São Paulo, 26 February 2026

Response to reviewers for the revised manuscript: “The Atlantic Ocean's Decadal Variability in mid-Holocene Simulations using Shannon's Entropy”.
Manuscript ID: egusphere-2025-921

We thank both reviewers for their careful reading of the manuscript and for their constructive comments and suggestions. In response, we have substantially revised the manuscript to sharpen its focus on the proposed methodological framework rather than on the interpretation of individual results.

In this revision, we emphasize the methodology and demonstrate its applicability using both principal components derived from a merged multi-model dataset and regional-box SST indices to define a physically motivated low-dimensional phase space for the tropical and South Atlantic region. We then use Shannon entropy to compare the different experiments within these phase spaces, alongside observational datasets.

This refocusing motivated additional analyses and clarifications, which led to new findings and refined conclusions that improve the robustness and clarity of the study. All corresponding changes have also been implemented in the code and are reflected in the updated GitHub repository. We believe that the manuscript has been significantly improved through these revisions and that all reviewers’ concerns have been addressed. We therefore hope that the revised manuscript is now suitable for publication in *Geoscientific Model Development*.

Below, we provide detailed responses (in blue) for all the reviewers' comments (marked as R1 and R2 in black). Line numbers and manuscript excerpts (highlighted in yellow) in our responses refer to the revised version of the manuscript.

On behalf of all authors,
Sincerely,
Iuri Gorenstein

Corresponding author.

Response to Reviewer #1 (Chris Brierly):

This piece of work applies a new methodology to look at the decadal variability to understand the response across an ensemble of idealised paleoclimate simulations. I do not see anything incorrect in the work. But I'm not sure that the current layout will result in many citations. This is because the manuscript jumps between trying to describe two things simultaneously: a new method and some scientific results. As the manuscript is submitted to GMD, I presume that the authors consider the new method to be the primary innovation and will address my comments and recommendations accordingly.

R1.1 I see that there are three important facets of the methodology that are outside of the standard approach. Firstly, there is the fact that the EOFs are computed by looking across the whole ensemble – rather than separately within each single model. This is a nice touch and should be made clearer within the text (currently this is mentioned briefly on L123, but not stressed as key aspect of the methodology). I happen to have adopted this approach myself before (Chandler et al, 2024, https://doi.org/10.1175/JCLI-D-23-0089.1) to look at regional models. However, we were looking at the mean climate, not variability, and the primary modes were detected between the models. I was therefore surprised that your approach does not pick up any inter-model variations. I guess that this comes from the application of the decadal filtering, which is in effect removing the mean climate from the individual models. You ought to explain in more detail what kind of filter is being applied, and its implications. I suspect that you are using a band-pass filter, and that if you instead used a low-pass filter you find fundamentally different EOF patterns (more akin to the EPP, we describe in Chandler et al). The labelling of Fig 1 and 2 implies that they show EC-Earth's EOF patterns – rather than full ensemble patterns. Only the PCs, directed graph and transition timeseries relate to the particular EC-Earth simulation.

We thank the reviewer for his comments. Indeed, our methodology uses a multi-model merged dataset to extract the EOFs and project each simulation's principal components. We find the cited paper that uses a similar approach very informative. As mentioned in the reviewer's comment, the decadal filter helps in maintaining these EOFs focused on the mean representation of decadal climate and not in particular anomalies that may dominate the tropical and South Atlantic variability in some simulations. We have added comments to clarify that we calculated the ensemble using a merged dataset in the methodology, figure labels, and where needed. Some paragraphs that address this comment added to the revised manuscript are shown below:

(line 252) "To make sure we extract physically meaningful patterns, we create the merged data set's phase-space using the Tropical South Atlantic three leading EOFs (Figure 3, 'a', 'b' and 'c')"

(line 142) "**Low frequency filters.**

To investigate decadal variability in the Atlantic Ocean, decadal filters were applied across all datasets, calculated as simple decadal means from the original monthly time series. These filters serve two primary purposes: first, to examine the decadal coupling between precipitation and SST variables; and second, to extract consistent multi-model patterns that serve as the foundation for a shared phase space. Decadal filtering is applied to the precipitation and SST datasets before pattern extraction (see Methods, section 2.2.2) to effectively smooth out fine-scale structures and disparities arising from different model architectures. Since we define a single phase space to encompass the PCs of the merged dataset, this filtering process is essential to ensure that the leading PCs capture substantial variance across the integrated multi-model ensemble."

(line 135) "... we extracted two distinct phase spaces (one for SST and another for precipitation) composed of PCs 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. This procedure is discussed and applied by Chandler et al. (2024), who construct a merged dataset across models and extract dominant modes of variability using principal component analysis. Their approach implicitly defines a shared low-dimensional representation for inter-model comparison, which is conceptually related to our phase-space construction."

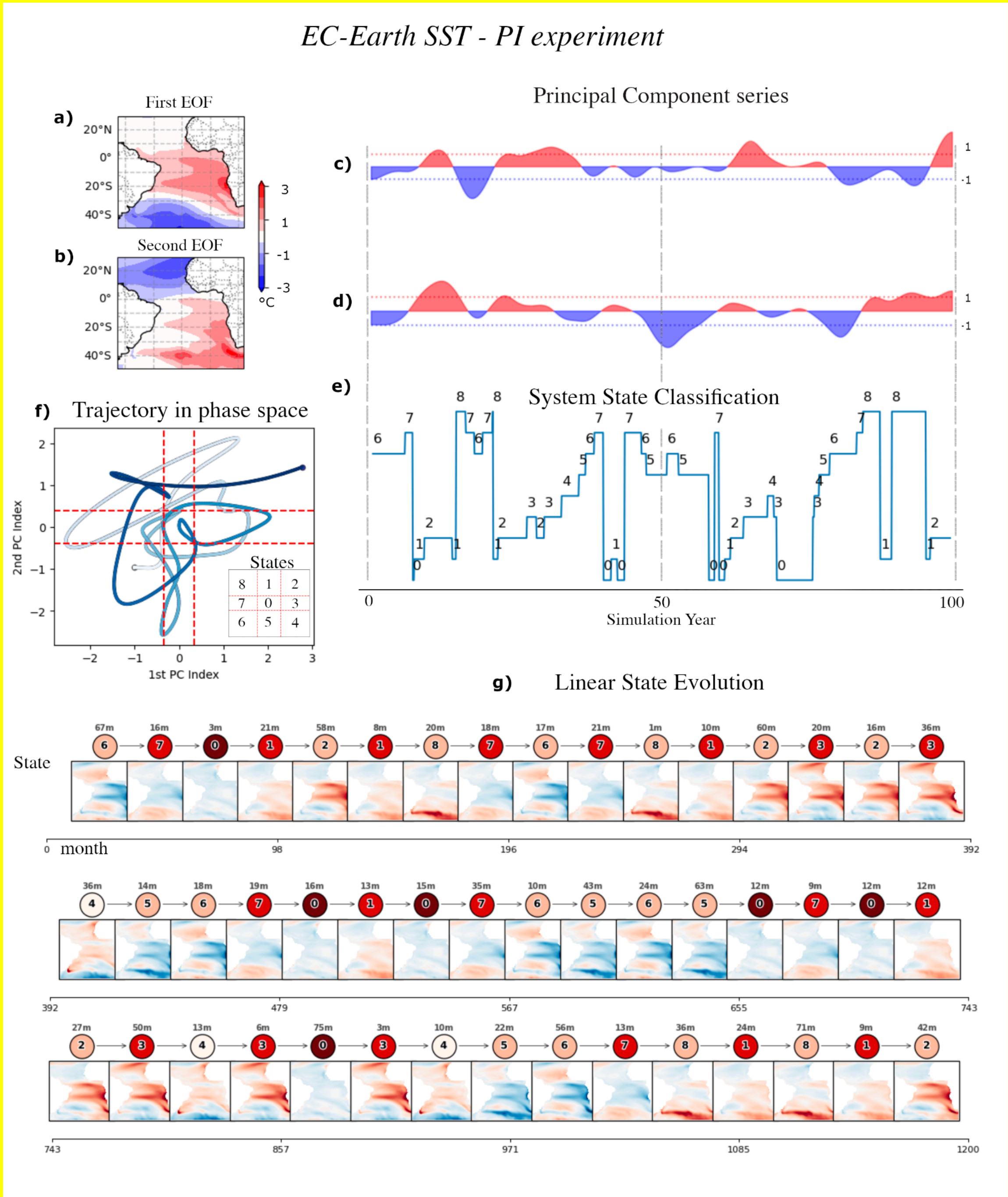
Furthermore, alongside the standard procedure of normalizing each PC by its standard deviation, our methodology uses the concept of maximum entropy threshold, which works as a filter of the principal components' amplitude. That way, Shannon's Entropy becomes a more reliable quantitative measure for multi-model comparison. A paragraph extracted from the revised manuscript that addresses this follows:

(line 233) "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."

R1.2 ***Your second innovation is the introduction of the directed diagrams. I confess that I find these hard to interpret, but I can see that they are important. Can you please spend a bit more time describing them? Perhaps thinking of them in a reduced dimension set would help; say by using the 2 ENSO modes of Ren & Jin (2011, https://doi.org/10.1029/2010GL046031) at then you can place the phase along the x and y axes. Can you also create some possible directed graphs for a system in which all the PCs are truly independent, through building a simple statistical model? These would then provide a suitable null-hypothesis and allow some statistical testing to be undertaken.***

The authors greatly appreciate the reviewers'suggestion. A new test example of the system's state evolution in a two-dimensional map was created using only the 2 main PCs of the Atlantic Ocean SST. Each axis represents one of the 2 dimensions from this simplified phase space, creating an obvious analogy to the direct graphs. The suggestion to shift the analysis from the Atlantic to the Pacific Ocean is also appreciated; however, we aim to retain the focus of this study on the Atlantic region. The new Methods section has the trajectory in phase space section updated as follows:

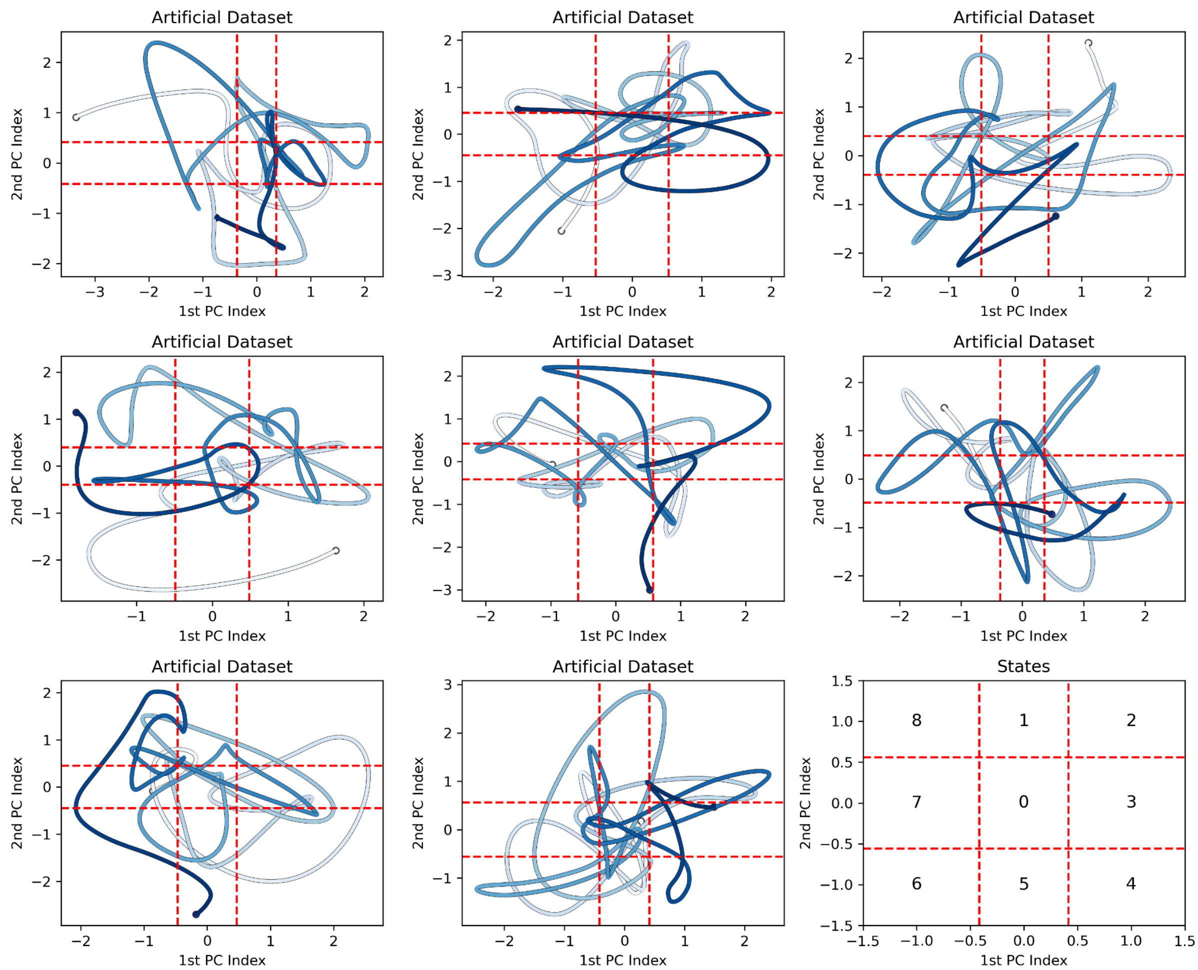
(line 163) "Projecting a simulation series onto the 'n' dimensions of our phase space generates a trajectory in that phase space. To elaborate more, we present a simplified 2-dimensional phase space for the SST decadal anomalies of the EC-Earth - PI experiment, using the two leading EOFs from the merged dataset. First, we apply the dimensionality reduction using the PC analysis to find the merged dataset's two leading EOFs (Figure 1 'a' and 'b'). Then, we project the EOFs into the simulation's Atlantic Ocean time series, generating its PCs (Figure 1 'c' and 'd'). The trajectory of our system in the 2-D continuous phase space is shown in Figure 1 'f'. Defining negative, neutral, and positive phases for each index, using a threshold linked to Shannon's Entropy (as described in Section 2.3), we coarse-grain this space into n states (9 possible states for this 2-D problem) and project our system on it (Figure 1 'e'). Finally, the trajectory can also be seen as a discrete system evolving in time, from one possible state to the next, as depicted in Figure 1 'g'.

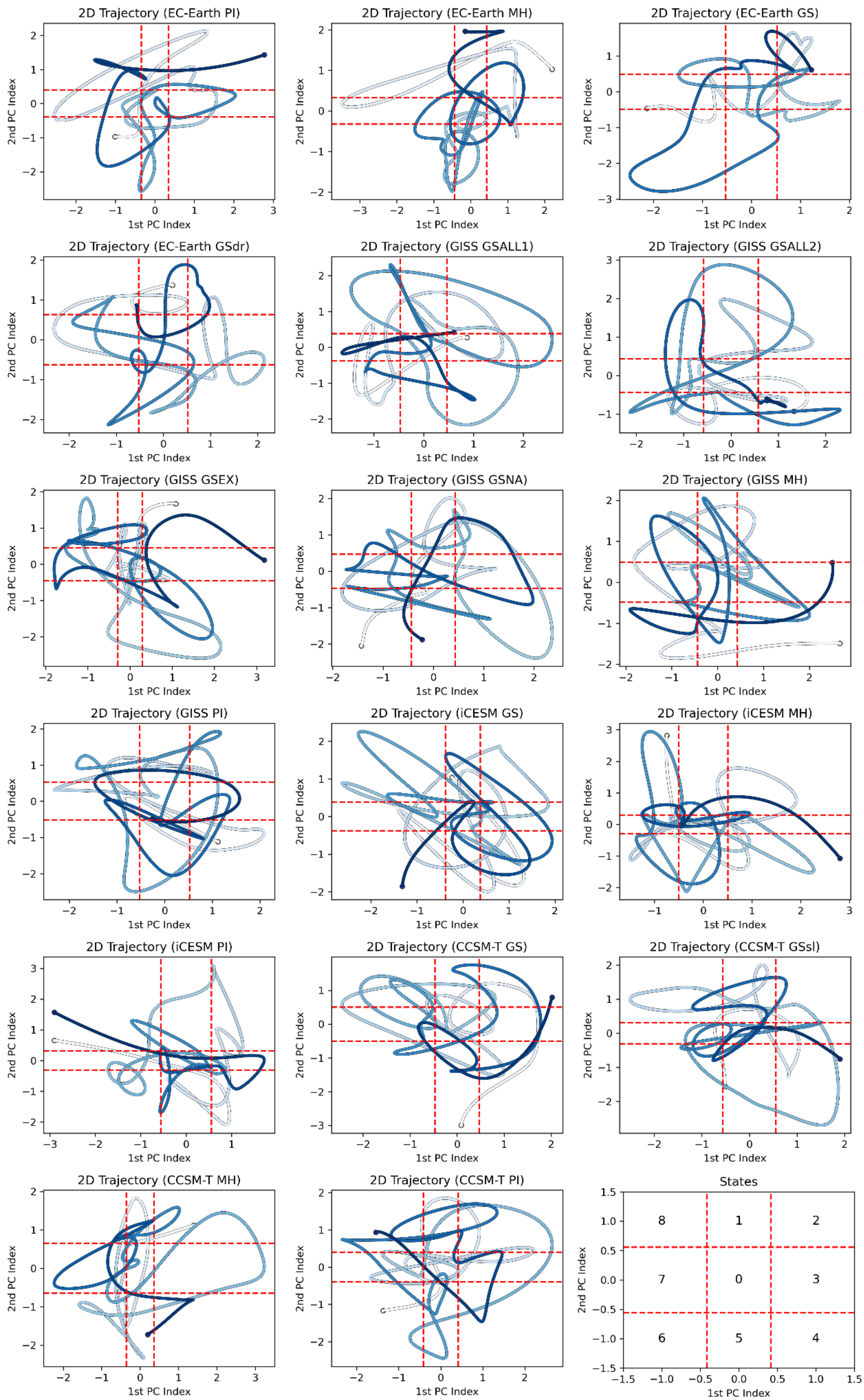


EC-Earth SST system state identification and evolution during the PI experiment. a) The first EOF; b) The second EOF; (c) and (d): their respective PC series, blue for negative and red for positive, dashed lines indicating the 1 standard deviation value. (e) The cluster identification of the discrete system state, given by the two PC indices above. (f) The system's trajectory in the continuous PC phase space. The line's color represents the time evolution, white for the start (month 0) and dark blue for its end (month 1200). (g) The diagram illustrates the state-by-state evolution of the system over time. Each node represents a specific SST pattern, identified by its cluster number (referenced in panels (e) and (f)) and its spatial pattern displayed below the node. The numeral positioned above each node indicates the duration, in months, that the system remained in that state before transitioning. Darker shading on a node signifies a higher frequency of transitions into that specific cluster throughout the entire time series.

Since the continuous trajectory in a high-dimensional phase space can be challenging to plot, the trajectories of a system in a higher-dimensional phase space can be depicted with directed graphs (Figure 2; 'h' in Figures 3 and 4; and Figures 5-8). These graphs can hold different information regarding the system's trajectory in phase space. Each node in these graphs represents a state of the system at a particular time step, with the node's size indicating the number of months the system remained in that state. Nodes that appear darker have a higher degree, meaning they are connected to more transitions to and from other states. This indicates that the system frequently returns to the same state, which results in a darker color for that node. The distance between nodes is irrelevant; it was adjusted for the graph. The information in a directed graph can be overwhelming to analyze for every simulation run; therefore, we utilize its information to calculate a macro-property that reflects its variability in phase space.

Furthermore, artificial sets were created with two independent random normally distributed indices. However, there was no notable difference between the evolution of these systems and their use of a 2-dimensional PC phase space. To satisfy the reviewer’s curiosity, we present some of the artificial trajectories created from independent 2-D random Gaussian series alongside the 2-dimensional trajectories of the leading SST modes of the Atlantic Ocean in different simulations used in the manuscript. These plots will also be available in the GitHub repository, along with the code used to generate them.




2D Trajectory (EC-Earth PI)
2D Trajectory (EC-Earth MH)
2D Trajectory (EC-Earth GS)
2D Trajectory (EC-Earth GSdr)
2D Trajectory (GISS GSALL1)
2D Trajectory (GISS GSALL2)
2D Trajectory (GISS GSEX)
2D Trajectory (GISS GSNA)
2D Trajectory (GISS MH)
2D Trajectory (GISS PI)
2D Trajectory (iCESM GS)
2D Trajectory (iCESM MH)
2D Trajectory (iCESM PI)
2D Trajectory (CCSM-T GS)
2D Trajectory (CCSM-T GSsl)
2D Trajectory (CCSM-T MH)
2D Trajectory (CCSM-T PI)
States
8
1
2
7
0
3
6
5
4
1st PC Index
2nd PC Index

R1.3 Your third innovation involves the use of Shannon's entropy. You provide an equation for this and then describe some of its properties in the methods section. However, when showing the results, all the values are approximately 3. I didn't get why that should be so. I guess it's related to either the fact there are 3 EOFs or the fact that each PC is divided into 3 states, but I don't know which and you didn't explain.

The entropy values are close to 3. This is due to the structure of our phase space: with three defined phases (positive, neutral, and negative) for each of the three principal components, the system has $3^3 = 27$ possible states. The maximum entropy (defined by the manuscript's equation 1) of a discrete space such as this would result in $\ln(27) \approx 3.296$. Each simulation's threshold is tuned to yield its maximum possible entropy, which naturally approaches $\ln(27)$. The following paragraph was added to the manuscript:

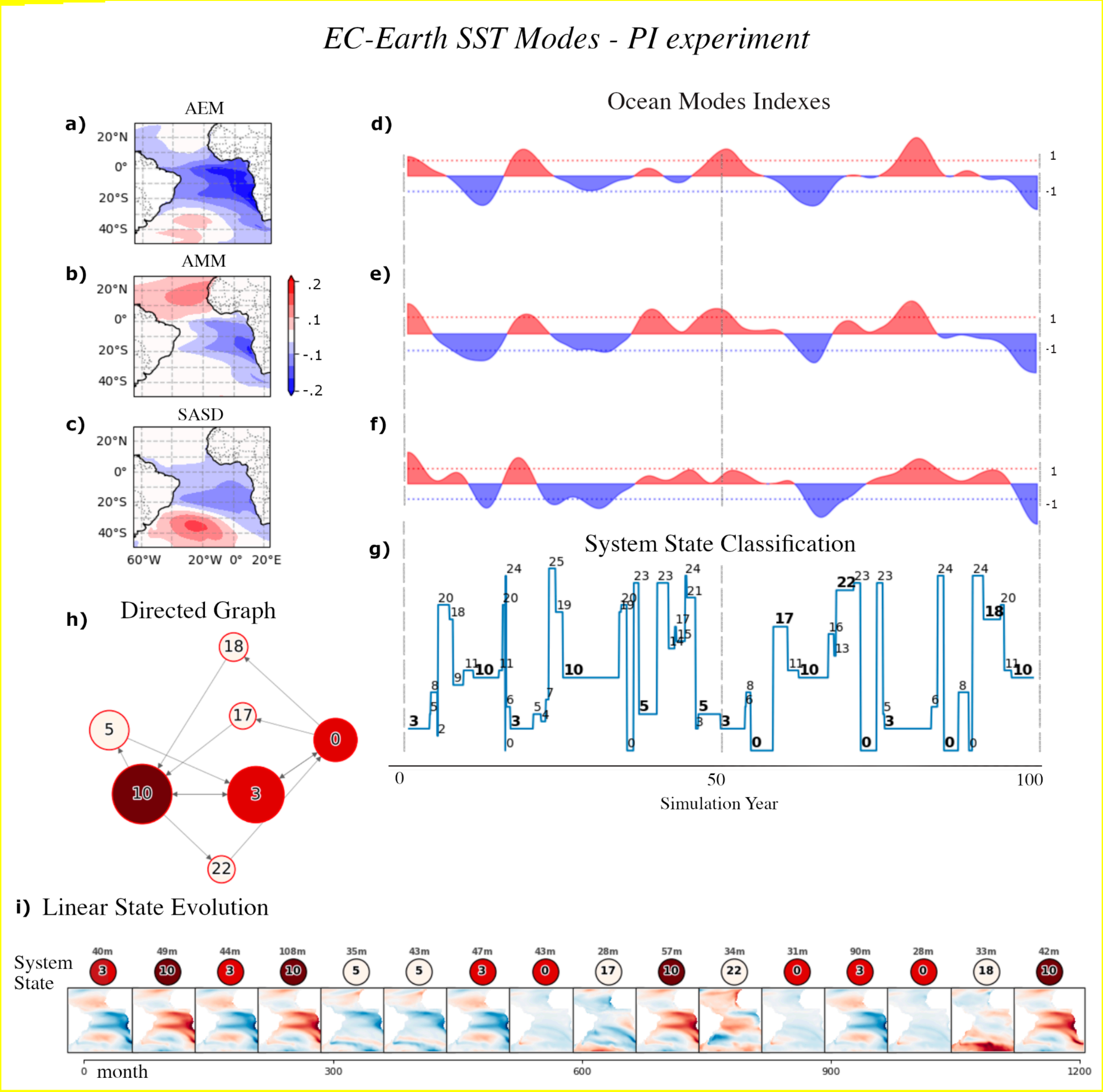
(line 289) "As seen in Figure 9 and Table 1, the entropy values are approximately 3. This is due to the structure of our phase space: with three defined phases (positive, neutral, and negative) for each of the three principal components, the system has $3^3 = 27$ possible states. The maximum entropy (equation 1) of a discrete space such as this would result in $\ln(27) \approx 3.296$. Since the simulation's thresholds in phase space are individually tuned to yield its maximum possible entropy, it naturally approaches $\ln(27)$."

I have 3 other substantive comments:

R1.4 Both the AMM and AZM/AEM are more often defined by SST indices, rather than through principal components. It would be informative to compare your new method against this style of definition.

This revised manuscript's Result section uses the traditional AMM, SASD, and AEM modes to define an additional 3-dimensional phase space, yielding different comparable Entropy results. A new discussion of the PCs and Atlantic Ocean Modes approaches to defining the phase space has been added. To highlight some of the additions made in the rewritten Results section, the new figure in Results follows:

(line 315) "... An alternative approach to reducing Atlantic SST dimensions into a 3D phase space uses regional SST indices. These averages provide simple, physically interpretable metrics of climate mode strength. To test how phase space definitions affect Shannon's Entropy, we repeat our calculations using indices for the AMM, AEM, and SASD."



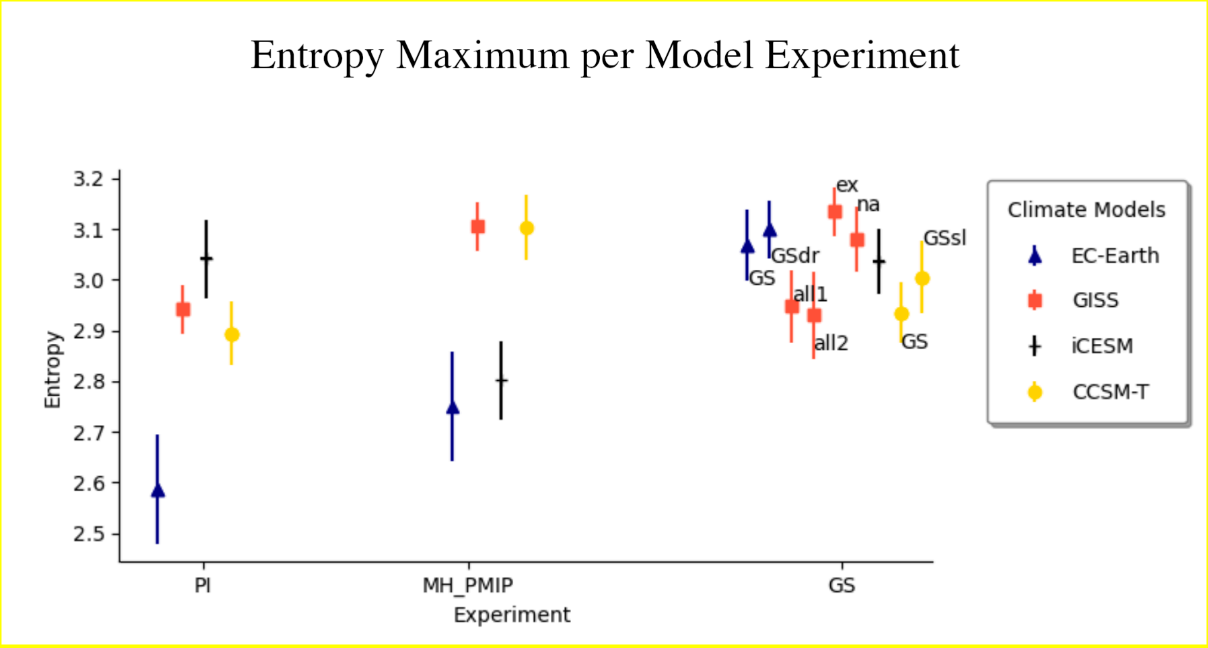
EC-Earth SST system state identification and evolution during the PI experiment. (a),(b), and (c): The merged ensemble AMM, AEM, and SASD, respectively; (d),(e), and (f): their respective index series, blue for negative and red for positive, dashed lines indicating the 1 standard deviation value. (g) The cluster identification of the discrete system state, given the three above Ocean Mode 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.

In the merged dataset, the first PC correlates strongly with the AEM (70%) and SASD (64%), while the second PC correlates with the AMM (71%). However, while the SST indices are significantly intercorrelated (39–43% in the merged dataset), the PCs maintain negligible correlation (below 3%). As an orthonormal basis, PCs inherently ensure minimal correlation between dimensions, unlike regional indices.

This index-based phase space yields generally lower entropy values than the PC-based space. While Shannon Entropy does not directly measure correlation, the dependencies between the regional indices significantly impact the system's total entropy.

Furthermore, this exercise helped debate the sensitivity of Shannon's Entropy to the phase space. We show the following paragraph from the revised manuscript to highlight some of the additions made:

(line 327) "



Maximum Entropy values and uncertainty for each model experiment for SST using the Atlantic Ocean modes phase space. From left to right: Pre-Industrial PI; mid-Holocene only orbital forcing (MH_PMIP); and mid-Holocene with Green Sahara boundary conditions (MH_GS) experiments. All models include PI, MH_PMIP, and MH_GS runs. Under Green Sahara conditions, additional experiments include: EC-Earth (ECE) with northern African vegetation and dust reduction (dr); GISS with both extratropical and northern African vegetation (GISSall1 and GISSall2), with only extratropical vegetation (GISSex), and with only north African vegetation (GISSna); CCSM-T with soil and lakes (sl).

EC-Earth

The EC-Earth PI experiment (Figure 10 - 'd', 'e' and 'f' and first line of Table 2) best illustrates how high correlation results in low entropy, showing the lowest entropy of all simulations within the Atlantic Ocean Modes phase space. In this specific setup, the AEM and AMM are correlated at 88%, while the AMM and SASD show a 76% correlation. Conversely, all MH scenarios exhibit higher entropy than the PI (6% for MH_PMIP and 19% for GS with/without dust) as these Atlantic modes decouple.

GISS

Compared with the PC phase space, GISS PI entropy reduces using the Atlantic ocean modes indices (Table 2). The entropy from the remaining MH scenarios mostly decreases; however, they show no significant changes to the entropies calculated in the PCs phase space.

iCESM

For this model, the SST entropies show a general decrease, while holding the same internal biases seen in the PC phase space entropies (Table 2).

CCSM-Toronto

The only significant difference is still between PI and MH_PMIP; however, in this phase space, MH_PMIP holds the largest entropy amongst this model's different experiments (7% higher than PI)."

R1.5 There is no comparison to observations contained within your analysis. Can you project the resultant EOFs onto a reanalysis dataset and look at the real world?

This revised manuscript's Result section uses observational data to compare the model's PCs and SST Atlantic Ocean modes to the satellite era precipitation and SST patterns.

(line 242) "We test our methodology in three steps. First, we analyze the outputs of four different numerical climate models in PI and various MH climate scenarios (see Methods). We extract the tropical and South Atlantic SST and precipitation three leading Principal Components (PCs), calculate their trajectory in this 3-dimensional phase space, and compute their respective Shannon's Entropy. In the second, we explore the classical Oceanography SST-based indexes to create the Tropical and South Atlantic Ocean's SST phase space (see Methods), calculate its trajectory, and Shannon's Entropy once again. The third step is a comparison between the calculated Entropy values from the model simulations and observational data from satellites using the previously discussed phase spaces."

(line 343) "A primary goal of numerical climate models is the faithful reproduction of low-frequency decadal variability. Notable differences exist between the PCs extracted from observational data and model data, nor do they represent the same variance as the principal components from the leading components seen in the observational satellite era (Figure S1). However, the EOFs extracted from the merged simulations dataset can be projected onto the observational data time series. As long as they have the same length, the use of maximum Entropy enables a comparison between the observational data and the simulations' trajectories in a common phase space. Since the observational datasets range from 1980 to 2015, the results shown in the Table ahead are the Entropy values of the first 420 months (35 years) of each simulation. Within these time windows, the resulting entropy values are lower, indicating that shorter intervals do not fully capture the decadal variability envelope of the tropical and South Atlantic system.

Table 3: Entropy mean and Standard Deviation from the 420-month-long series model runs and the observation data set

Model	Scenario	Entropies in PCs' phase space H_{sst}	Entropies in PCs' phase space H_{ppt}	Entropies in Modes' phase space H_{sst}
ECE	PI	2.60 ± 0.11	2.70 ± 0.08	2.26 ± 0.13
ECE	MH_{PMIP}	2.79 ± 0.09	2.91 ± 0.13	2.45 ± 0.10
ECE	MH_{GS}	2.63 ± 0.07	2.58 ± 0.09	2.80 ± 0.11
ECE	MH_{GSdr}	2.62 ± 0.09	2.80 ± 0.08	2.67 ± 0.09
GISS	PI	2.73 ± 0.08	2.95 ± 0.08	2.76 ± 0.09
GISS	MH_{PMIP}	2.75 ± 0.08	2.63 ± 0.10	2.81 ± 0.09
GISS	MH_{GSall1}	2.83 ± 0.10	2.78 ± 0.14	2.64 ± 0.10
GISS	MH_{GSall2}	2.85 ± 0.09	2.68 ± 0.10	2.62 ± 0.11
GISS	MH_{GSex}	2.95 ± 0.12	2.74 ± 0.10	3.04 ± 0.09
GISS	MH_{GSna}	2.80 ± 0.10	2.81 ± 0.11	2.72 ± 0.10
iCESM	PI	2.71 ± 0.10	2.72 ± 0.10	2.77 ± 0.09
iCESM	MH_{PMIP}	2.85 ± 0.09	2.79 ± 0.08	2.74 ± 0.12
iCESM	MH_{GS}	2.90 ± 0.11	2.71 ± 0.10	2.84 ± 0.09
CCSM-T	PI	2.90 ± 0.09	2.56 ± 0.17	2.72 ± 0.11
CCSM-T	MH_{PMIP}	2.57 ± 0.10	2.70 ± 0.07	2.57 ± 0.10
CCSM-T	MH_{GS}	2.57 ± 0.07	2.92 ± 0.13	2.73 ± 0.12
CCSM-T	MH_{GSsl}	2.65 ± 0.09	2.76 ± 0.10	2.83 ± 0.08
Observation	$1980 - 2015$	2.83 ± 0.09	2.78 ± 0.09	2.40 ± 0.13

"

"

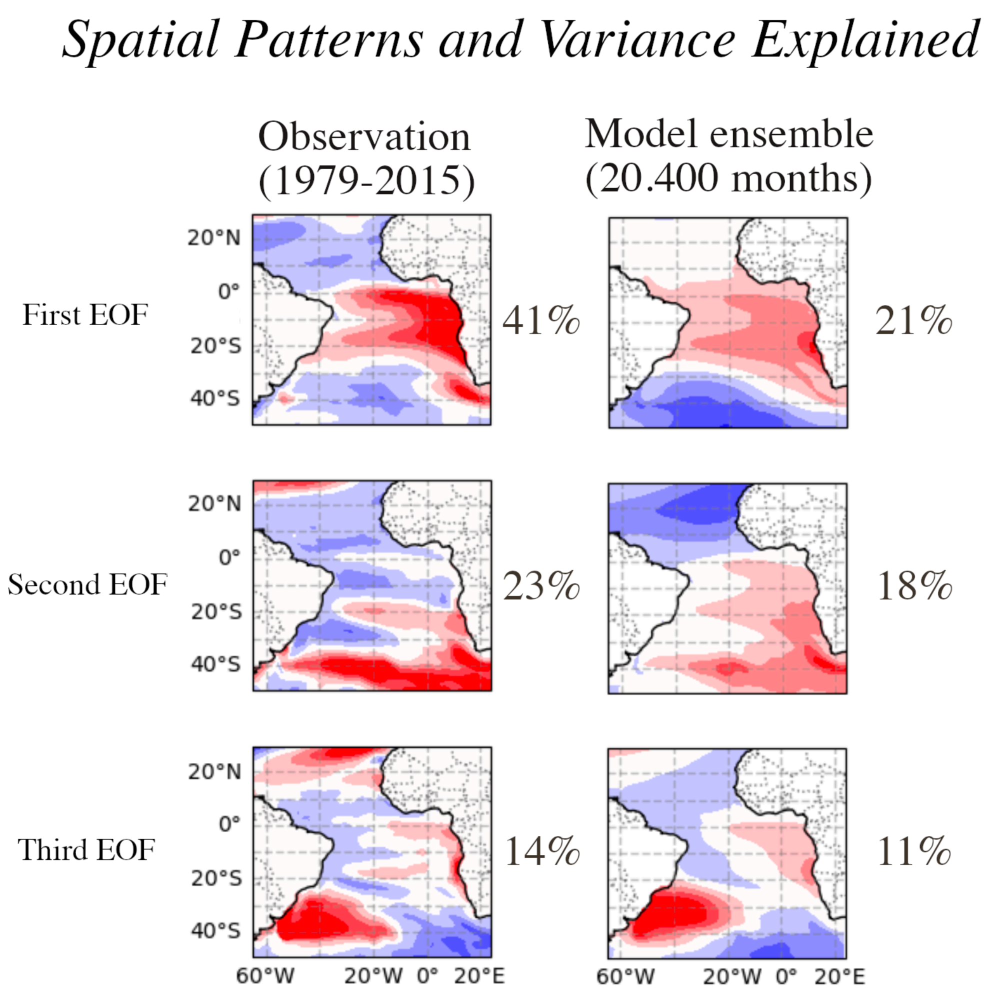


Figure S1. The three leading empirical orthogonal functions (EOFs) from the principal component analysis. From left to right, the observational EOFs, the merged model ensemble EOFs, and the percentage of explained variance in their respective dataset. From top to bottom, the first, second, and third leading modes."

Furthermore, we do not wish to overstate results about the mid-Holocene experiments using the observational dataset alone; when using numerical climate models, it is wisest to compare satellite era data with historical experiments. However, we try to use these results to compare the mean Entropy from the multi-model ensemble in each phase space, discussing what the observation dataset's entropy indicates is the most reliable procedure (the PC or the regional SST boxes). The discussion made in the manuscript follows:

(line 352) "Typically, satellite-era observations are compared against historical model runs. Previous research categorized the tropical and South Atlantic systems into a discrete evolution of patterns, tracking the SST and precipitation cycles using reanalysis data from 1890-2015 with a directed graph format (Gorenstein et al., 2023). However, as an attempt to quantitatively compare the values emerging from the numerical climate models' entropy in these MH experiments, we calculate the entropy from two observational datasets (the HadISST and the GPCP datasets - see Methods).

In the PC phase space, the ensemble model mean (2.76 for SST and 2.75 for precipitation) falls within the observational entropy uncertainty interval. However, in the Atlantic Ocean Modes phase space, the ensemble model mean (2.70) is significantly higher than the observational entropy value (2.40 ± 0.13). While this could suggest that historical variability is lower than most of the PI and MH experiments presented, such a direct comparison assumes the models represent these modes accurately. Instead, this discrepancy may indicate that, for this methodology, studying Atlantic decadal variability in simulations is more reliable when using principal components rather than traditional SST indices."

Finally, these changes are also reflected in the conclusion section as follows:

(line 417) "Across four models, the results show that mid-Holocene forcings can significantly alter the degree of organization of Atlantic decadal variability, with model-dependent entropy changes, and different SST and precipitation responses. Comparisons with observational datasets indicate that PC-based phase spaces provide a more consistent and robust basis for evaluating low-frequency Atlantic variability than traditional SST indices, highlighting entropy as a powerful metric to diagnose how external forcings reshape the structure, persistence, and transitions of dominant climate modes."

(line 437) "This methodology characterizes system dynamics and variability in terms of state occupancy and transitions rather than absolute anomaly magnitudes. By discretizing the climate system in a unified low-dimensional phase space and computing Shannon's entropy, the underlying climate variability can be directly compared between different models, experiments, and observations."

R1.6 You have chosen a case study (idealised green Sahara experiments during the mid-Holocene) that is not readily accessible to the general reader, and therefore requires a substantial amount of additional introduction and methods that get in the way of the methodological innovations you are trying to explain.

We agree that other types of data or model experiments could be equally interesting to investigate using our methodology. In the previous version of the manuscript, the discussion focused too heavily on the results derived from the chosen dataset, which may have obscured a more direct presentation of the proposed methodology. To address the reviewers' concerns, the revised manuscript has been substantially rewritten to emphasize the methodological framework rather than the specific results of the Green Sahara. In particular, the Introduction and Discussion sections were revised to remove mid-Holocene-specific details that may not be of broad interest, while retaining this period as a motivating example for applying a novel methodology to study climate variability in the tropical and South Atlantic region. The revised paragraphs in the Introduction section now read as follows:

(line 35) "The mid-Holocene (MH) period, approximately 5000-7000 years Before Present, provides a natural framework to investigate how external forcings reshape climate variability. Orbital changes during this period altered the seasonal and latitudinal distribution of insolation, leading to profound hydroclimatic changes, most notably the African Humid Period and the expansion of vegetation over the Sahara. Proxy studies and climate model simulations suggest that these boundary-condition changes affected not only mean climate states but also the dynamics of tropical and Atlantic variability through land–atmosphere–ocean feedbacks. Rather than simply amplifying or damping anomalies, quantifying how such forcings reorganize decadal variability remains an open problem, particularly when comparing across models with differing physics and parameterizations.

The geophysical mechanisms behind the Atlantic Ocean modes coupling with pressure and wind driving decadal precipitation anomalies were unraveled using observational data in Gorenstein et al. (2023). When examining the dynamics of this region in climate simulations, a more fundamental question arose regarding how to assess its decadal variability. Since numerical models present biased climate representations compared to observational data and among themselves, their climate variability is not typically measured in relation to large-scale climate patterns, such as ocean modes. This motivated us to create a new method using ocean modes and their precipitation counterparts to quantify decadal variability in numerical climate models.

Furthermore, the Discussion section is used to explore the broader applicability of the proposed methodology. There, we draw on previous proxy and numerical modeling studies to discuss how our information-theoretic definition of climate variability can be compared with traditional variance-based metrics used in the literature. In this context, it is necessary to address mid-Holocene-specific features and to assess whether our results are consistent with earlier findings for this period. While we have aimed to keep the manuscript direct and concise, removing this discussion would limit the paper's ability to evaluate the methodology in relation to existing studies and to illustrate its potential applications.

Other comments:

R1.7 You need to label each panel of a figure rather than collect them together (so Fig 1 has panel a-h, not just a-b).

The figures have been changed according to the reviewer's suggestions.

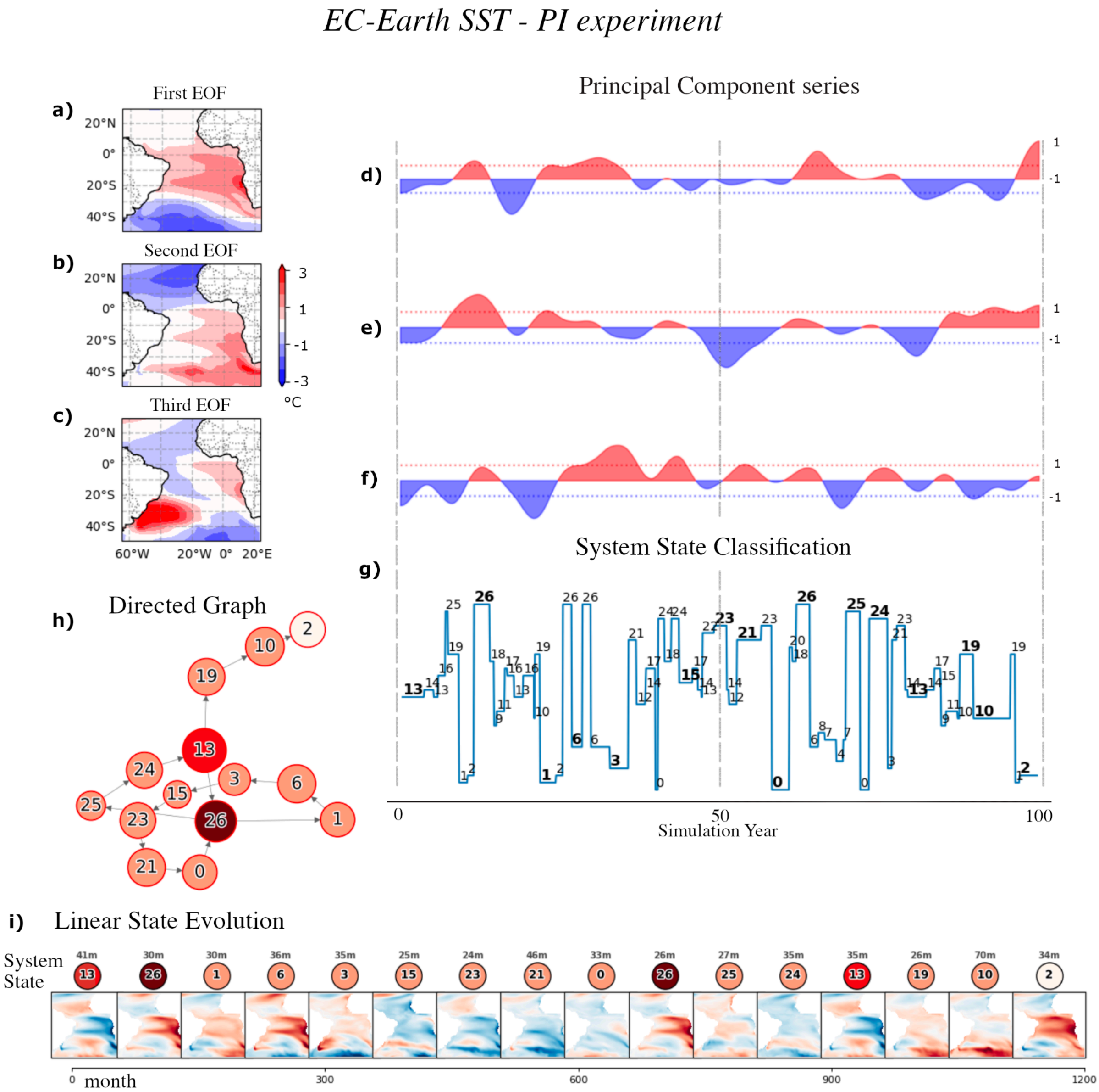
R1.8 Can you make the system-state classification centre around the neutral state of the 3 modes? This is the most intuitive way of thinking about them, but I don't see where that state is - presumably it is the most common mode in the directed-graph (but not necessarily).

The direct graphs can be controlled until a certain point. For example, it could be interesting to fix the most visited state in the center of the plot and the remaining states accordingly. However, in the revised version of the manuscript, state zero does not seem to be the most visited in the direct graphs, nor is there an evident commonality in the ensemble of simulations. We believe the previous version of the manuscript had Figures and graphs around state zero, as they were plotted using system discretizations with a fixed threshold (thresholds of 1 standard deviation) instead of the maximum entropy threshold.

Furthermore, the networkx library used to plot them is not very adaptable, and forcing a design can compromise the final graph visualization. It is not simple to represent densely connected graphs, maintaining connected nodes close to each other, and neither nodes nor edges overlapping. The graphs presented in the manuscript are a product of extensive testing using Python's library and are found to be feasible representations of the simulation trajectories.

To improve the graph readability, we introduced two new concepts in all the figures that contain graphs (Figures 1, 3-8, and 10). First, the most persistent states concept: the only states used to plot the graphs are those that show a minimum persistence of months; this way, the graphs become more focused and compact. Second, a frame-by-frame depiction of the most persistent system state's evolution, presenting the same information contained in the directed graphs in a linear format. For example, the new Figure 3 follows:

(line 265) “



EC-Earth SST 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 (red/blue 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.”

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:

“

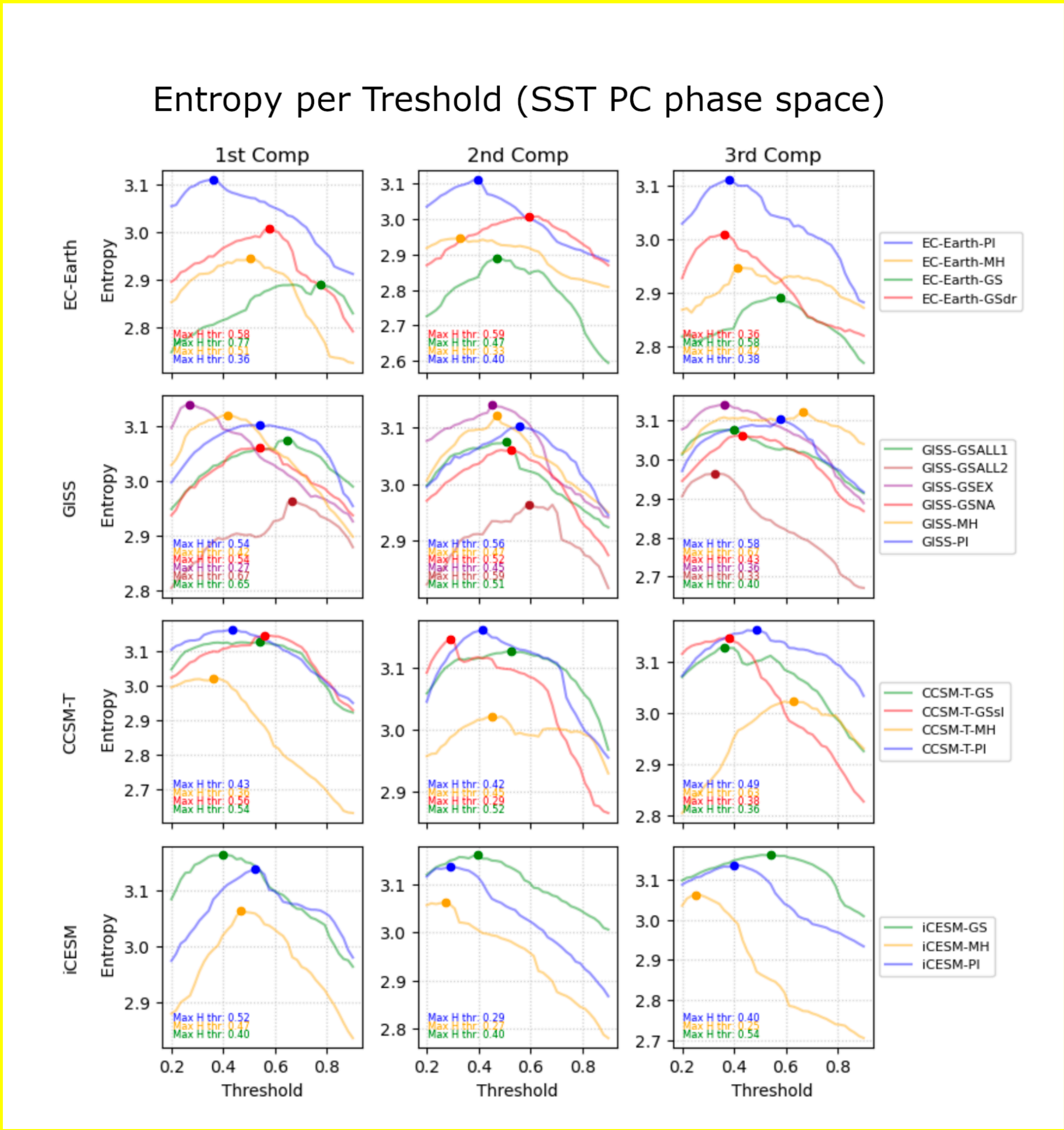


Figure S3. SST entropy values for each model scenario with different thresholds. The y-axis is the Entropy value, calculated from equation 1. The x-axis is the threshold used to define the positive, negative, and neutral phases of the three main components (a.k.a possible states) in the PCs' space. At 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)

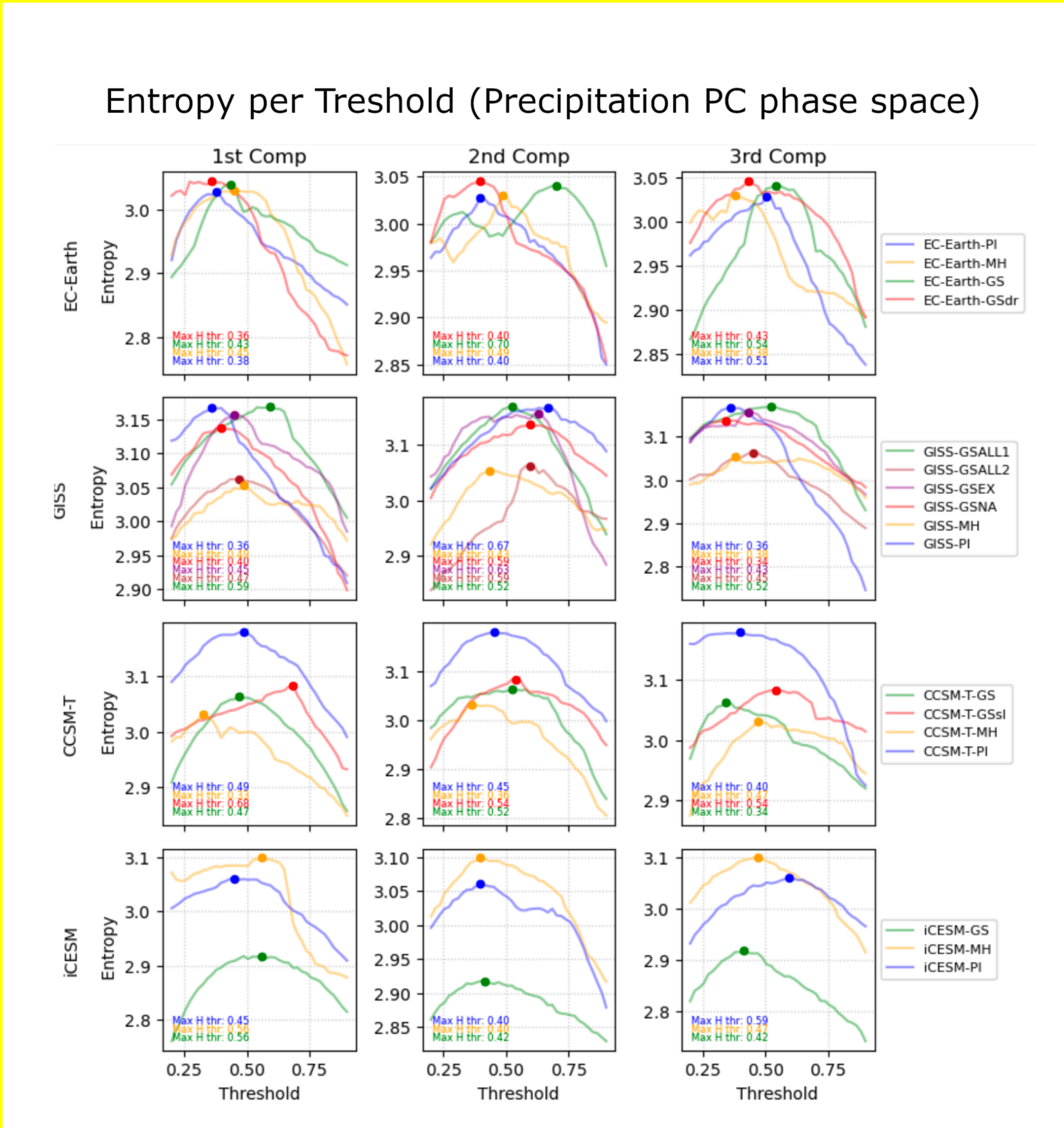
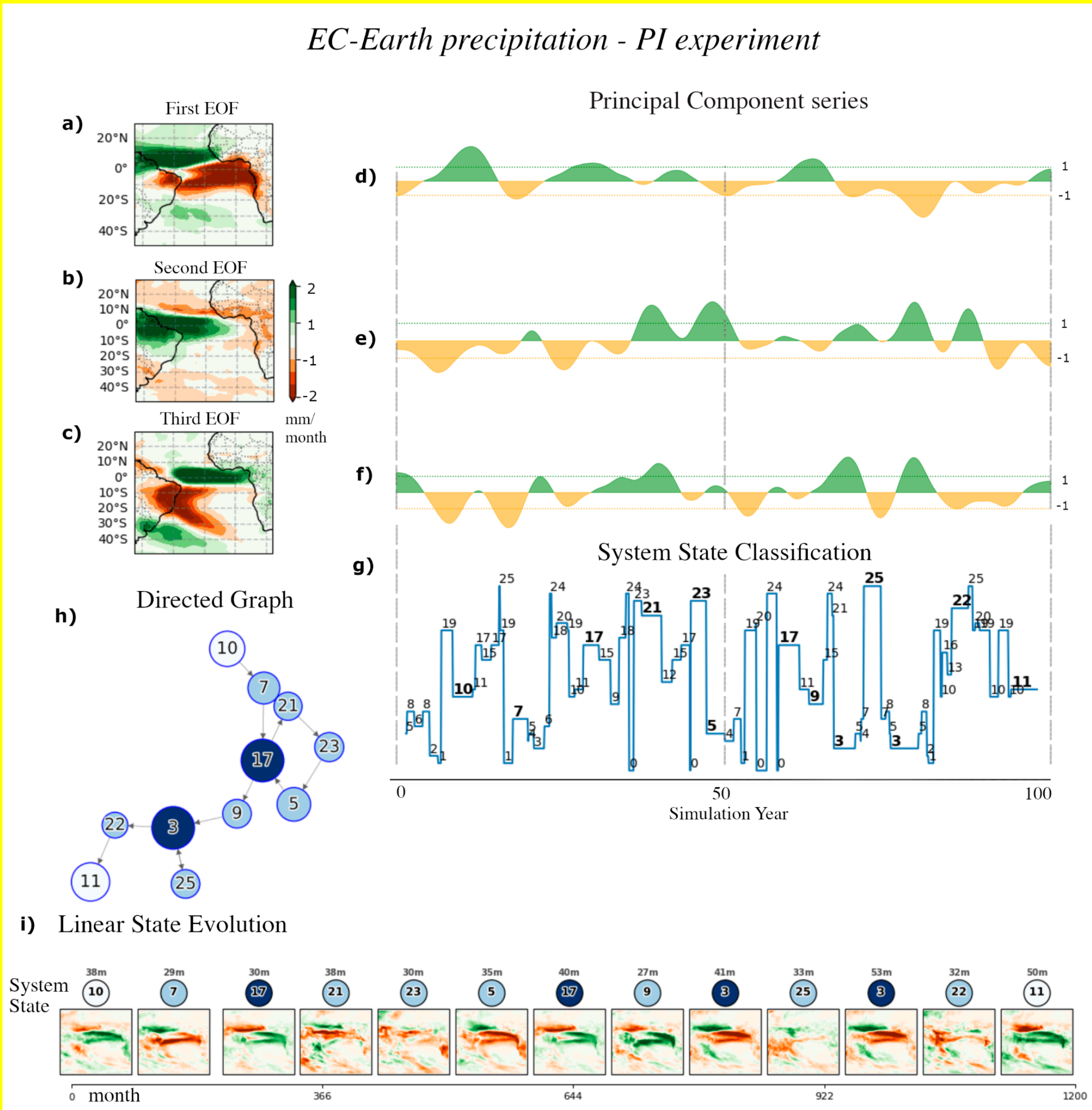


Figure S4. Precipitation entropy values for each model scenario with different thresholds. The y-axis is the Entropy value, calculated from equation 1. The x-axis is the threshold used to define the positive, negative, and neutral phases of the three main components (a.k.a possible states) in the PCs' 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)."