General comment

This paper presents METEORv1.6, an extension of the METEOR climate emulation framework that adds monthly variability generation (METEOR-NOISE). The paper is clearly written, well-structured, and addresses a genuine need in the climate modeling community: bridging the gap between computationally expensive Earth System Model simulations and the sectoral demand for sub-annual, spatially resolved climate projections with uncertainty quantification.

The overall methodology is robust and the validation effort (which focuses on out-of-sample generalization across five CMIP6 models and multiple scenarios, rank histograms, and skill ratios) is impressive. The paper demonstrates that METEORv1.6 can reproduce key statistical properties of monthly climate variability at multiple spatial scales with modest computational cost. The method is innovative, particularly in its use of harmonics and in the way the residual is decomposed.

However, a key concern emerges from this methodology: the PCA decomposition is performed over the entire 250-year training record (historical and SSP2-4.5 or other scenario), 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. There is substantial evidence in the literature that key modes of climate variability (e.g., ENSO spatial structure, jet-stream position, Arctic amplification patterns) can shift under strong forcing. Even if the training phase trend to get the best compromise for stationnary PCA than represent well past and future, but it could give an high error in the emulation. This limitations should be discussed at least in the discussion part (only the hypothesys on staionnary on VARX is mentionned but not on PCA). A 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 (even though the corresponding spatial modes may conflate several physical processes and thus become partly abstract), the METEOR-NOISE component is fundamentally a statistical construction. PCA/EOF decomposition maximizes variance, not physical interpretability. The resulting modes are basis-dependent and their identification with physical processes (teleconnections, circulation patterns) is not guaranteed without dedicated analysis (e.g., rotated EOFs, physical validation). The manuscript should be more careful in distinguishing the physical grounding of the pattern scaling component from the statistical nature of the variability model. If the emulator is to be used for physical interpretation, a validation step linking the PCA modes to their putative physical counterparts should be performed before drawing any such conclusions. Of course, compare to an machine learning, the link between a physical mode and the PCA is possible.

Specific comments

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). These claims conflate statistical constructs with physical processes. As say previously, EOF modes maximize explained variance and their spatial structure depends on the domain, variable, and period used, they do not necessarily correspond to physically distinct modes. The authors should either remove such identifications or make a disclaimer that you should verify the identification before to do this use (e.g., correlation with standard climate indices, comparison with rotated EOF solutions).

5-year smoothing window (line 160): The choice of a 5-year running mean for smoothing the global mean temperature is stated without justification. This window seems as a good compromise between too short with the window leaves interannual variability in T_glob that contaminates the harmonic regression and too long with the window suppresses genuine decadal modulation of seasonality. However, it should be demonstrate with a sensitivity analysis showing that results are robust to this choice (e.g., testing 3-year, 5-year, and 10-year windows) or cite an established rationale for this particular value.

Scope of harmonic regression (around line 165): 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.

Section 3.3 : 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. Given that METEOR-NOISE aims to capture spatiotemporal climate variability including its frequency structure (seasonal, interannual), a brief discussion of why a time-domain PCA+VARX approach was preferred over frequency-domain alternatives would strengthen the methodological positioning.

Rank histograms (Figure 5): 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. Visual inspection of Figure 5 suggests non-negligible underdispersion in several cases (Northern Europe precipitation, Beijing precipitation, North South America or Delhi temperature), which would benefit from quantification.

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. Moreover, this assumption necessitates transforming precipitation values through a beta distribution to approximate Gaussianity. While the Gaussian framework is understandably retained for its mathematical convenience in constructing the covariance structure, alternative approaches exist and should at least be cited.

Gamma distribution fitting and fallback to method of moments: The paper states that when maximum likelihood estimation fails for the gamma distribution parameters, the code falls back to method-of-moments estimation ($k = \mu^2/\sigma^2$, $\theta = \sigma^2/\mu$). Method-of-moments estimators for the gamma shape parameter are known to be inefficient and can be unreliable, particularly when the shape parameter leads to a strongly skewed distribution. The authors should consider more robust alternatives, such as Bayesian estimation with weakly informative priors on the shape parameter, which can regularize the estimation without the instability of MLE or the inefficiency of moments. At minimum, the authors should report how frequently the MLE fallback is triggered across their grid and whether this correlates with regions of poor precipitation emulation.

VARX model and stationarity of lag structure: The VARX model (Eq. 8) assumes that the coefficient matrices A_1, A_2 and B are constant over the full training period. This implies that teleconnection strengths, persistence timescales, and the relationship between PCs and external forcing do not change over a period spanning from pre-industrial to late 21st-century conditions. Under very strong forcing (SSP5-8.5), this assumption is more likely to break down. While the Discussion briefly mentions this (lines 515–518), a simple diagnostic such as fitting the VARX on rolling sub-periods and examining coefficient stability would considerably strengthen confidence in the approach.

Technical comments

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 $\sin(4\pi t/12)$.