Response to Reviewer #2 (Bernard Twaróg):

> ***The presented study has many strengths. First and foremost, it introduces an innovative approach that is rarely seen in the literature and aligns with established methodologies for assessing climate variability. The use of Shannon entropy to analyze climate variability based on trajectories in phase space is a novel method that goes beyond traditional metrics. Constructing a phase space using the first three principal components (PCs) of SST and precipitation, derived from PCA, is consistent with physically justified modes of variability (AEM, AMM, SASD). The comparative value of the study is enhanced by the use of four different models (EC-Earth, GISS, iCESM, CCSM-Toronto) and multiple scenarios (PI, MHPMIP, MHGS, etc.). Analyzing SST and precipitation separately enables the identification of potential decoupling in their response to different forcings. The application of a bootstrap approach to estimate confidence intervals for entropy is methodologically sound.***

R2.1 ***However, the study is not without flaws. One contradiction lies in the implicit assumption that high entropy equates to high physical variability. Shannon entropy measures the diversity of system states, but not necessarily the amplitude of fluctuations. A simulation with low-amplitude variability but frequent state changes may yield high entropy, despite low physical variability.***

We appreciate the reviewer's comments. Indeed, Shannon entropy does not account for the amplitude of variability. As in standard practice, principal components obtained from PCA are normalized by their standard deviation, removing amplitude information and providing a common basis for comparing variability across simulations. Our approach follows this same principle by constructing a phase space from three normalized principal components, such that each dimension contributes equally to the entropy calculation and variability is characterized independently of amplitude. This is a deliberate methodological choice. While the amplitude and intensity of physical variables such as precipitation or sea surface temperature are important aspects of climate variability, the correct representation, persistence, and structure of the underlying modes of ocean–atmosphere variability are more subtle yet equally critical.

Following the reviewer's comment, we revised the manuscript to clarify that large EOF variance reflects high-amplitude variability, whereas high entropy indicates increased variability in the temporal organization and transitions among the modes defining the phase space. The following paragraph was added to the revised manuscript:

(line 232) "In climatology, variability encompasses the temporal amplitude fluctuations of a given variable. In contrast, Shannon Entropy, as calculated here, accounts for these variations with an amplitude filter. When we define a phase space using the leading PCs, we select a domain characterized by the patterns representing the system's greatest variance. By normalizing the indices and determining the threshold that maximizes Shannon Entropy, we effectively isolate temporal dynamics from the influence of amplitude variations. Our approach establishes a leveled ground for analyzing the temporal evolution of the system's state across its defining patterns, ensuring that the results remain independent of the specific amplitude differences simulated by various numerical models. In other words, although two models may reproduce the 1st PC with different amplitudes, they are both considered representations of the same climatic pattern. Consequently, Shannon's Entropy evaluates the system's persistence and transitions between states independently of these amplitude variations."

R2.2 Another notable shortcoming is the lack of validation against observational data, even though the authors acknowledge that such a comparison would be possible. This is a critical point — without observational benchmarks, we cannot determine whether the models' entropy values are realistic or merely reflect internal simulation dynamics.

We thank the reviewer for this comment. This suggestion prompted further discussion and led to new analyses and discussion in the revised manuscript. The Results section now includes observational datasets to compare the dominant model-derived principal components and Atlantic SST modes with satellite-era precipitation and SST patterns. As discussed in our response to Reviewer 1, observational data are not used to directly validate the pre-industrial or mid-Holocene simulations. Rather, they have been used as a reference for comparing mean entropy values across different simulations that have realistic observational-based standards. Furthermore, these observational entropy values were used for assessing whether principal-component-based or regional SST-index-based phase spaces provide a more reliable framework for characterizing Atlantic decadal variability. Since the comment overlaps with the concern raised by Reviewer 1, we refer the reviewer to our response to comment R1.5, where the corresponding revisions to the manuscript are described in detail.

R2.3 A further difficulty is the inconsistency in model parametrization. The models differ in terms of the factors they include (e.g., vegetation, dust, lakes), making comparisons challenging. The study lacks an attempt to isolate partial effects — for example, what specifically causes changes in entropy: dust, vegetation cover, or their combination?

Anticipating our response to Comment R2.8 regarding PCA and cross-model comparability, we note that principal components are extracted from the merged dataset, which enables direct comparison across different models and experiments. This key aspect may not have been sufficiently clear in the previous version of the manuscript. With this in mind, we discuss entropy differences between experiments while recognizing that such differences may arise from initial conditions or from boundary conditions and parameterized processes specific to each simulation. Without a large ensemble, it is not statistically robust to attribute entropy changes solely to boundary or parameterized factors. We therefore agree that applying this methodology to large ensembles could yield additional insights not addressed in this study. In particular, ensembles of numerical models or emulators that explore small perturbations in initial conditions under varying forcings, such as dust or vegetation, could reduce sensitivity to initial conditions and enable a more systematic investigation of the dependence of Shannon entropy on different climate variables and scenarios. However, the results presented in this study are based on a multi-model analysis using single realizations from distinct experiments. The primary objective of this work is to introduce a physically motivated phase-space framework for applying Shannon entropy to the characterization of climate variability in numerical models. A more systematic exploration using large ensembles or finely tuned forcing parameters is therefore beyond the scope of this introductory study. To make this clear, we have rewritten the following paragraph and added it to the manuscript:

(line 200) "The results presented in this study are based on a multi-model analysis using single realizations from each experiment. Because we do not analyze large ensembles of simulations, the conclusions drawn here are strictly conditional on the specific models and experiments considered. Within this context, the uncertainty associated with the entropy estimates arises primarily from the discretization of the principal-component phase space and the representation of its 27 possible states."

Furthermore, we added suggestions for future work in this theme to the Conclusion section as follows:

(line 432) "Because this study is based on single realizations, differences in entropy between experiments may reflect a combination of sensitivity to initial conditions and differences in boundary conditions or parameterized processes. Applying this methodology to ensemble simulations or emulators that systematically perturb initial conditions and external forcings (e.g., dust or vegetation) represents a natural extension of this work and would enable a more rigorous assessment of how Shannon entropy responds to different climate variables and scenarios."

R2.4 Moreover, the study does not quantitatively separate different sources of uncertainty. Although three types are mentioned — internal variability, discretization, and scenario-based uncertainty — their individual contributions to total variability are not assessed.

We thank the reviewer for this comment. We acknowledge that the previous version of the Methods section, mentioning uncertainties, was not sufficiently clear and combined several concepts, such as model uncertainty, entropy estimation, and threshold selection, and may have suggested an intention to quantitatively separate different uncertainty sources.

In the revised manuscript, we have reorganized and clarified this section to emphasize that its primary purpose is to justify the choice of an entropy-dependent discretization of the phase space. The discussion of internal variability, model formulation, and scenario-based uncertainty is now explicitly framed as motivation for adopting a maximum-entropy threshold, rather than as quantities to be separately quantified. Because climate models differ in their mean state, variance, and representation of variability, using a fixed threshold across simulations would bias entropy estimates toward amplitude differences. The maximum-entropy approach is therefore introduced as a means to enable consistent multi-model comparisons of decadal variability within a unified phase space.

The revised text now clearly distinguishes between the role of model uncertainties in motivating the methodology and the uncertainty associated with estimating entropy itself, which is quantified using a bootstrap procedure. We believe this reorganization resolves the ambiguity noted by the reviewer and more accurately reflects the scope and intent of the methodology.

Some paragraphs reflecting the changes made follow:

(line 208)"... climate models differ markedly in their representation of variability due to internal climate fluctuations, differences in numerical formulation and parameterizations, and uncertainties associated with imposed boundary conditions and forcings (Lehner et al., 2020). Applying a single fixed threshold across all models, therefore, risks producing entropy values that reflect differences in simulated amplitudes rather than differences in the temporal organization of variability.

To address this limitation, we adopt an entropy-centered discretization strategy in which the threshold is determined by the requirement of maximizing entropy, rather than prescribing entropy as a consequence of an arbitrary threshold choice. In this formulation… although two models may reproduce the 1st PC with different amplitudes, they are both considered representations of the same climatic pattern. Consequently, Shannon Entropy evaluates the system's persistence and transitions between states independently of these amplitude variations."

R2.5 While the selection of three principal components may be reasonable, the study does not examine the sensitivity of results to the inclusion of additional components.

In the merged dataset, the third principal component explains a relatively small fraction of the variance (approximately 11% for SST). Because each dimension in the phase space contributes equally to the Shannon entropy calculation, including lower-variance components would effectively assign the same weight to weak or potentially noise-dominated patterns as to the leading modes of variability. From a heuristic standpoint, this would imply treating low-variance fluctuations (e.g., patterns explaining less than 10% of the variance) as equally important as the dominant PC when characterizing tropical and South Atlantic decadal variability. For this reason, we do not consider a higher-dimensional phase space to be sufficiently physically grounded for this analysis. Since the phase-space construction intentionally removes amplitude information, it is particularly important to restrict the representation to physically meaningful and dynamically relevant patterns. This discussion is mentioned throughout the text, and it is now emphasized in the following paragraph of the Discussion section:

(line 401) "Modes such as El Niño, the AMM, and the AEM influence climate across the globe; they are known to impact society (McGowan et al., 2012; Lam et al., 2019; Anderson et al., 2018), the atmosphere (Xie and Carton, 2004; Gorenstein et al., 2023), and climate equilibrium (Pillai et al., 2022; Cai et al., 2021). The temporal evolution of these modes provides a conceptual framework for measuring decadal climate variability in numerical models using the same metrics applied to observational datasets. Accordingly, the phase space used to compute Shannon entropy is constructed to explicitly reflect the variability associated with these climate modes."

R2.6 The use of maximum entropy for each simulation as a reference point is statistically understandable but may lead to non-comparable thresholds and obscure differences stemming from less dynamic models. This approach might favor models that "artificially" gain entropy through threshold adjustments.

We appreciate the reviewers' comments on this topic, as they address central aspects of our analysis that we aim to clarify as clearly as possible. In standard practice, principal component time series are normalized, and applying a fixed threshold to define their phases is indeed a valid approach. However, entropy is inherently sensitive to how the system is coarse-grained. In the literature, thresholds ranging from 0.5 to 1.5 standard deviations are commonly adopted depending on the authors' definition of extreme events in their datapoints; however, small variations within this range can lead to substantially different entropy estimates, introducing considerable uncertainty associated with the choice of threshold. This issue is explicitly discussed in the manuscript as follows:

(line 204) "The value of Shannon's entropy depends critically on how the phase space is discretized. Small changes in the thresholds used to define the positive, negative, and neutral phases can substantially alter a simulation's trajectory through phase space and, consequently, its entropy (see Figures S2 and S3 in the Supplementary Material). A common approach in the literature is to normalize each principal component by its standard deviation and apply a fixed threshold to define these phases (typically ranging between 0.5 and 1.5). However, climate models differ markedly in their representation of variability due to internal climate fluctuations, differences in numerical formulation and parameterizations, and uncertainties associated with imposed boundary conditions and forcings (Lehner et al., 2020). Applying a single fixed threshold across all models, therefore, risks producing entropy values that reflect differences in simulated amplitudes rather than differences in the temporal organization of variability.

To address this limitation, we adopt an entropy-centered discretization strategy in which the threshold is determined by the requirement of maximizing entropy, rather than prescribing entropy as a consequence of an arbitrary threshold choice. In this formulation, the threshold is allowed to vary between simulations, ensuring that each model's variability is characterized using the discretization that best represents its exploration of the phase space."

Once our phase space is defined, the maximum entropy of each simulation emerges naturally. For this reason, the authors believe that the entropy-dependent threshold is a less biased approach to measuring a system's Entropy. To clarify which thresholds were used, Figures S2 and S3 have been modified to explicitly present the maximum entropy thresholds adopted for each component in each simulation as follows:

“

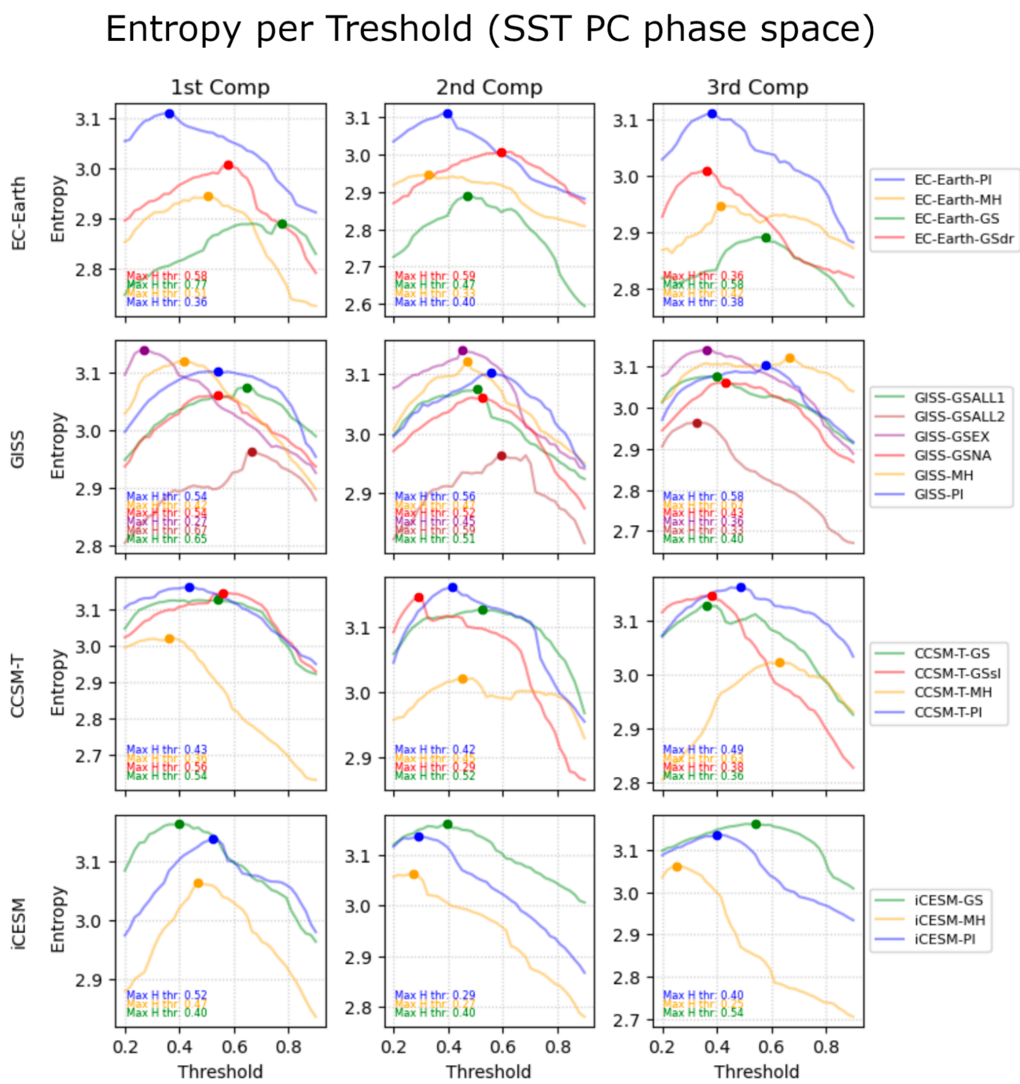


Figure S3. SST entropy values for each model scenario with different thresholds. The y-axis is the Entropy value, calculated from equation 1. The x-axis is the threshold used to define the positive, negative, and neutral phases of the three main components (a.k.a possible states) in the PCs' space. At each column, the thresholds from two of the main components are maintained fixed at 0.5, while the remaining component threshold varies. From left to right: the first, second, and third components' thresholds are varied. From top to bottom, the model runs are: EC-Earth - Pre-Industrial (PI, blue), mid-Holocene (MH_PMIP, orange), Green Sahara (MH_GS, green), and Grenn Sahara with dust reduction (MH_GSrd, red); GISS - Full vegetation (MH_GSALL1, green; MH_GSALL2 brown), Extra-tropical vegetation (MH_GSEX, purple), North Africa vegetation only (MH_GSNA, red), MH_PMIP (orange) and PI (blue); iCESM - MH_GS (green), MH_PMIP (orange) and PI (blue); CCSM-T - MH_GS (green), GS with soil and lake inputs (MH_GSsl, red), MH_PMIP (orange) and PI (blue)

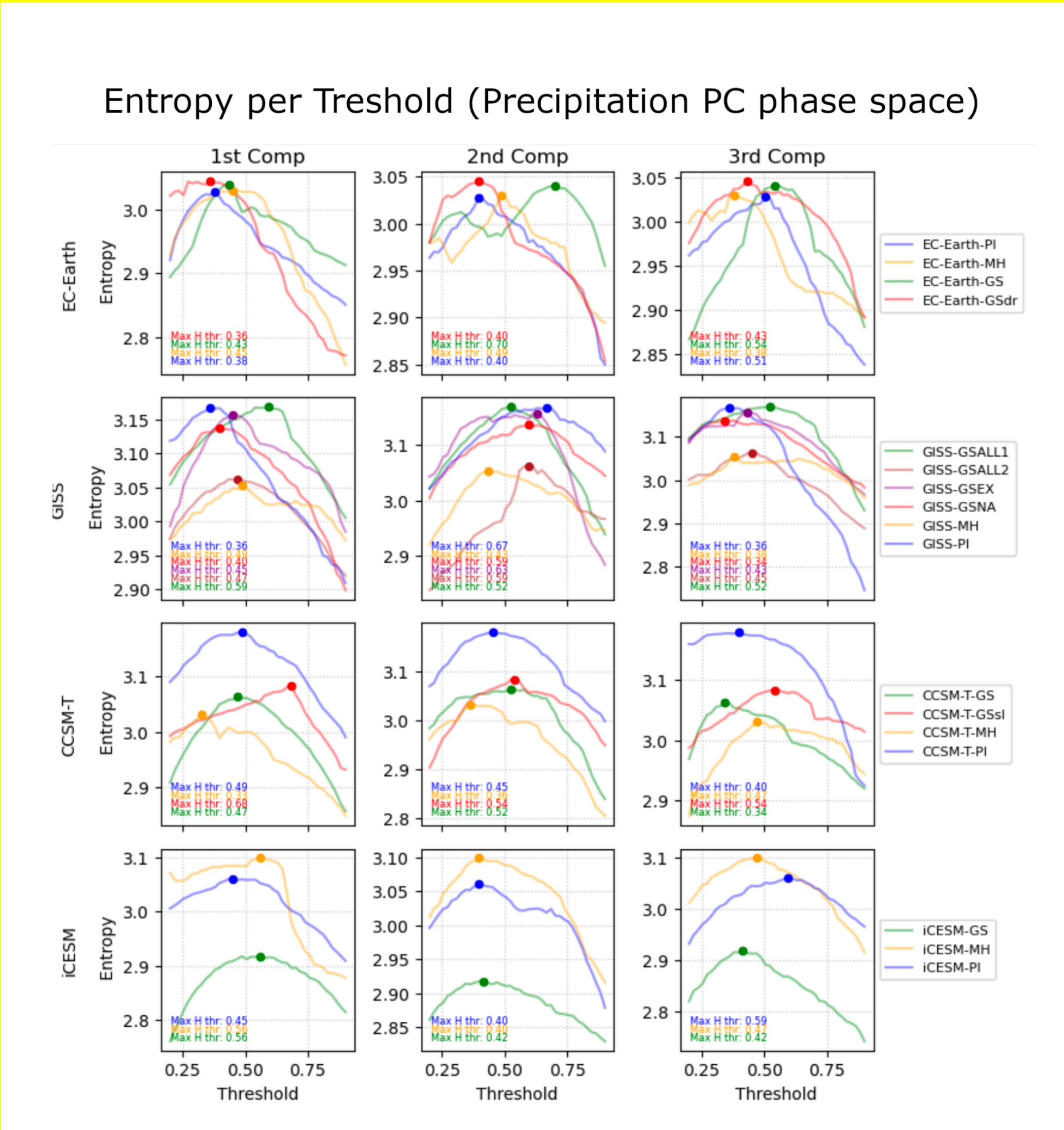
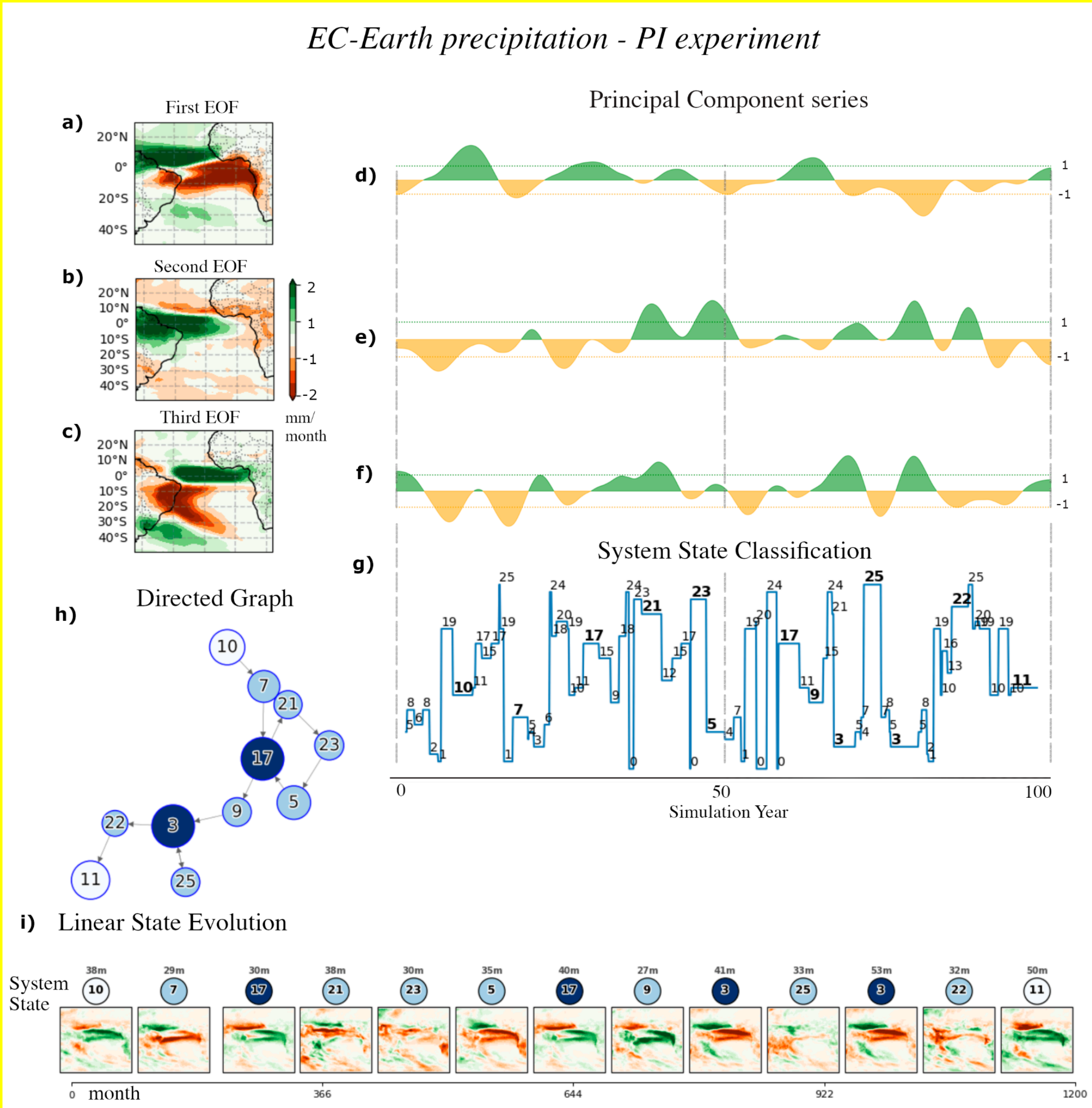


Figure S4. Precipitation entropy values for each model scenario with different thresholds. The y-axis is the Entropy value, calculated from equation 1. The x-axis is the threshold used to define the positive, negative, and neutral phases of the three main components (a.k.a possible states) in the PCs' space. At each column, the thresholds from two of the main components are maintained fixed at 0.5, while the remaining component threshold varies. From left to right: the first, second, and third components' thresholds are varied. From top to bottom, the model runs are: EC-Earth - Pre-Industrial (PI, blue), mid-Holocene (MH_PMIP, orange), Green Sahara (MH_GS, green), and Grenn Sahara with dust reduction (MH_GSrd, red); GISS - Full vegetation (MH_GSALL1, green; MH_GSALL2 brown), Extra-tropical vegetation (MH_GSEX, purple), North Africa vegetation only (MH_GSNA, red), MH_PMIP (orange) and PI (blue); iCESM - MH_GS (green), MH_PMIP (orange) and PI (blue); CCSM-T - MH_GS (green), GS with soil and lake inputs (MH_GSsl, red), MH_PMIP (orange) and PI (blue)”

R2.7 ***Lastly, the graphical representation of results as directed graphs is visually complex and difficult to interpret.***

The directed graphs are intended to illustrate the cyclicity of the climate system and the modes that define the phase space. To address the reviewer’s concern, Figures 1,3 and 4 have been redrawn to more clearly highlight the connection between the directed graphs and the underlying phase space. To facilitate interpretation, we also added a frame-by-frame depiction of the system’s state evolution along a linear time axis, presenting the same information in an alternative graphical structure. For example, the new Figure 4 follows:

“



EC-Earth precipitation state identification and evolution during the PI experiment. Panels (a)–(c) show the merged ensemble’s first three EOF patterns, with (d)–(f) displaying their corresponding PC series (green/orange indicating positive/negative phases; dashed lines showing 1 standard deviation). (g) identifies discrete system states based on these PC indices, bold font used in the most persistent states (lasting more than 24 months); (h) illustrates state evolution of the most persistent states. (i) tracks the state evolution in linear form. Each node displays its cluster number, spatial pattern, and duration in months before transitioning.”

Furthermore, following the suggestions of Reviewer 1, a 2D phase space solution of the Atlantic SST system has been added to the manuscript. Using only the two main PCs in a simplified problem, the system's trajectory in phase space can be plotted in x and y coordinates, and the directed graph takes an obvious structure. More details about this simple direct graph construction are in response to comment R1.2.

R2.8 ***One final comment, offered with all due respect and goodwill: a common PCA analysis should be performed for all models and for each variable (SST and precipitation) using a merged dataset from all models and experiments. This would ensure a shared phase space and resolve the issue of cross-model comparability.***

A common phase space is indeed used, and we acknowledge that this was not stated with sufficient clarity in the previous version of the manuscript. In the revised manuscript, all results based on the principal-component phase space are derived from a unified PCA constructed from the merged dataset encompassing all models and experiments. The phase space is defined by the three leading components extracted from this combined dataset. Together with the use of the maximum-entropy threshold, this approach ensures cross-model and cross-scenario comparability. The revised manuscript now states this explicitly in the paragraph describing the phase-space construction.

(line 135) "In this approach, we extracted two distinct phase spaces (one for SST and another for precipitation) composed of EOFs derived from the combined simulations of all models and scenarios. This process yields a unified phase space with a shared spatial structure for the entire ensemble, providing a consistent framework for analyzing and comparing variability across different simulations (Chandler et al., 2024)."