Temperature Variability Projections Remain Uncertain after Constraining Them to Best Performing Large Ensembles

Laura Suarez-Gutierrez [1,2] & Nicola Maher [3,4]

1. Wageningen University & Research, The Netherlands
2. ETH Zurich, Switzerland
3. Australian National University, Canberra, Australia.
4. ARC Centre of Excellence for the Weather of the 21st Century, Australia

Article: Suarez-Gutierrez & Maher. Nat Comms (2026)

Contact: Laura.suarez@wur.nl

Constraining Temperature Variability Projections

Models show large disagreement on the magnitude and even the change of temperature variability changes under future warming.

Can we reduce uncertainties in future changes by constraining projections to models that capture historical variability correctly?

Why: Variability change projections are crucial to determine changes in the frequency and intensity of extremes, and the range of temperatures ecosystems and society need to adapt to in a warming world; yet they remain highly uncertain. [IPCC AR6, 2021; van der Wiel & Bintanja (2021); Bathiany et al., 2018; Schär et al., 2004]

How: We constrain variability changes projections for future warming levels based on the Single Model Initial-Condition Large Ensembles (SMILEs) that best represent historical temperature variability. By using SMILEs we ensure a robust sampling of internal variability as well as a more reliable model evaluation. We use rank-frequency analysis [Suarez-Gutierrez et al., 2021] on 11 CMIP5 and CMIP6 SMILEs against GISTEMPv4 over land and ERSSTv5 observations over the oceans.

Constraining projections to best-performing SMILEs still leaves large uncertainties in the magnitude and even in the sign of variability change.

The constrained ensemble decreases model spread in some regions. However, in many regions good performing models still disagree in their projections, or too few models capture historical variability adequately.

Historical variability is captured poorly by most models in Australia, India, South America, and Africa, particularly in the local summer, and Polar and Southern Oceans. In these regions constrained change is larger, and unconstrained multi-model means underestimate future variability change.

The SMILEs that best represent isolated historical temperature variability or temperature variability and forced changes jointly are:
→ **CESM-LE & CESM2-LE, with GFDL-SPEAR-MED and MPI-GE5 third best in each category, respectively.**

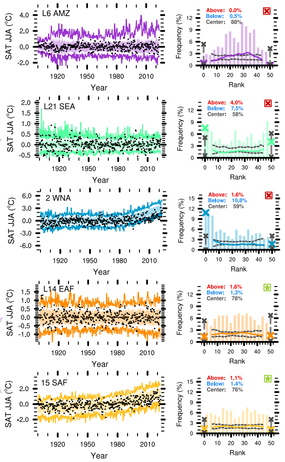
Variability Evaluation Framework

We use **rank-frequency analysis** to assess biases in the models simulated variability. Biases include:

a. **Observations cluster within central ensemble spread** (∩-shaped Rank Histogram)
→ **variability overestimation** bias

b. **Observations do exceed ensemble minimum or maximum excessively** (Slanted or U-shaped Rank Histogram)
→ **variability underestimation** bias
or (for non-detrended data)
→ **trend over/underestimation** bias

When none of these biases are present for a given season and region, the model is considered to **capture historical variability adequately** (Flat rank histogram within perfect model range)



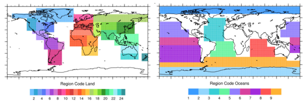
▲ **Fig. 1: Rank-Frequency Time Series Evaluation** Evaluation example for CanESM5 and selected regions. Time series (left) show ensemble maxima and minima (lines) and 75th central ensemble spread (shading; 12.5-87.5th percentiles) against observations (black dots). Rank histograms represent the place that observations take in a list of ensemble members ordered by ascending anomaly values (color bars; line 10-rank rolling mean). Crosses mark the frequency of the rank minimum (0) and maximum (number of members). Perfect-Model Rank Ranges (gray, 5-95th percentiles) represent the range of ranks each ensemble member takes in a list of the remaining ensemble members, as if it were the observations.

We use 2 criteria to evaluate model performance:

1. **Regional Perfect-model Rank Range:** Observations rank histogram for regionally averaged temperatures lies within the perfect-model rank range
2. **Grid-cell Threshold-based Range:** 50% or more of the grid-cells in the region are unbiased.

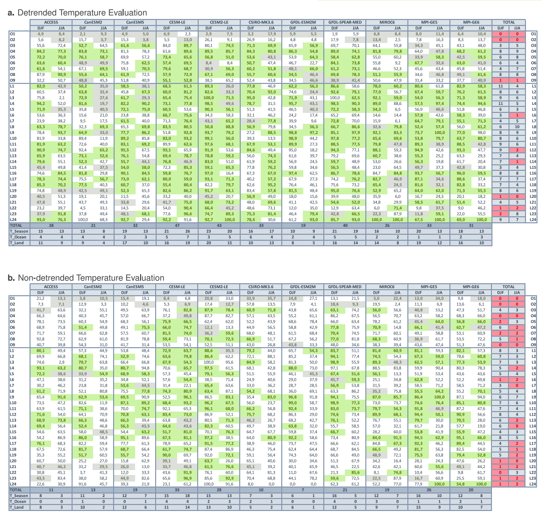
Models that capture isolated historical variability adequately for detrended data are selected for the constrained ensemble for the given region and season (Fig. 2 a)

We also show evaluation results for all models for non-detrended data (Fig. 2b)



Full evaluation results for all regions and 11 SMILES:

Supp. Information Suarez-Gutierrez & Maher (2026)



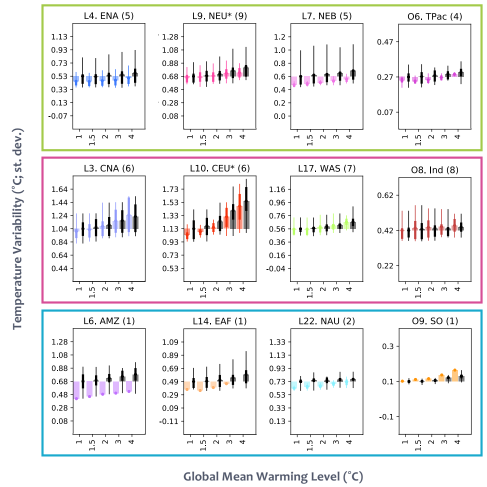
▲ **Fig. 2: Temperature Variability Evaluation results** for land (L1–L24; top left map) and ocean (O1–O9; top right map) regions for 11 SMILES for (a) detrended and (b) non-detrended monthly temperature anomalies in DJF and JJA. Shading marks the fulfilment of Criteria 1 (rank histogram within perfect-model rank range). Numbers mark the percentage of unbiased grid cells (exceeding the ensemble maxima or minima less than 8% of the months and occurring within the central 75th percentile ensemble bounds less than 80% of the months). A model is adequate when it fulfils Criteria 1 (green and gray shading) and Criteria 2 (at least 50% of unbiased grid cells).

Variability Changes under Warming: Unconstrained versus Constrained Projections

The constrained ensemble yields increased agreement (less inter-model spread) in the magnitude and the sign of variability change in some regions, while others still exhibit similar or even larger uncertainty in the magnitude and even sign of the constrained projections of temperature variability change.

Over some regions, historical variability is captured poorly by most models. In these regions, constrained variability change is larger than unconstrained projections, suggesting unconstrained multi-model means underestimate future variability change.

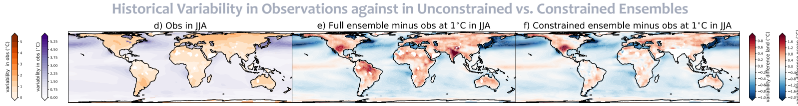
Variability Change Projections in Unconstrained vs. Constrained Ensembles



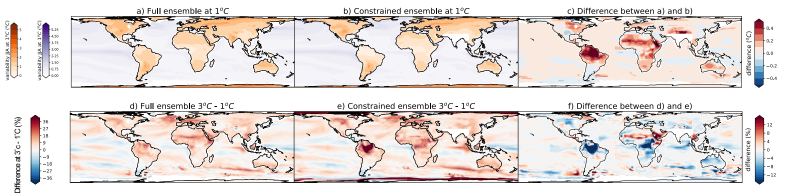
◀ **Fig. 3: Unconstrained vs. Constrained multi-model JJA temperature variability projections** (st. dev.) for selected regions at different warming levels for the unconstrained (black bars) and constrained ensembles (color bars), relative to the variability at 1 degree of warming in the unconstrained ensemble. Error lines show the full model spread (thin lines) and the 25th and 75th ensemble percentiles (thick lines). Number of adequate models in the constrained ensemble shown in brackets. Polar oceans are excluded due to sampling issues at ice edges.

Historical variability biases are largely removed in the constrained ensemble, by design (Fig. 4, top). The full unconstrained ensemble overestimates historical variability over large regions (Fig. 4, middle). Constrained projections show variability changes that are mostly larger (especially in Amazon, East Central Africa) and in some regions smaller (e.g., Arctic NA, Central Africa) than unconstrained projections (Fig. 4, bottom)

Historical Variability in Observations against in Unconstrained vs. Constrained Ensembles



Historical Variability & Projected Changes in Unconstrained vs. Constrained Ensembles



▲ **Fig. 4: Observed (top) and simulated historical variability estimates (middle) and future variability changes (bottom) for the full unconstrained and constrained ensembles** for temperature variability magnitude at 1°c of mean global warming in observations vs. models (top rows; multi-model means of st. dev. across monthly temperature anomalies), and temperature variability changes between 3 °C and 1 °C of global warming (bottom rows; shown as percentage of variability at 1°C)

ETH zürich

Funded by the European Union

WAGENINGEN UNIVERSITY & RESEARCH