



Supplement of

METEORv1.6: Spatial climate variability and integrated impact emulation

Benjamin M. Sanderson et al.

Correspondence to: Benjamin M. Sanderson (benjamin.sanderson@cicero.oslo.no)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Overview

This supplement collects the validation and diagnostic material supporting the monthly variability model, METEOR-NOISE, described in Section 3 of the main text. We begin with the fraction of variance the model accounts for in each variable (Table S1), then present out-of-sample validation for the five CMIP6 models used in the main text, rank histograms at point, regional and global scale, a demonstration of the global-mean scaling procedure, and finally two groups of diagnostics that probe the assumptions the model rests on: whether the spatial patterns and temporal dynamics it learns stay fixed as the climate warms, and whether the statistical distributions it assumes hold up against the training data.

METEOR-NOISE builds its representation of monthly variability in two stages, and Table S1 separates their contributions. The model first fits a seasonal cycle whose amplitude and phase depend on global temperature; the “Seasonal” column reports the fraction of the raw monthly variance this stage removes. It then represents the remaining anomalies with a set of leading spatial patterns and their month-to-month evolution, and the “PCA” column reports the fraction of that residual variance recovered by the leading 40 patterns. The “Total” column combines the two stages. Temperature is dominated by its seasonal cycle and is reproduced almost completely, above 97% in every model. Precipitation, whose seasonal cycle explains a much smaller share of its variance, retains considerably more that is unexplained, between roughly 35 and 50% depending on the model.

Table S1: Variance decomposition for the METEOR noise model across five CMIP6 models. Seasonal R^2 shows variance explained by the temperature-dependent harmonic model; Anomaly R^2 (PCA) shows variance of residuals explained by 40 principal components; Total R^2 combines both components.

Model	Temperature			Precipitation		
	Seasonal	PCA	Total	Seasonal	PCA	Total
CanESM5	91.6%	76.8%	98.1%	34.4%	33.1%	56.1%
CESM2-WACCM	90.7%	80.3%	98.2%	33.5%	45.4%	63.7%
CNRM-ESM2-1	91.0%	75.3%	97.8%	30.9%	37.5%	56.8%
IPSL-CM6A-LR	91.3%	76.0%	97.9%	31.3%	28.5%	50.9%
UKESM1-0-LL	90.6%	75.7%	97.7%	31.7%	34.9%	55.5%

S2 Stationarity of PCA and VARX components

METEOR-NOISE learns two things once, from a single training simulation, and then reuses them unchanged for every scenario it generates: a set of fixed spatial patterns (the PCA basis, or empirical orthogonal functions, of Section 3.3 in the main text), and the coefficients that govern how the amplitude of each

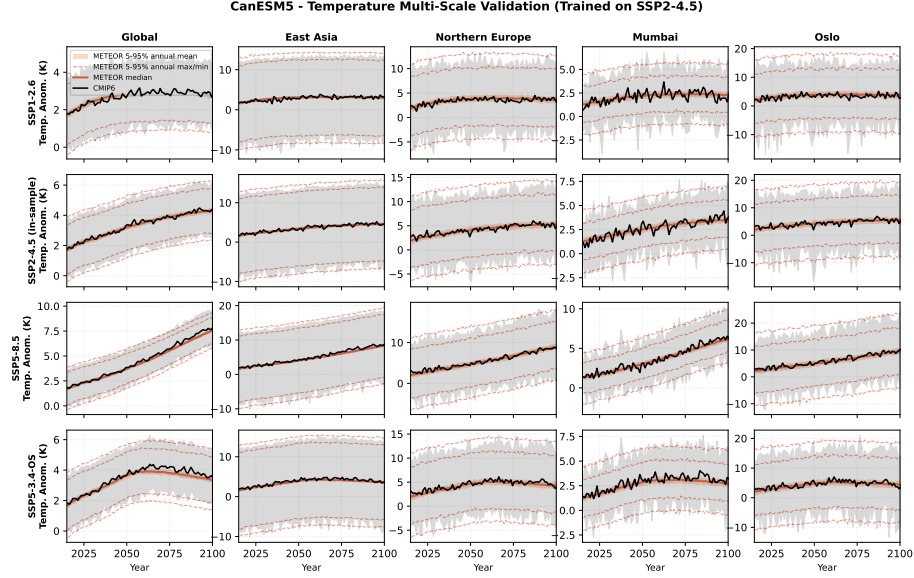


Figure S1: Out-of-sample temperature validation for CanESM5 trained on SSP2-4.5. Layout as in Figure 7.

pattern evolves from one month to the next (the VARX matrices of Section 3.4). Reusing them this way keeps generation cheap, but it works only if both are genuinely stable features of the model’s climate. They could instead drift as the world warms, and this is the assumption statisticians call stationarity. If it were to fail badly, variability generated for a warm end-of-century scenario would inherit the structure of a cooler training period. This section tests the assumption directly for both components, using the NorESM2-MM SSP2-4.5 training record. The analysis is contained in the notebook `notebooks/notebooks_v16_description_paper/Stationarity_Diagnostics.ipynb` in the METEOR code repository.

The PCA basis is tested by refitting the noise model separately on the early (1850–1974) and late (1976–2100) halves of the record and comparing the resulting EOFs by their absolute spatial pattern correlation, allowing best-match reordering. Panels (a) and (b) of Supplementary Figures S15 and S16 show the correlation matrix and the top-three EOF pair maps for temperature and precipitation respectively. The leading pattern is essentially unchanged in both variables ($|r| = 0.92$ and 0.98) and visually captures the same physical structure — Arctic amplification for temperature, an ENSO-related tropical Pacific pattern for precipitation. Mid-rank modes exhibit reordering (same-index correlations average 0.39 for tas and 0.43 for pr) but retain broadly similar patterns under best-match pairing (top-10 best-match means of 0.81 and 0.77).

The VARX coefficients describe how the pattern amplitudes in a given month depend on those of the previous two months, together with the influence of global

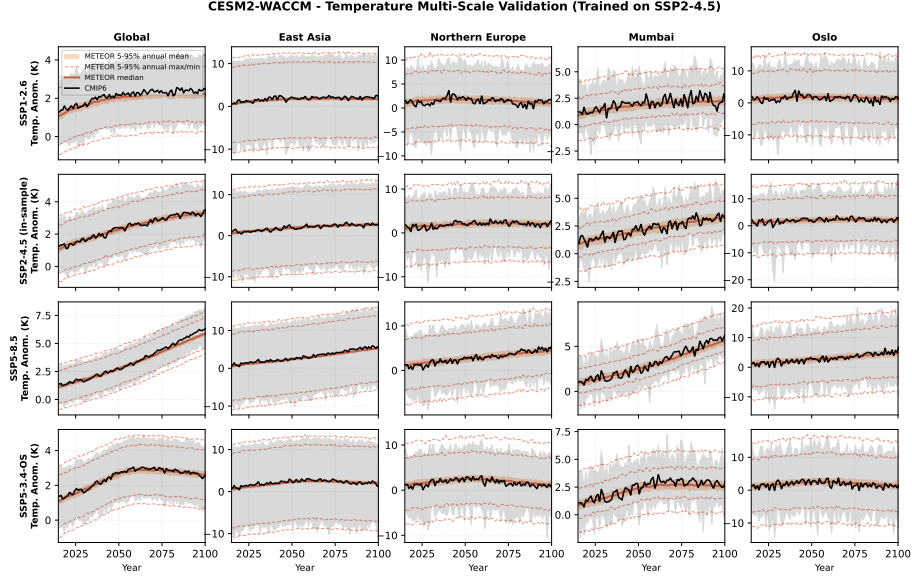


Figure S2: Out-of-sample temperature validation for CESM2-WACCM trained on SSP2-4.5. Layout as in Figure 7.

temperature. Their stability matters in two ways: the variability the model generates should neither decay away to nothing nor grow without bound, and its persistence and cross-mode structure should keep roughly the same character across the record. We test this by fitting the noise model on seven overlapping 100-year windows sliding through the record in 25-year steps, with center years 1900, 1925, ..., 2050, and summarise the resulting coefficient stability in panels (c)–(e) of the same figures. The spectral radius of the companion matrix is a single number for how quickly a perturbation to the modeled fluctuations decays; a value below one, the unit-root threshold, means the process is stationary, so shocks fade rather than amplify. It stays within a narrow band across all seven windows (0.935–0.961 for temperature, 0.902–0.914 for precipitation), comfortably below one. The overall size of the two autoregressive matrices A_1 and A_2 , measured by their Frobenius norm (the square root of the summed squares of every entry), varies by less than 5% for both matrices in precipitation and by 5–10% for temperature. The exogenous coefficient matrix B varies more, but this largely reflects the changing variance of the smoothed T_{glob} input rather than genuine coefficient drift, and it does not affect the stability of the autoregressive process; it is not shown in the summary panels.

The one systematic pattern is a rise in the self-lag coefficient of the leading temperature mode, $A_1[0,0]$, from 0.32 in the earliest window to 0.62 in the modern period. Two interpretations are consistent with the data. First, the leading tas EOF may progressively absorb more of the forced warming signal in later windows: the pattern is dominated by Arctic amplification, whose variance

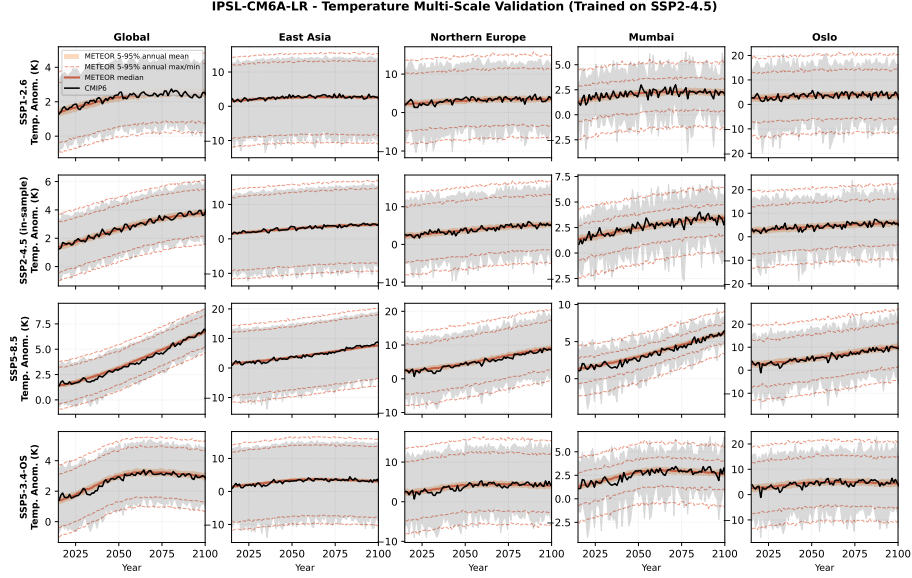


Figure S3: Out-of-sample temperature validation for IPSL-CM6A-LR trained on SSP2-4.5. Layout as in Figure 7.

in early windows is small compared to internal fluctuations but whose modern-period signal is increasingly monotonic. On this reading the apparent increase in self-persistence reflects a shift in what the leading EOF captures, not a change in the underlying dynamics. Second, the underlying persistence timescale of the leading mode may itself evolve with warming. The present diagnostic cannot distinguish between these interpretations; we flag this as an open question for future work.

Overall, the PCA and VARX components are strongly stationary at aggregate level. Leading spatial patterns are preserved and the VAR system remains uniformly stable across two centuries of window centers. The fixed-basis, fixed-coefficient design is thus a working approximation rather than a rigorous invariant: for scenarios well beyond the training envelope (e.g. SSP5-8.5 with training on SSP2-4.5), or for applications sensitive to specific mid-rank modes or to the persistence of the leading mode, users may wish to retrain on scenario-specific windows, or to explore the time-varying or state-dependent parameterisations planned for future versions.

S3 Additional diagnostics

The following three figures address reviewer-requested diagnostics on three separate choices in the noise model. The first concerns the smoothing applied to the global-temperature input T_{glob} before it enters the seasonal fit. The second

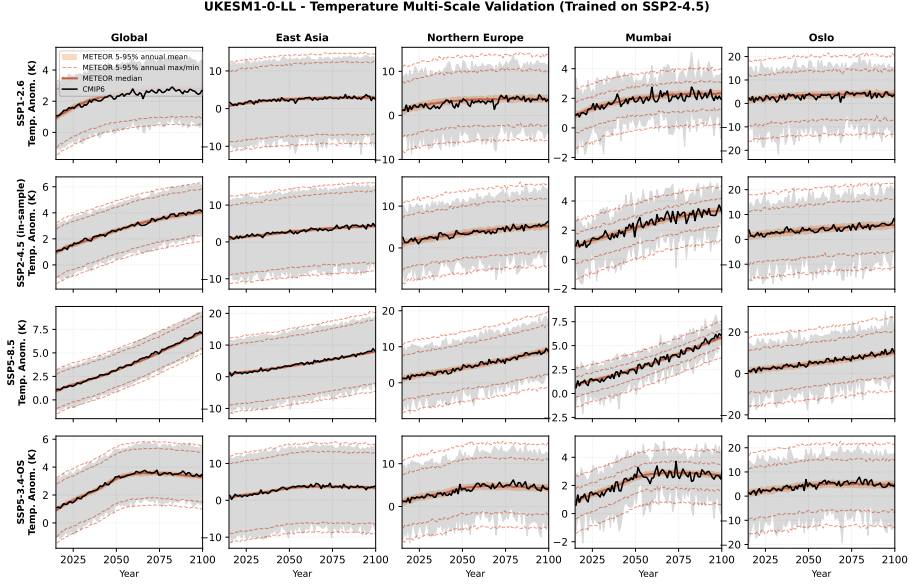


Figure S4: Out-of-sample temperature validation for UKESM1-0-LL trained on SSP2-4.5. Layout as in Figure 7.

concerns the random monthly shocks that drive the VARX process, the “innovations” $\varepsilon(t)$, which the model draws from a Gaussian distribution: here we ask how closely the shocks inferred from the training data actually follow that assumption. The third concerns the routine that fits a gamma distribution to precipitation at each gridpoint, and specifically how often it has to fall back on a fixed default when the fit cannot be computed. All are produced by the notebook `notebooks/notebooks_v16_description_paper/Additional_Reviewer_Diagnostics.ipynb` in the METEOR code repository. Chi-squared uniformity p-values for the reviewer-requested formal test on the rank histograms are added directly to Figures 5 and 9.

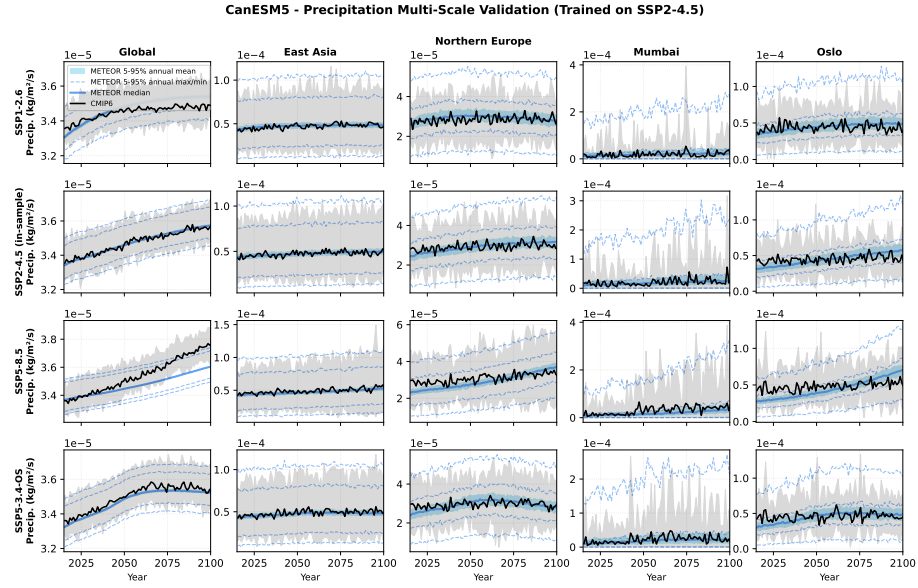


Figure S5: Out-of-sample precipitation validation for CanESM5 trained on SSP2-4.5. Layout as in Figure 8.

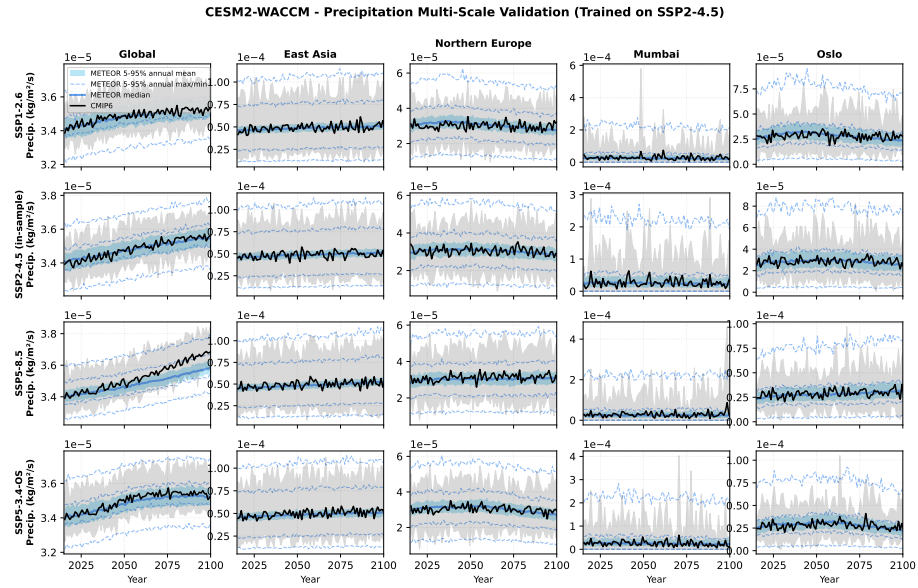


Figure S6: Out-of-sample precipitation validation for CESM2-WACCM trained on SSP2-4.5. Layout as in Figure 8.

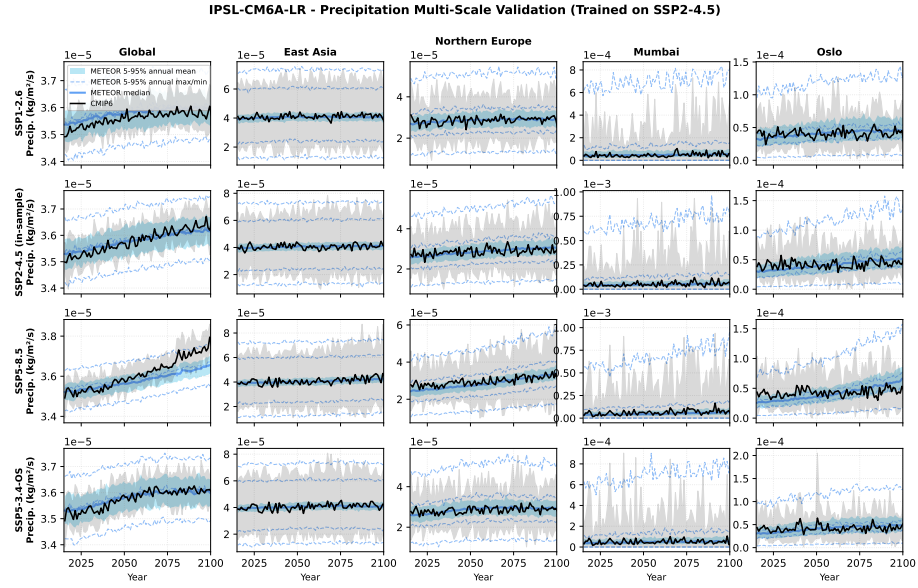


Figure S7: Out-of-sample precipitation validation for IPSL-CM6A-LR trained on SSP2-4.5. Layout as in Figure 8.

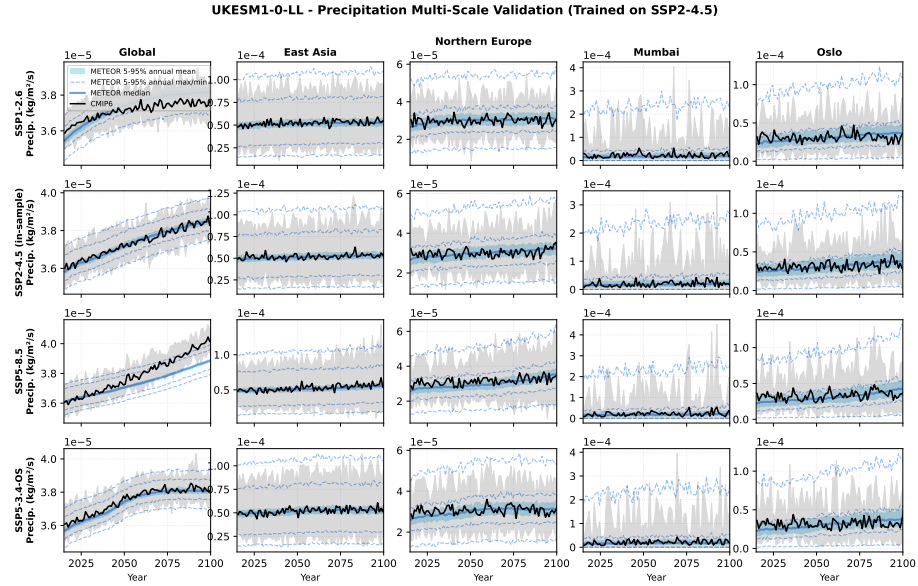


Figure S8: Out-of-sample precipitation validation for UKESM1-0-LL trained on SSP2-4.5. Layout as in Figure 8.

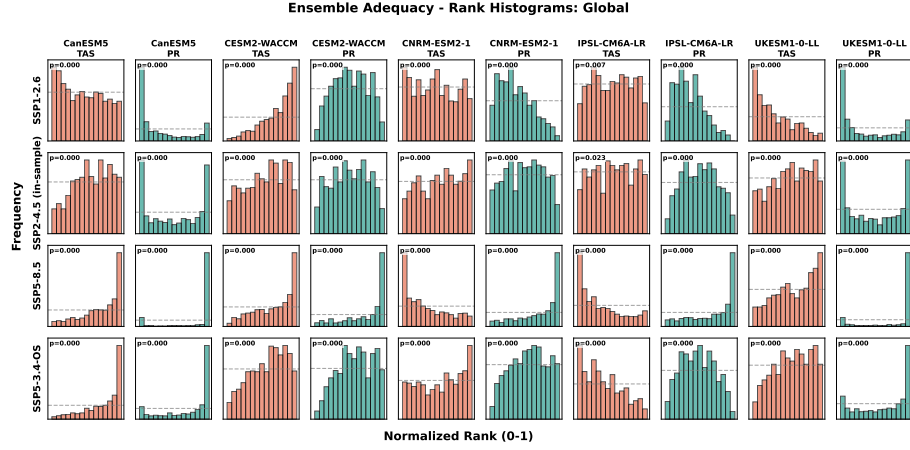


Figure S9: Out-of-sample rank histograms for global mean. Layout as in Figure 9. Note the more pronounced underdispersion in precipitation compared to point-scale results, suggesting the precipitation distributional transform may be less appropriate for large-scale spatial aggregations.

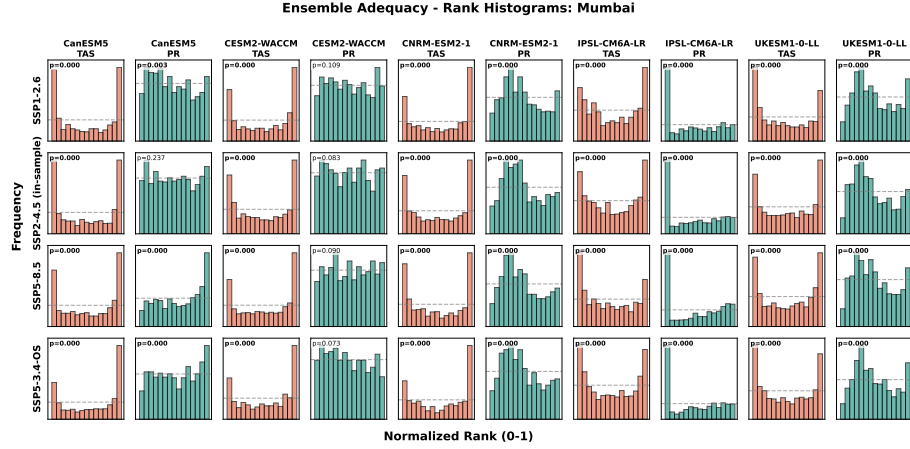


Figure S10: Out-of-sample rank histograms for Mumbai, India. Layout as in Figure 9.

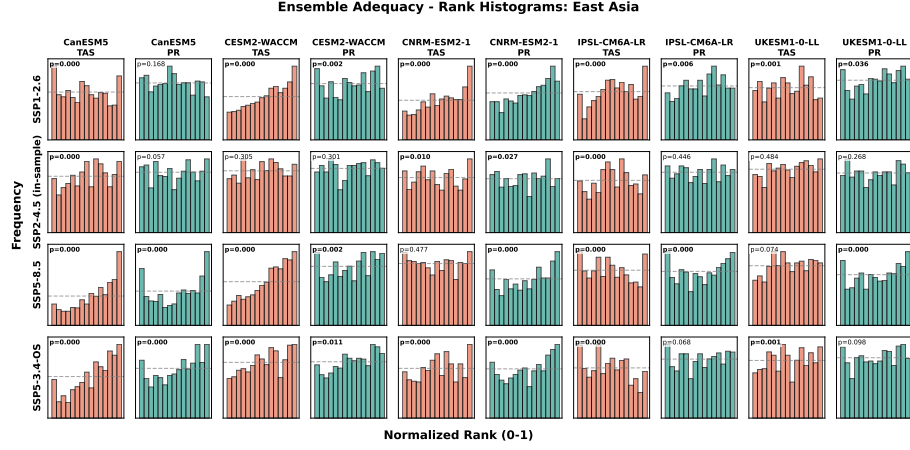


Figure S11: Out-of-sample rank histograms for East Asia (AR6 region). Layout as in Figure 9.

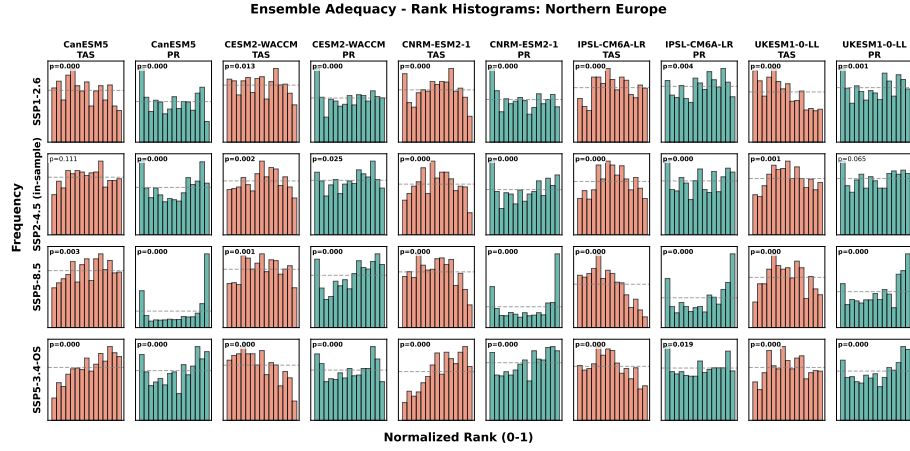


Figure S12: Out-of-sample rank histograms for Northern Europe (AR6 region). Layout as in Figure 9.

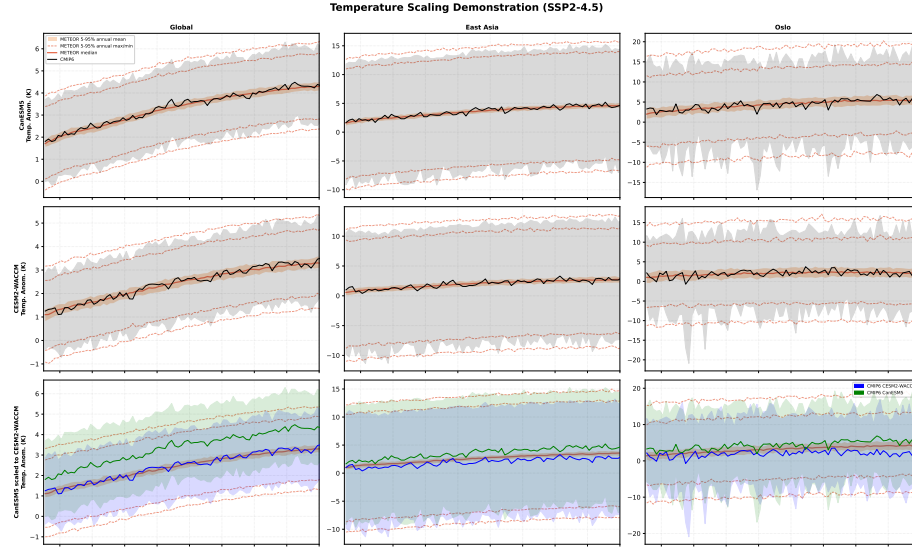


Figure S13: Temperature anomaly results for ssp245 with a model trained on CanESM5, but scaled to the global mean timeseries of CESM2-WACCM, for global, regional and point scale. Results for the models trained directly on the two models are shown for comparison in the top two rows.

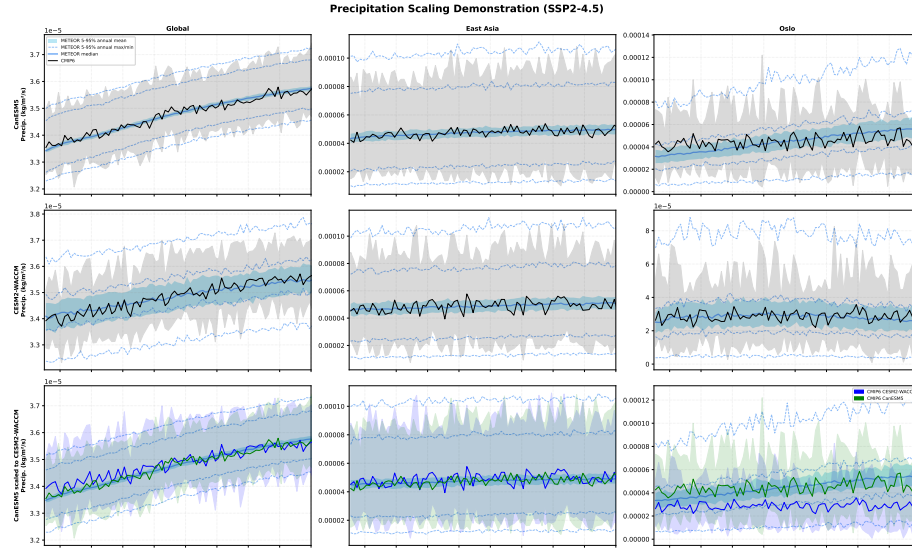


Figure S14: Precipitation anomaly results for ssp245 with a model trained on CanESM5, but scaled to the global mean timeseries of CESM2-WACCM, for global, regional and point scale. Results for the models trained directly on the two models are shown for comparison in the top two rows.

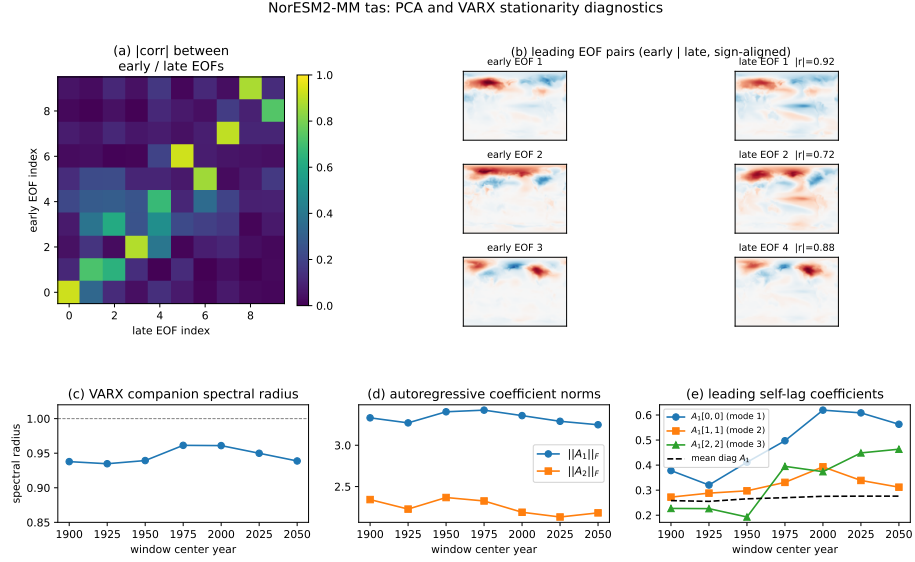


Figure S15: Stationarity diagnostics for NorESM2-MM temperature. (a) Absolute spatial pattern correlation between the leading 10 EOFs of a noise model refit on the early half (1850–1974) versus the late half (1976–2100) of the training record. (b) Top-three EOF pair maps (early | late, sign-aligned for display); the paired late EOF is chosen by best-match correlation. (c) Spectral radius of the VARX companion matrix as a function of window center year for fits on rolling 100-year windows. Values below 1 (dashed) indicate stationarity of the AR process. (d) Frobenius norms of the autoregressive coefficient matrices A_1, A_2 . (e) Individual self-lag coefficients $A_1[i, i]$ for the top-three modes and their diagonal mean.

NorESM2-MM pr: PCA and VARX stationarity diagnostics

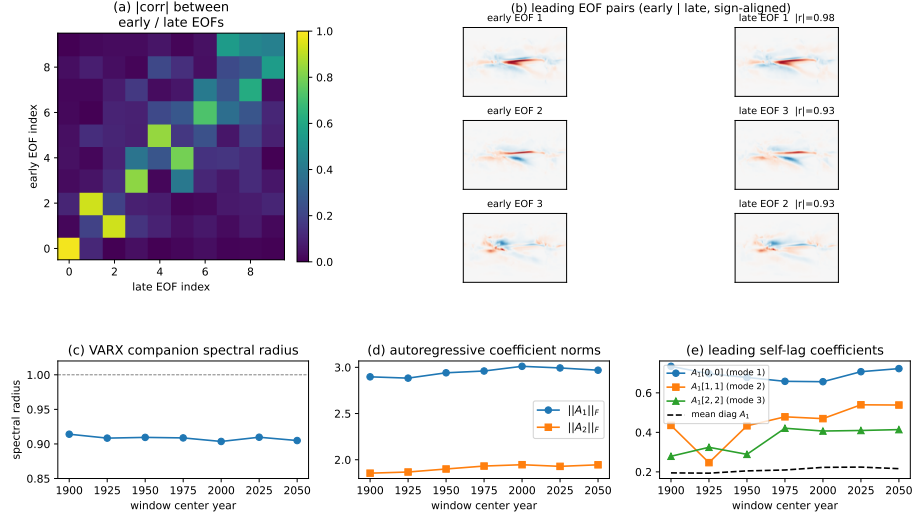


Figure S16: Stationarity diagnostics for NorESM2-MM precipitation. Layout as in Supplementary Figure S15.

NorESM2-MM tas: sensitivity to T_{glob} smoothing window

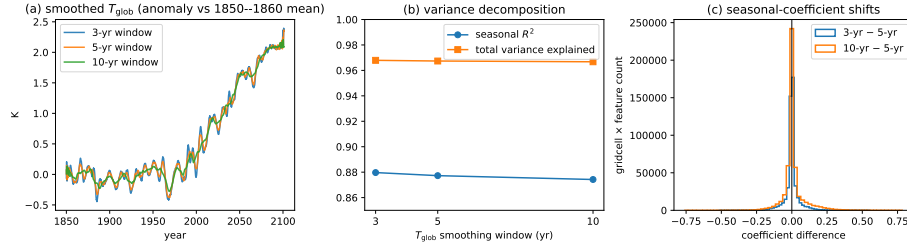


Figure S17: Sensitivity of the seasonal-cycle fit to the T_{glob} smoothing window. (a) Smoothed T_{glob} curves for 3-, 5- and 10-year windows, shown as anomalies relative to their respective 1850–1860 means; the three curves are virtually indistinguishable at multi-decadal timescales. (b) Seasonal-model R^2 and total variance explained (seasonal + PCA) as a function of window length. Both are essentially flat, changing by less than 0.6 and 0.1 percentage points respectively. (c) Distribution of per-gridcell seasonal-regression coefficient differences relative to the 5-year reference fit, pooled across all nine features. The distributions are tightly clustered around zero.

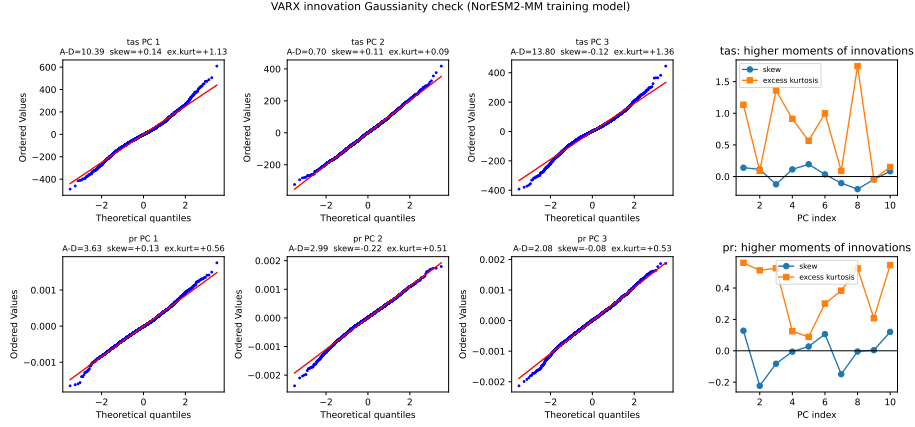


Figure S18: Empirical check of the Gaussian assumption on the VARX innovations $\varepsilon(t)$ for the NorESM2-MM training model. Columns 1–3 show quantile–quantile (QQ) plots, which order the sampled shocks and plot them against the values a Gaussian of the same variance would give, so points lying on the diagonal indicate Gaussian behaviour and systematic curvature away from it indicates a departure. Plots are shown for the top three PCs of temperature (top) and precipitation (bottom), each annotated with an Anderson-Darling statistic (a goodness-of-fit measure weighted towards the tails), the skewness, and the excess kurtosis (how much heavier the tails are than a Gaussian’s, zero for a perfect Gaussian). Column 4 shows the skewness and excess kurtosis across the leading ten PCs. Skew is small in absolute value across all leading modes for both variables; excess kurtosis is systematically positive at 0.5–1.4 for temperature and 0.1–0.6 for precipitation, indicating slightly heavier tails than a strict Gaussian, most visible in the outer $\pm 3\sigma$ regions of the QQ plots.

NorESM2-MM SSP2-4.5 pr: vectorized gamma MLE fallback fires at 0 of 55,296 gridpoints (0.00%)

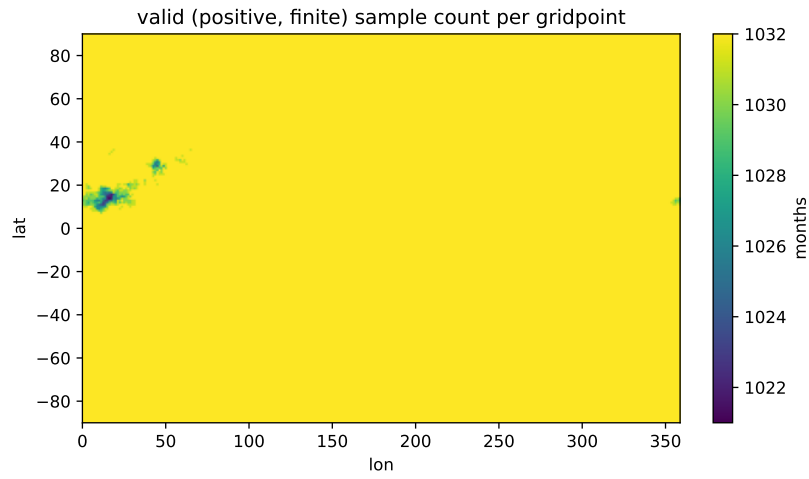


Figure S19: Frequency of the gamma maximum-likelihood fallback across the NorESM2-MM SSP2-4.5 precipitation grid (55,296 gridpoints, 1032 monthly samples each). The map shows the number of valid (positive, finite) monthly samples at each gridpoint, which is what determines whether the fit is possible. Every gridpoint retains at least 1021 of the 1032 samples; the only appreciable shortfall is a small cluster of arid cells over the Sahara and Arabian peninsula that record zero precipitation in roughly ten months per century, still far above the two positive samples the fit requires. The fallback consequently does not trigger in any gridpoint on this training record.