

Author Response to Reviewers

Climate emulation framework METEORv1.6: Monthly climate variability
and integrated impact assessment extensions

B. M. Sanderson et al.

September 8, 2026

We thank both reviewers for their careful and constructive reading of the manuscript. The comments have helped us sharpen the framing of METEOR-NOISE as a statistical construction distinct from the physically motivated METEOR-CORE, acknowledge the stationarity assumptions that sit inside the PCA and VARX steps, and add several quantitative diagnostics that strengthen the validation.

This document first summarises the changes made to the manuscript and supplement, then responds to each reviewer comment in turn. Reviewer text is set in *grey italic*; our responses follow. Section numbers refer to the revised manuscript. A marked-up version of the manuscript showing all changes is provided alongside this response.

1. Summary of changes to the manuscript

1.1 Distinction between the physical and statistical components

- C1. Section 2.2 (overview).** We added two sentences added stating the METEOR-CORE / METEOR-NOISE distinction outright, prompted by reviewer 2: METEOR-CORE’s impulse-response formulation has a physical motivation; METEOR-NOISE is a statistical construction whose PCA basis is variance-optimal rather than physically defined.
- C2. Section 2.2 (PCA description).** The claim that the patterns “include known modes of variability like ENSO or the North Atlantic Oscillation” has been softened to note that EOFs maximise explained variance rather than physical separability, and that the PCs are treated here as a compact statistical basis, not as physically labelled modes.
- C3. Section 3.4 (VARX motivation).** “Interactions between modes, e.g. tropical patterns influence extratropical circulation” reframed as “statistical dependencies between modes (which can, in principle, reflect physical teleconnections such as tropical–extratropical coupling)”.
- C4. Section 3.4 (teleconnection example).** The hypothetical tropical-Pacific / North-America example has been rewritten so that the physical reading is explicitly conditional on those processes projecting onto the leading EOFs of the training data.
- C5. Section 8.3 (comparison with ML approaches).** The attribution of “physically grounded” is now split between the two components: METEOR-CORE provides a physically motivated impulse-response formulation; METEOR-NOISE extends this with a statistical variability model.

C6. Section 9 (Conclusions, opening). Rewritten to attribute the physical motivation specifically to the impulse-response core rather than to the framework as a whole.

1.2 New diagnostics and analysis

C7. Section 8.2 — stationarity of the PCA basis and VARX coefficients. Two new paragraphs report a split-sample EOF comparison (early 1850–1974 vs. late 1976–2100) and a rolling 100-year-window refit of the VARX coefficients, with quantitative results for both temperature and precipitation.

C8. New supplementary section, “Stationarity of PCA and VARX components” (Supplemental Figures 15 and 16), giving the full diagnostic per variable: EOF pattern correlation matrices, top-three EOF pair maps, companion-matrix spectral radius, Frobenius norms of A_1 and A_2 , and individual self-lag coefficients across seven rolling windows.

C9. Section 3.1 — smoothing-window sensitivity. A sentence added reporting that refitting with 3- and 10-year T_{glob} smoothing windows changes the seasonal R^2 by less than 0.6 percentage points and the total variance explained by less than 0.1 percentage points, with the full analysis in Supplemental Figure 17.

C10. Section 3.2 — scope and provenance of the harmonic regression. A new paragraph states up front that the seasonal-cycle model applies to both temperature and precipitation (and in principle any other supported variable), and cites prior harmonic representations in monthly-resolution emulation (Nath, 2022; Schöngart, 2024), noting that our extension is the GMT-modulated amplitude and phase terms.

C11. Section 3.3 — MTM-SVD. A new paragraph acknowledges Multi-Taper Method Singular Value Decomposition (Mann, 1994) as a frequency-domain alternative and explains why a time-domain PCA+VARX approach was preferred given the generative requirement, flagging a frequency-domain comparison as future work.

C12. Figures 5 and 9 — formal uniformity tests. Every panel of the in-sample and out-of-sample rank histograms is now annotated with a chi-squared uniformity p -value (21-bin, $\text{df} = 20$ and 15-bin, $\text{df} = 14$ respectively). Panels with $p < 0.05$ are labelled in bold, and both captions state the convention.

C13. Section 8.2 and new Supplemental Figure 18 — Gaussianity of the VARX innovations. QQ plots and Anderson–Darling / higher-moment diagnostics for the leading PCs of both variables, reported in the text as small skew ($|\text{skew}| \leq 0.25$) with systematic positive excess kurtosis (0.5–1.4 for temperature, 0.1–0.6 for precipitation). Alternatives (t-distributed innovations, GARCH-style time-varying variance, empirical resampling of residuals) are now cited as future directions.

C14. Section 4.5 and new Supplemental Figure 19 — gamma MLE fallback frequency. We instrumented `_vectorized_gamma_mle` over the full NorESM2-MM SSP2-4.5 record and now report that the fallback path does not fire at any gridpoint. The new supplemental figure shows the valid-sample count per gridpoint, which is what governs whether the fit is possible. The description of the fallback in the text has been softened to reflect that it is a safety net rather than a routinely exercised path.

1.3 Corrections and editorial changes

- C15. Equation 4.** The fifth component is corrected from $\sin(2\pi t/12)$ to $\sin(4\pi t/12)$. This was a typesetting error only; the code builds both the annual and semi-annual sine/cosine terms correctly, and no results are affected.
- C16. Figure 1.** The “DegreeDaysCalulator” typo in the impacts box has been corrected to “DegreeDaysCalculator”.
- C17. Data availability.** The section now states that all input data are CMIP6 output available through ESGF and the PANGEO mirrors and identifies the fields used (`tas` and `pr` at monthly resolution). The model output is formally cited in the reference list at the granularity of the twelve ESGF model–activity collections drawn on, and a new table gives the citation DOI of each of the 39 individual simulations used, across the six ESMs and their `abrupt-4xC02`, `historical`, `piControl` and SSP experiments; every entry in the table hyperlinks to its DOI. All DOIs were retrieved from, and checked against, the WDCC/DKRZ CMIP6 Citation Service. This is in addition to the existing code DOI.
- C18. Supplement — introductory material.** A new opening paragraph explains what the supplement contains and how the variance decomposition in Table 1 separates the seasonal and PCA stages of the model.
- C19. Supplement — caption correction.** The caption of the precipitation scaling demonstration figure incorrectly read “Temperature anomaly results”; this is fixed.
- C20. Sections 4.3 and 6.1.2 — GMT scaling.** The description of the GMT-scaling procedure has been rewritten and cross-referenced explicitly to the METEOR-CORE integration section and to the ensemble generation equation. A malformed cross-reference at the end of Section 4.3 has also been repaired.
- C21. Acknowledgements.** The statement on AI tool use has been revised to describe their role more precisely (initial code development and initial literature review, with all text and code subsequently reviewed, edited and validated by the authors).
- C22. Throughout.** Minor clarifications and wording edits: brief glosses added for “multicollinearity” and for the VARX innovation term, em-dash constructions replaced, and several sentences in Sections 8 and 9 tightened.
-

2. Response to Reviewer 1

Many thanks to the reviewer for taking the time to read the paper. The reviewer’s original comments follow in italics together with our responses.

This is an exceptional manuscript that was an intellectual joy to read. The authors extend the already published METEOR model to incorporate monthly variability and to the impact space. They evaluate/validate the METEORv1.6 in-sample and out-of-sample. I recommend prompt publication.

Response. Many thanks indeed for the positive response.

2.1 Equation 4

Eq 4: is the 5th entry supposed to be $4\pi/12$ or $2\pi/12$?

Response. Thanks for picking this up — it’s an issue in the L^AT_EX only. The code builds both the annual and semi-annual sin/cos terms correctly. Equation 4 is corrected, and no results are affected.

2.2 Data availability

Data: I briefly checked the zenodo repo and I think you only have code there, right? I think you should release the underlying data of the work as well.

Response. Thanks — all data is public and can be automatically fetched by the code from ESGF or PANGEO mirrors. We’ve now revised to include the DOI for every CMIP simulation and field used in the analysis. The model output is cited formally in the reference list at the granularity of the ESGF model–activity collections, and a new table in the Data and Code Availability section lists the DOI of each of the 39 individual simulations, each hyperlinked to its landing page.

2.3 Consistency and typographical issues

L586: I would prompt Claude to double check some consistency issues and phrasings throughout the manuscript ;) e.g., “DegreeDaysCalulator” in Fig 1

Response. This figure was hand-made ;) . But thanks for catching! Typo fixed, and we have carefully checked for others before resubmission.

3. Response to Reviewer 2

We thank the reviewer for a careful and constructive assessment of the manuscript. We respond to each comment below.

3.1 General comments

3.1.1 Physical grounding vs. statistical construction

Related concern is the repeated characterization of the framework as “physically grounded” (e.g., Section 8.3, Discussions). While METEOR-CORE’s impulse response formulation has a clear physical motivation ..., the METEOR-NOISE component is fundamentally a statistical construction. PCA/EOF decomposition maximizes variance, not physical interpretability...

Response. We agree with the distinction, and we’ve gone through the manuscript to keep the physical motivation of METEOR-CORE (an impulse response to forcing) separate from the statistical character of METEOR-NOISE (a variance-optimal EOF basis plus a VARX model). The updated framing shows up in several places.

1. **§2.2, end of overview paragraph.** Added two sentences that state the METEOR-CORE / METEOR-NOISE distinction outright: *“METEOR-CORE’s impulse-response formulation has a physical motivation. METEOR-NOISE is a statistical construction: the PCA basis is variance-optimal rather than physically defined, and its purpose is to reproduce the target ESM’s covariance and persistence structure, not to represent physical processes.”*
2. **§2.2, PCA description** (previously: “These patterns include known modes of variability like ENSO or the North Atlantic Oscillation.”). Softened to flag that the EOFs are variance-optimal, not physical. The revised passage reads: *“The leading patterns often project onto recognizable modes of variability such as ENSO or the North Atlantic Oscillation. But EOFs maximize explained variance rather than physical separability, and do not necessarily isolate individual physical processes without dedicated validation (e.g. correlation with standard climate indices or comparison with rotated EOF solutions). Here we treat the PCs as a compact statistical basis for reproducing spatiotemporal covariance, not as physically labelled modes.”*
3. **§3.4, VARX motivation** (previously: “interactions between modes, e.g. tropical patterns influence extratropical circulation”). Reframed as *“statistical dependencies between modes (which can, in principle, reflect physical teleconnections such as tropical–extratropical coupling)”*.
4. **§3.4, hypothetical tropical-Pacific / North-America example** (was: “This allows for richer teleconnections: if tropical Pacific warming represented in one PC typically precedes North American circulation changes...”). Rewritten to make the physical reading conditional. The revised text reads: *“Where such dependencies reflect physical teleconnections (for example, tropical Pacific SST anomalies preceding extratropical circulation responses by roughly one month), they are captured to the extent that these processes project onto the leading EOFs of the training data.”*
5. **§8.3, comparison with ML approaches** (was: “METEOR provides physically grounded and fully explainable pattern scaling...”). The attribution is now split. As revised: *“METEOR-CORE provides a physically motivated impulse-response formulation with interpretable structure and significantly lower training data requirements; METEOR-NOISE extends this with a statistical variability model calibrated to reproduce the target ESM’s covariance and persistence.”*
6. **§9, Conclusions opening** (was: “lightweight enough for interactive use, yet grounded in the physical dynamics that govern Earth system model responses”). Rewritten to attribute the physical motivation to the impulse-response core. The revised opening reads: *“METEORv1.6 is lightweight enough for interactive use, with a physically motivated core (impulse-response pattern scaling) and a statistical model of sub-annual variability calibrated against ESM output.”*

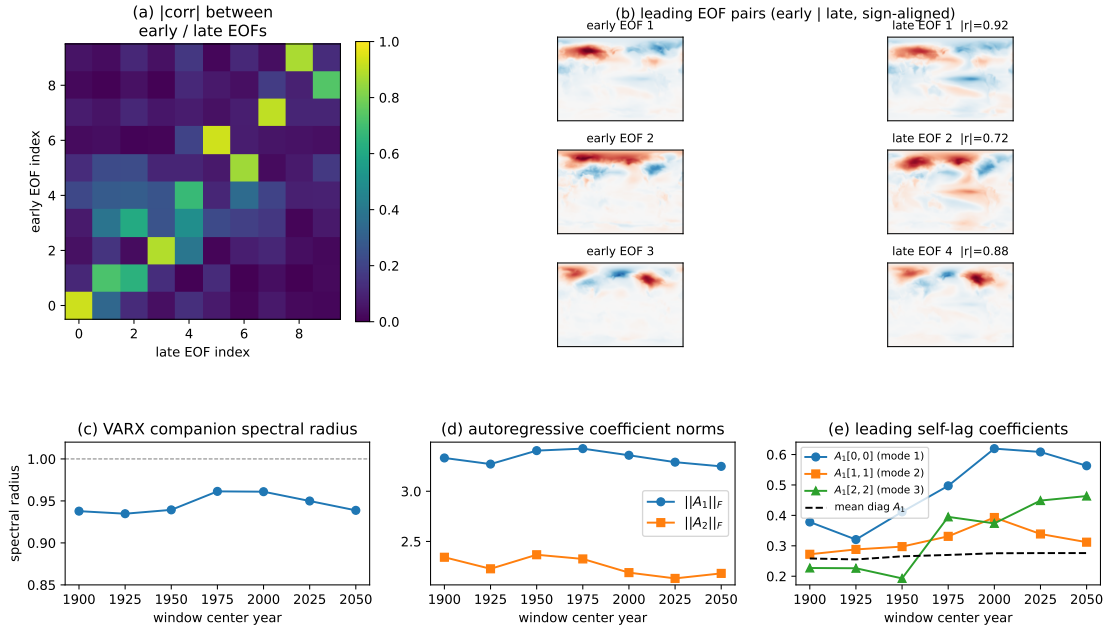
3.1.2 PCA stationarity

...the PCA decomposition is performed over the entire 250-year training record ... yielding a fixed set of Empirical Orthogonal Functions (EOFs) that are then assumed to remain valid across all climate states. This stationarity assumption is a strong hypothesis that is not discussed in the manuscript.

Response. The concern is fair, and it deserves treatment alongside the VARX stationarity discussion. The worry is concrete: we compute the EOF basis once, from the whole record, and

then hold it fixed in every climate state we simulate. If the dominant patterns of variability reorganise as the planet warms, that fixed basis stops being the right one. We’ve done two things. First, expanded Discussion §8.2 with two new paragraphs on the stationarity of both the PCA basis and the VARX coefficients. Second, added a supplementary section with the full diagnostic and a combined figure per variable (Supplemental Figures 15 and 16). The §8.2 text now reports the split-sample result directly: *“The leading EOFs are broadly preserved across the split-sample halves (top-10 best-match $|r| = 0.81$ for temperature, 0.77 for precipitation), with the first mode particularly stable ($|r| = 0.92$ and 0.98 respectively) but some reordering at intermediate ranks.”*

NorESM2-MM tas: PCA and VARX stationarity diagnostics



Supplemental Figure 15 asks whether the noise model’s basis and lag structure drift across the training record. We refit on the early half (1850–1974) and the late half (1976–2100), and on rolling 100-year windows. (a) Absolute pattern correlation between the top 10 EOFs of the early and late fits. A bright diagonal means the same patterns show up in the same rank; bright off-diagonal means they persist but get re-ordered. (b) The top three EOFs side by side (early | late). The modes are near-identical, so the same spatial structures dominate both halves. (c) Spectral radius of the VAR companion matrix at each 100-year window centre — a single number for how persistent the fitted process is, where below 1 means shocks decay rather than grow. It stays in the same range across the record, so the process is stable, and stable in the same way throughout. (d) The overall magnitude of the autoregressive coefficient matrices is essentially flat, so the temporal patterns don’t strengthen or fade in global intensity. (e) Self-persistence of the top three individual modes plus their average (dashed). The average is flat, but the leading mode’s self-lag — how strongly it predicts its own value the following month, the $A_1[0,0]$ term — rises from 0.32 to 0.62 across the record. Either its persistence is genuinely lengthening under warming, or the leading EOF is progressively absorbing more of the forced trend, which is highly autocorrelated by construction. The diagnostic can’t separate the two, and we say so.

In the revised text we state that the assumption holds well for the leading modes, which carry most of the variance, but isn’t uniformly satisfied at intermediate ranks — consistent with the reviewer’s concern.

3.2 Specific comments

3.2.1 Physical interpretation of PCA modes

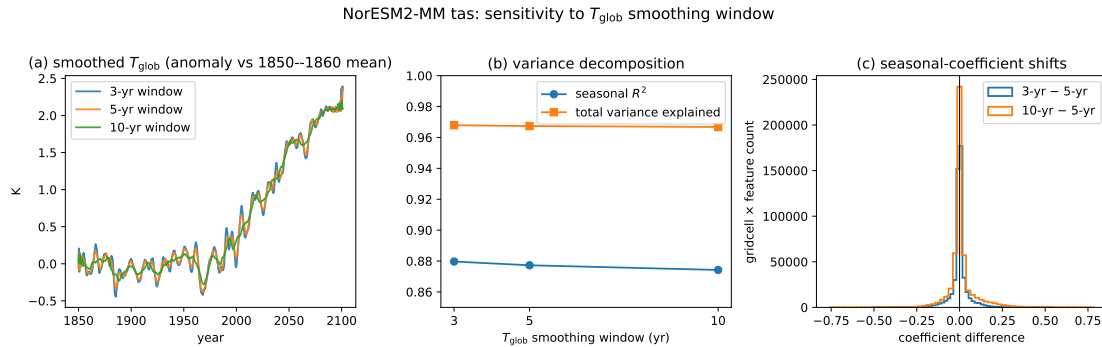
The text states for example that PCA patterns “include known modes of variability like ENSO or the North Atlantic Oscillation” (line 131) and that “tropical patterns influence extratropical circulation” (line 204)...

Response. Addressed through the framing revisions above. The two sentences the reviewer flags — the ENSO/NAO line and the tropical–extratropical line — are exactly the ones quoted under General Comment 1, items 2 and 3. Both now state that the EOFs are variance-optimal and carry no guarantee of physical separability.

3.2.2 5-year smoothing window

The choice of a 5-year running mean for smoothing the global mean temperature is stated without justification ... it should be demonstrated with a sensitivity analysis showing that results are robust to this choice (e.g., testing 3-year, 5-year, and 10-year windows).

Response. Sensitivity analysis added. We refit the noise model on NorESM2-MM SSP2-4.5 with T_{glob} smoothing windows of 3, 5, and 10 years, and compared the seasonal-cycle fit quality and the coefficient distributions. The seasonal R^2 moves by less than 0.6 percentage points across the three windows (0.880, 0.877, 0.874), and the total variance explained by the seasonal + PCA model is stable to within 0.1 percentage points (0.968, 0.967, 0.967). The seasonal-coefficient distributions sit almost on top of each other (Supplemental Figure 17, panel c). So the 5-year value isn’t a critical tuning point; we keep it as a sensible middle ground between suppressing interannual noise, which would otherwise contaminate the harmonic regression, and preserving the decadal modulation of the forcing. The added justification in §3.1 reads: “This choice is not critical: refitting the noise model with 3- and 10-year windows changes the seasonal R^2 by less than 0.6 percentage points and the total variance explained by less than 0.1 percentage points.”



Supplemental Figure 17 asks whether the noise model is sensitive to whether we smooth global temperature over 3, 5, or 10 years before fitting the seasonal cycle. (a) The three smoothed T_{glob} curves, plotted as anomalies from each curve’s own 1850–1860 mean. At multi-decadal timescales they’re indistinguishable — the window only affects the year-to-year wiggle that the seasonal fit doesn’t rely on. (b) Seasonal-model R^2 and total variance explained (seasonal + PCA) as a function of window length. Both are essentially flat, moving by less than 0.6 and 0.1 percentage points across the range. (c) Distribution of per-gridcell seasonal-regression coefficient differences relative to the 5-year fit, pooled across all nine harmonic features. Both the 3-yr and 10-yr fits cluster tightly around zero — the coefficients barely move.

The takeaway is that the 5-year window isn’t a critical assumption. Anything from 3 to 10 years gives essentially the same seasonal fit.

3.2.3 Scope of harmonic regression

The feature vector X_{harm} is introduced for the seasonal cycle model, but it is not made clear early enough that this regression applies to both temperature and precipitation (and potentially other variables). Moreover, this formalism is not new, if you find cite few article thant do it.

Response. Both are good points. Two changes in §3.2:

- A short new paragraph, right after the motivating sentences, states that the seasonal-cycle model applies to both temperature and precipitation (and in principle any other supported variable), with a pointer that the variable-specific distributional treatment comes later.
- The same paragraph cites prior work in monthly-resolution climate emulation that uses harmonic representations of the seasonal cycle (Nath, 2022; Schöngart, 2024), and notes that the harmonic representation itself isn’t new — our contribution is the GMT-modulated amplitude and phase terms.

The added §3.2 text reads: *“The model described below applies to both temperature and precipitation, and in principle to any other supported variable... Harmonic representations of the seasonal cycle are well established in monthly-resolution climate emulation (Nath, 2022; Schöngart, 2024); the formulation here extends this by allowing the amplitude and phase of the harmonics to modulate linearly with global mean temperature.”*

3.2.4 Section 3.3: MTM-SVD alternative

The paper does not discuss Multi-Taper Method Singular Value Decomposition (MTM-SVD), which provides a frequency-domain approach to identifying spatiotemporal variability modes that can naturally handle non-stationary spectral characteristics.

Response. We’ve added a paragraph at the end of §3.3 that acknowledges MTM-SVD (Mann, 1994) as a frequency-domain alternative and explains our choice. Our requirement here is generative: time-domain PCA+VARX simulates pattern amplitudes forward with GMT as an exogenous predictor, whereas frequency-domain modes need an extra resynthesis step to produce ensembles under an arbitrary temperature trajectory. MTM-SVD is still a useful independent diagnostic, and we flag a frequency-domain comparison as possible future work. The added paragraph opens: *“An alternative to this time-domain approach is a frequency-domain decomposition such as Multi-Taper Method Singular Value Decomposition (MTM-SVD) (Mann, 1994), which identifies spatiotemporal modes across a spectrum of frequencies and naturally accommodates non-stationary spectral characteristics.”*

3.2.5 Rank histogram uniformity tests

The rank histogram evaluation is appreciated, but the interpretation remains largely qualitative (“broadly uniform”, “slight underdispersion”). For a more rigorous assessment, the authors should provide a formal uniformity test with associated p -values.

Response. Agreed. We now annotate every panel of the in-sample rank histogram (Figure 5) and the out-of-sample one (Figure 9) with a chi-squared uniformity p -value, computed on the 21-bin ($df = 20$) and 15-bin ($df = 14$) count vectors respectively. A flat histogram is

what a well-calibrated ensemble should produce, so the test measures how unlikely the observed departure from flat is under that null. Panels where $p < 0.05$ — a formal rejection of uniformity — are labelled in bold, and the captions state the convention. The Figure 5 caption now reads: “Each panel is annotated with the chi-squared uniformity p -value on the 21-bin count vector ($df = 20$); values shown in bold indicate $p < 0.05$, i.e. formal rejection of the uniformity null.”

3.2.6 Gaussianity of VARX innovations

The assumption that the errors follow a normal distribution is imposed by construction but not verified empirically. The authors should acknowledge that this constitutes a strong modelling hypothesis. ... alternative approaches exist and should at least be cited.

Response. Empirical check added as Supplemental Figure 18. The assumption under test is that the VARX innovations — the month-to-month random shocks that drive the noise model — are Gaussian. We check it two ways: QQ plots for the top three PCs, which compare the shape of the empirical distribution against a normal one, and Anderson–Darling and higher-moment diagnostics across the top ten. The picture is mostly, but not entirely, Gaussian.

- Asymmetry is small. $|\text{skew}| \leq 0.25$ for both tas and pr across all the leading modes, so the distributions aren’t visibly lopsided.
- The tails are a little heavy. Excess kurtosis is positive throughout — 0.5–1.4 for the leading tas PCs and 0.1–0.6 for pr — so there’s slightly more weight out in the extremes than a Gaussian puts there.
- The formal test rejects, as it was always going to at this sample size. Anderson–Darling rejects the strict Gaussian null at $n \approx 3000$ for most PCs, but the QQ plots only peel away from the reference line out in the $\pm 3\sigma$ tails.

We acknowledge this in §8.2 and cite alternatives (t-distributed innovations, GARCH-style time-varying variance, or empirical/non-parametric resampling of residuals) as directions for future work. We keep the Gaussian assumption for now because it’s what makes the multivariate covariance Σ analytically tractable. It’s fit for purpose for monthly-mean fields, but it underestimates the probability of extremes — consistent with the caveat already in the paper. The §8.2 text now reports: “Anderson-Darling tests and quantile-quantile (QQ) diagnostics on the fitted VARX innovations show small skew ($|\text{skew}| \leq 0.25$ across leading modes) but systematic positive excess kurtosis of 0.5–1.4 for temperature and 0.1–0.6 for precipitation, indicating slightly heavier tails than a strict Gaussian.”

3.2.7 Gamma distribution MLE fallback

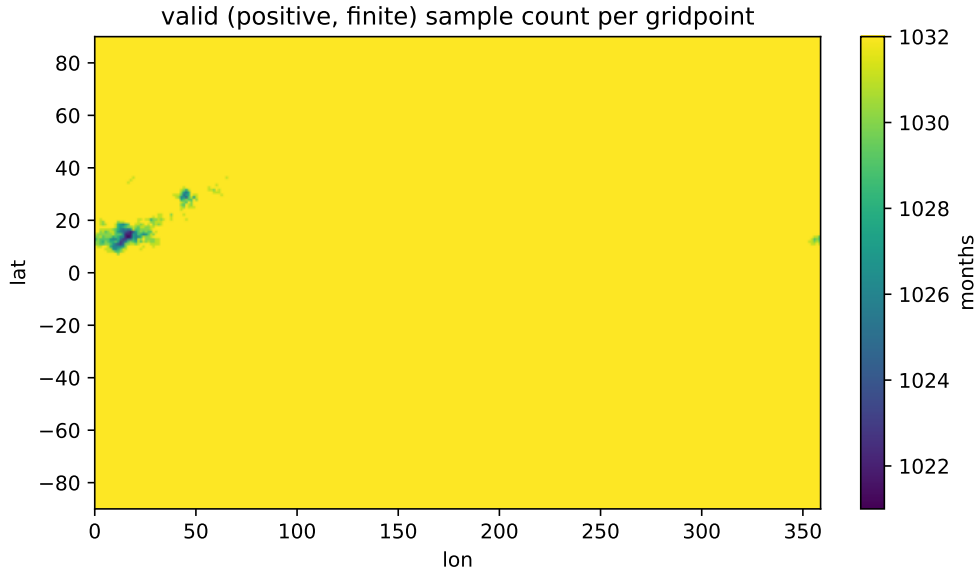
...the authors should report how frequently the MLE fallback is triggered across their grid and whether this correlates with regions of poor precipitation emulation.

Response. We instrumented `_vectorized_gamma_mle` — the routine that fits a gamma distribution to gridded precipitation, described in §4.5 — and ran it over the full NorESM2-MM SSP2-4.5 record (55,296 gridpoints, 1032 monthly samples each). The fallback fired at zero gridpoints: 0.0000% (Supplemental Figure 19). The degenerate path the paper calls “method-of-moments” (`shape = 1`, `scale = mean`) never actually triggers on this training data.

The figure shows why. Every gridpoint has at least 1021 valid (positive, finite) samples out of 1032. The worst case is a small cluster of Saharan cells that record zero precipitation in about 10

months per century — still well above the $n \geq 2$ the MLE needs to fit. The vectorised Newton iteration converges everywhere.

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



Supplemental Figure 19 reports how often the gamma maximum-likelihood fit fails over the NorESM2-MM SSP2-4.5 precipitation grid. The map shows the number of valid (positive, finite) monthly samples at each gridpoint, out of 1032, since that is what determines whether the fit is possible. The only appreciable shortfall is the arid cluster over the Sahara and Arabian peninsula, and even there at least 1021 samples remain — far above the two the fit needs. The fallback therefore fires nowhere, so the companion map of fallback gridpoints would be blank and is not shown.

We’ve added this fraction and its spatial pattern to §4.5, and softened the paper’s description of the fallback to what it is — a safety net. §4.5 now states: “*When fitting fails (e.g., at gridpoints with fewer than two positive samples), the code falls back to a canonical setting equivalent to method-of-moments estimation... In practice this fallback path does not fire on the NorESM2-MM SSP2-4.5 training grid used in this paper.*”

3.2.8 VARX coefficient stationarity

The VARX model (Eq. 8) assumes that the coefficient matrices A_1 , A_2 and B are constant over the full training period ... a simple diagnostic such as fitting the VARX on rolling sub-periods and examining coefficient stability would considerably strengthen confidence in the approach.

Response. We’ve done this. We refit the noise model on overlapping 100-year windows that slide through the NorESM2-MM SSP2-4.5 record in 25-year steps (centres 1900 to 2050), and asked a simple question of each window: do the fitted dynamics stay put? Three quantities answer it. The spectral radius of the companion matrix says whether the process is stationary at all: a value below 1 means shocks decay, above 1 means they amplify. The overall size of the lag matrices (their Frobenius norm) says whether the month-to-month coupling strengthens or weakens across the record. And the individual self-lags say how persistent each mode is on its own.

On the first two, nothing really moves. The spectral radius stays comfortably below 1 in every window, for both variables, so the process never drifts toward instability, and the lag matrices

change by under 10% end to end. The exogenous-forcing term looks more variable, but that just tracks the changing variance of the temperature input we feed it; it isn't drift in the model.

The one quantity that genuinely moves is the leading temperature mode's self-lag — how strongly it predicts its own value the following month — which roughly doubles over the record, from 0.32 to 0.62. This is the interpretively interesting case, and it's genuinely ambiguous. Either that mode's physical persistence is lengthening under warming, or the leading EOF is progressively absorbing more of the forced trend (which is highly autocorrelated by construction) and inheriting its persistence. The rolling-window diagnostic can't separate the two, so we flag it in the supplement and don't read a mechanism into it. Precipitation shows none of this: every one of these diagnostics is flatter still.

The §8.2 summary in the paper reports the rounded figures directly: *“The spectral radius of the companion matrix... stays within 0.94–0.96 for temperature and 0.90–0.91 for precipitation across all seven rolling windows... The self-lag of the leading temperature mode ($A_1[0,0]$...) rises from 0.32 in the earliest window to 0.62 in the modern period.”*

The combined supplemental figures (Supplemental Figures 15 and 16) show both the PCA and VARX diagnostics per variable. The full analysis lives in `notebooks/notebooks_v16_description_paper/Stat`

3.3 Technical comments

3.3.1 Eq. 4 typo

Eq. 4: the fifth component is written as $\sin(2\pi t/12)$, which duplicates the third component. This appears to be a typo, it should presumably be $\sin(4\pi t/12)$.

Response. Confirmed - LaTeX typo only. The code builds both the annual and semi-annual sin/cos terms correctly (see `noise_generator.py`, lines 160–163). Equation 4 is corrected, and no results are affected.