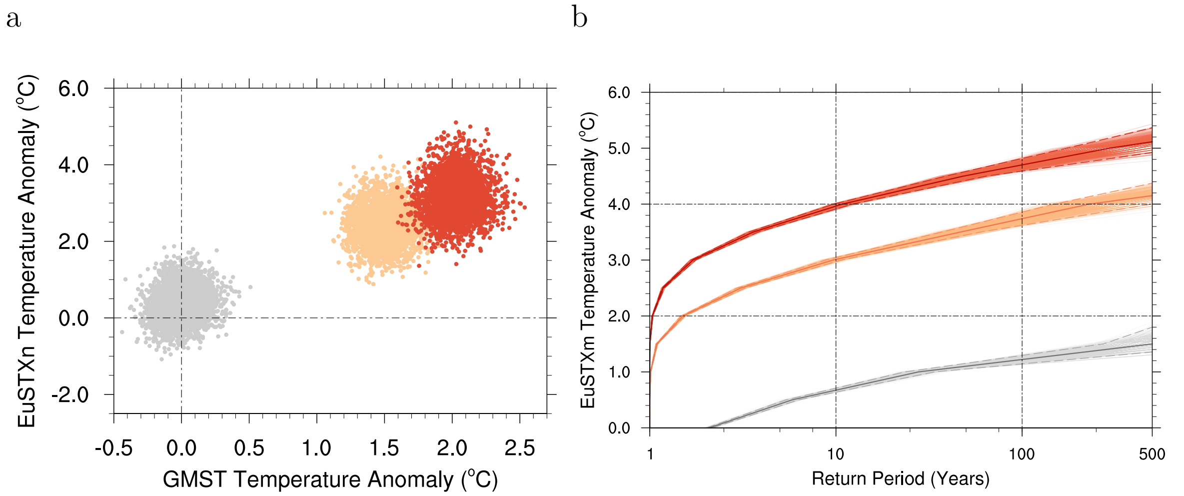


**Figure S5. European summer minimum value of daily minimum temperature (EuSTXm) at different global warming levels. (a)** EuSTXm anomalies against GMST anomalies for pre-industrial conditions (gray), and for global warming levels of 1.5°C (orange) and 2°C (red) above pre-industrial conditions, simulated by the MPI-ESM Grand ensemble. **(b)** Return levels of EuTXm summer block minima against their return period, represented by the thick solid lines in gray for pre-industrial conditions, in orange for global warming levels of 1.5°C and in red for 2°C above pre-industrial conditions. Uncertainty in these return levels is estimated by bootstrap-resampling with replacement. The coloured thin lines represent 1000 individual bootstrap estimates; the coloured dashed lines represent the 95% confidence intervals.