How large does a large ensemble need to be?

Sebastian Milinski, Nicola Maher, Dirk Olonscheck
Max Planck Institute for Meteorology, Hamburg, Germany

Manuscript: https://doi.org/10.5194/esd-2019-70 **Contact:** sebastian.milinski@mpimet.mpg.de

This display will also be presented in a live streaming session on May 8 at 16:30 CEST (link to programme)

Motivation

Large ensemble studies have demonstrated that the ensemble size needed to characterise a change in climate depends on the variable, the model, the signal that we want to detect, and the acceptable error. We develop an objective method to estimate the necessary ensemble based on the 100-member MPI Grand Ensemble and a long control run.

A recipe for estimating the required ensemble size

1. Define the question.
2. Choose an error metric and an acceptable error magnitude.
3. Estimate the error for different ensemble sizes by subsampling a long control run or a large ensemble with transient forcing.
4. Determine the minimum ensemble size that is required to reduce the error below the threshold chosen in step 2.
5. The required ensemble size derived in this way is robust if it is below 50% of the sample size used in the analysis.

The resampling problem

We can estimate which ensemble size is sufficient to represent the forced response by comparing the ensemble mean of a smaller ensemble to the full ensemble.
By construction, this difference will always converge to zero when approaching the full ensemble size as shown in figure 3. When approaching the full ensemble size, the estimate is biased low and the required ensemble size could be substantially larger than the estimate. The resampling problem affects the estimate when more than 50% of the full sample are used for generating subsamples.

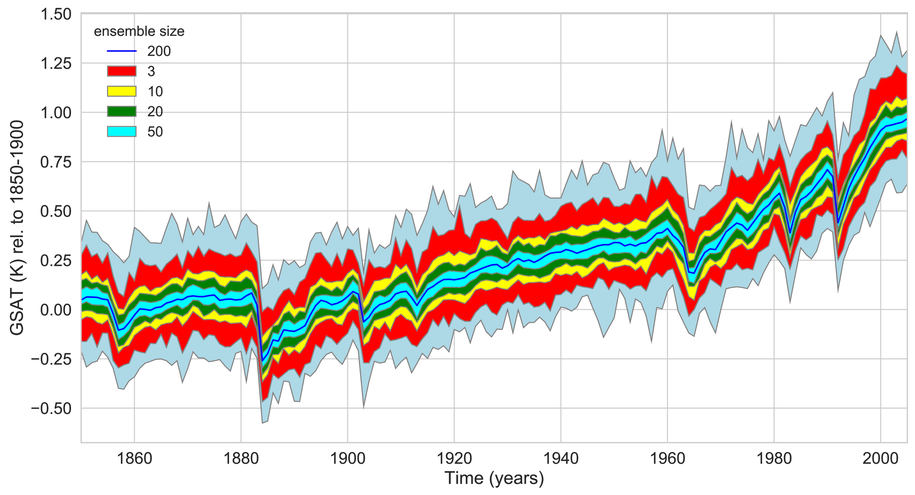

Figure 1: Ensemble mean GSAT for 200 members (blue line) and ranges of ensemble means for different ensemble sizes (1000 random samples)

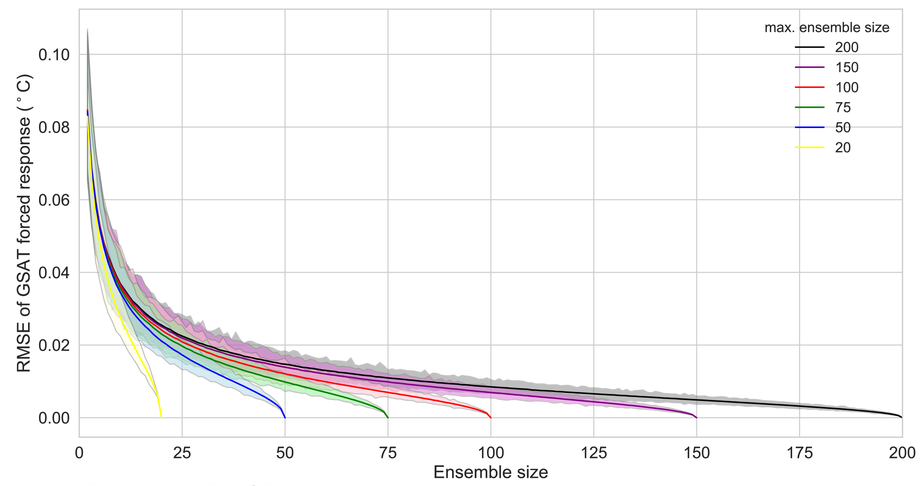

Figure 2: RMSE for ensemble means from smaller ensemble sizes compared to the ensemble mean for the full ensemble. Different colours represent different sizes of the full ensemble.

Example 1: estimating the historical forced response in temperature

Depending on the region and acceptable error, fewer than 10 or more than 50 ensemble members are required: more members are required over land than over the ocean, and more in high latitudes than in the tropics.

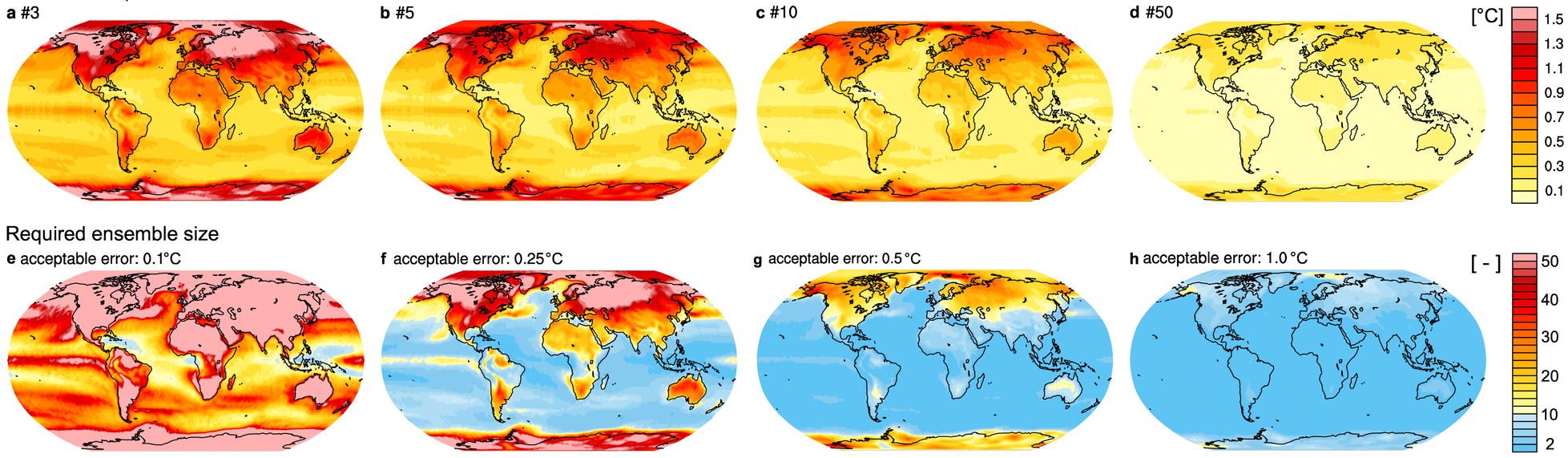

Figure 3: a-d, The mean RMSE for the forced response in historical monthly mean near-surface air temperature of MPI-GE for a, 3,b, 5, c, 10, and d, 50 ensemble members relative to the 100-member mean, globally. e-h, Required ensemble size to capture the 100-member mean forced response in historical monthly mean near-surface air temperature dependent on the acceptable error of a, 0.1, b, 0.25, c, 0.5, and d, 1.0 °C.

Example 2: quantifying internal variability

A region with large variability, such as the tropical Pacific, requires more members to quantify internal variability than a region with less variability, such as central America. In a 5-member ensemble, the Niño 3.4 standard deviation for a 30-year period can randomly vary between 0.5°C and 0.9°C. Forced changes in Niño 3.4 variability that fall within this range can not be detected in a 5-member ensemble.
The ensemble size needs to be chosen based on the the acceptable error or expected magnitude of an externally forced change.

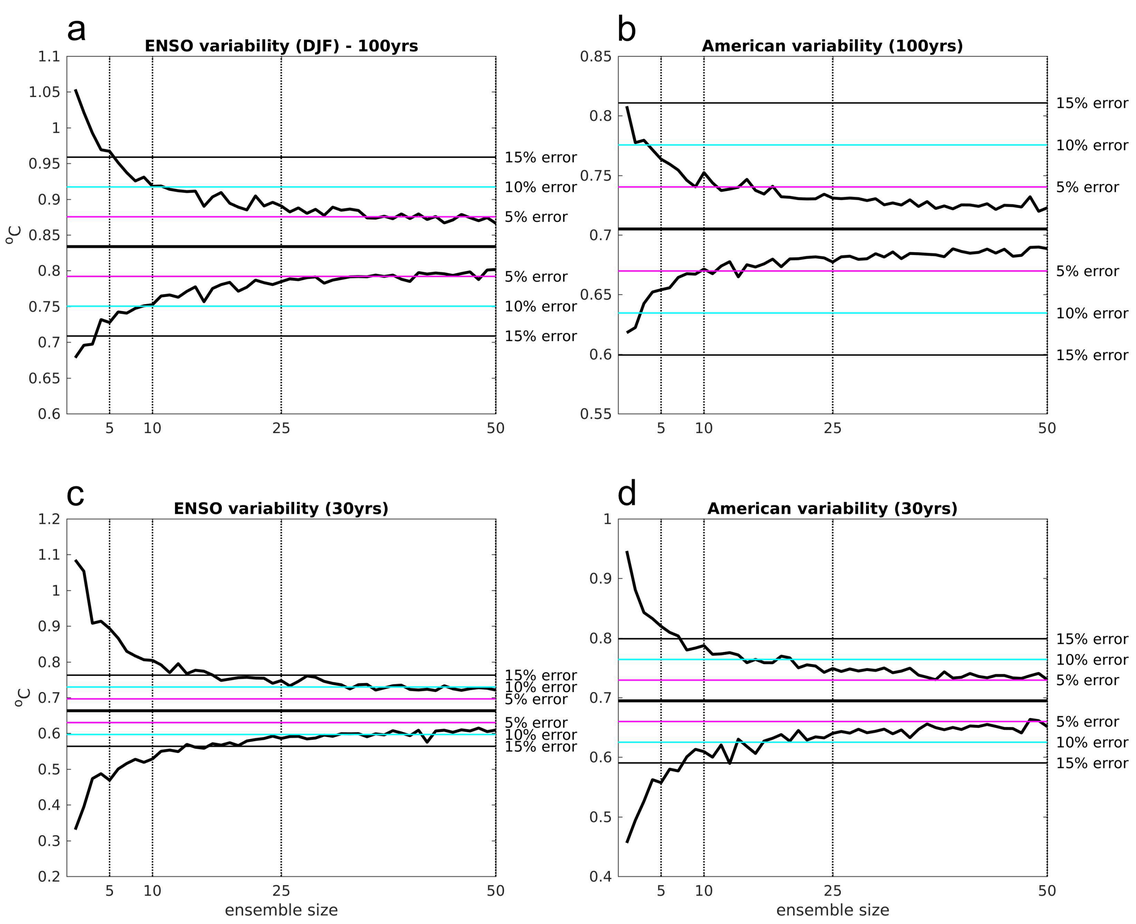

Figure 4: We show for increasing ensemble sizes the: a) ENSO variability in the Niño3.4 box calculated over 100 year periods, b) Central American variability calculated over 100 year periods, c) ENSO variability in the Niño3,4 box calculated for DJF over 30 year periods, d) Central American annual mean variability calculated over 30 year periods. All indices are calculated from the 2000 year MPI-GE control run. Each index is calculated as a running value at each time-step in the control. Ensembles of 1 to 120 members are created by randomly sampling the control simulation without replacement. For each ensemble size we create 1000 artificial ensembles. The estimated true value is calculated by using the entire 2000 years of the control and is shown in the horizontal black line. The maximum and minimum values of each index from the 1000 samples are shown in the solid black lines.

Max-Planck-Institut für Meteorologie

MAX-PLANCK-GESELLSCHAFT