Supplementary Information for "A low-order data-driven model of ENSO diversity"

Cristian Martinez-Villalobos[1,2], Antonietta Capotondi[3,4], Clara Deser[5,6], Boris Dewitte[7,8,9], Neil J. Holbrook[10,11], Matthew Newman[4], Cécile Penland[4], Daniel J. Vimont[12], and Andrew T. Wittenberg[13]

[1]Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez
[2]Data Observatory Foundation, ANID Technology Center No. DO210001
[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[4]Physical Sciences Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
[5]NSF National Center for Atmospheric Research, Boulder, CO, USA
[6]Climate and Global Dynamics Laboratory, NCAR, 1850 Table Mesa Drive, Boulder, CO, USA
[7]Center for Ecology and Sustainable Management of Oceanic Islands (ESMOI), Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Chile
[8]Centro de Estudios Avanzados en Zonas Aridas (CEAZA), Coquimbo, Chile
[9]CECI, Université de Toulouse III, CERFACS/CNRS, Toulouse, France
[10]ARC Centre of Excellence for the Weather of the 21st Century, University of Tasmania, Hobart, Tasmania, Australia
[11]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
[12]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, USA
[13]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Contents of this file

- Text S1 to S3
- Table S1
- Figures S1 to S23

1 Text S1. Comparison with CMIP6 models

To contextualize the performance of the NG-LIM, we compare it against simulations from 30 global climate models participating in CMIP6: ACCESS-CM2, ACCESS-ESM1-5, BCC-CSM2-MR, CAMS-CSM1-0, CAS-ESM2-0, CESM2, CIEMS, CanESM5, CanESM5-1, CNRM-CM6-1-HR, CNRM-ESM2-1, EC-Earth3-CC, FGOALS-f3-L, FGOALS-g3, FIO-ESM-2-0, GFDL-CM4, GFDL-ESM4, GISS-E2-1-G, GISS-E2-1-H, HadGEM3-GC31-LL, IITM-ESM, INM-CM4-8, INM-CM5-0, KIOST-ESM, MCM-UA-1-0, MIROC-ES2L, MIROC6, MPI-ESM-1-2-LR, MRI-ESM2-0, TaiESM1, and UKESM1-0-LL. We used the same time period as observations (1948-2022), comprising the first available member of the historical simulation (1948-2014) and their corresponding SSP2-4.5 scenario simulation (2015-2022). Before analysis, each model was regridded to ERSST v5 resolution, and each model's SST anomalies were preprocessed identically to the observational data, including removal of the seasonal cycle using the first two harmonics and spatially local cubic detrending. Principal components were computed over the [20°S–20°N, 120°E–50°W] region, and the rotated EP and CP indices were defined following Takahashi et al. (2011), as in the main text. To correct for common degeneracies in EOF structure, especially the tendency of some models (7 out of 30) to simulate the analogous to the EOF2 mode in observations as the model EOF3 (e.g., Richter et al., 2022), we used pattern correlation with the observed PC2 regression map to identify the correct mode in each case.

Figure S20 presents the EP and CP SSTA patterns for observations, along with the LIM, NG-LIM, and the CMIP6 multi-model mean. As can be observed, both the standard LIM and NG-LIM show strong alignment with the observed structures. On the other hand, compared to observations and the LIM-based models, the CMIP6 EP and CP spatial patterns tend to be more spatially elongated, with broader east–west and north–south extensions. This elongation is not observed in the LIM or NG-LIM patterns and may reflect biases in model dynamics (e.g., Priya and Dommenget 2025). The variance explained by the first two EOFs in the LIM models is also more consistent with observations (both standard LIM and NG-LIM PC1 and PC2 explain about the same variance as observed; see Fig. S20 caption for details), in contrast to many CMIP6 models where the variance explained by each EOF tend to be underestimated, especially EOF2.

Figure S21a,b shows the SST variance patterns across the CMIP6 ensemble. While the multi-model mean is broadly consistent with expected ENSO-related variability, although with noticeable biases in the far eastern and western tropical Pacific in the multi-model mean, individual models differ substantially in both amplitude and zonal location of maximum variance. In contrast, NG-LIM reproduces the observed east Pacific maximum more reliably (Fig. S4, bottom row), further reinforcing its utility as a compact observational emulator.

Figure S22a shows the deviation from Gaussianity in the PC1–PC2 joint distribution for the CMIP6 multi-model mean. This can be compared with Fig. 2d,e in the main text, which shows the same structure for observations and NG-LIM. While CMIP6 models on average simulate a non-Gaussian joint structure reasonably well, the spread across models is substantial, and important deficiencies remain related to the representation of EP and CP patterns themselves (e.g., Fig. S20), as well as the amount of variance captured by the first two EOFs. These limitations affect how well the models capture key nonlinear features such as event asymmetry and skewness. The NG-LIM, in

contrast, reproduces the observed "boomerang"-shaped structure and its intensity with high fidelity and competitive skill (Fig. 2e), despite being based on a reduced-order linear framework.

We further quantified the curvature of this distribution using the α coefficient from a quadratic fit between PC2 and PC1. Across CMIP6 models, α values range from -0.42 to 0.03, with a median of -0.17, falling short of the observed $\alpha = -0.3$. The NG-LIM performs better in this regard, with a median α of -0.25 (Fig. 2c and Fig. S6). As a complementary metric, we computed the Kolmogorov–Smirnov (KS) distance between model-simulated and observed PDFs of the EP and CP indices. The smallest the value of this distance, the closer the simulated PDFs are to the respective observed ones (e.g., Martinez-Villalobos and Neelin 2021; Goldenson et al., 2021). The standard-LIM shows distances of 0.082 and 0.087 for the EP and CP indices, the CMIP6 multi-model mean scores of 0.067 and 0.056, and NG-LIM yields the best values: 0.027 and 0.046, respectively.

Beyond EP/CP statistics, we also examine how well the models capture the frequency and spatial structure of coastal El Niño and La Niña events, classified as ENSO-like and dipole-like following Martinez-Villalobos et al., 2024. Figures S8 and S9 show that the NG-LIM outperforms LIM and CMIP6 in capturing the observed spatial asymmetries of these events, particularly for the dipole-like type. CMIP6 models tend to misplace the coastal anomalies or produce overly symmetric patterns. Figure S10 shows that the NG-LIM also reproduces the observed asymmetry in event frequencies more accurately than LIM, while CMIP6 models not only fail to capture this asymmetry but also tend to simulate too few events overall. These results underscore NG-LIM's improved representation of coastal ENSO diversity compared to both the standard LIM and CMIP6.

Using NG-LIM as a CMIP6 model emulator

Beyond benchmarking, we trained separate NG-LIMs on the leading 10 principal components of each CMIP6 model. These emulated NG-LIMs provide a reduced-order summary of each model's internal ENSO dynamics and allow for process-level comparison across models. As can be seen in Fig. S22b, the NG-LIM emulated model versions capture the CMIP6 simulated deviations of Gaussianity of the joint PC1/PC2 PDFs as effectively as for the observed case (Fig. 2d,e), including the characteristic inverted U-shape of the simulated joint distribution. Similarly, Fig. S23 shows a scatterplot comparing the EP and CP skewness coefficients directly simulated by each CMIP6 model and by its NG-LIM emulator. The strong correlation across models confirms that the NG-LIM emulator captures much of the underlying asymmetry present in the GCMs, suggesting that these features are largely encoded in the leading variability modes. Deviations from the 1:1 line may reflect higher-order nonlinearities or stochastic processes not captured in the linear inverse framework.

In summary, the NG-LIM trained on observations provides a surprisingly competitive benchmark for ENSO diversity and asymmetry, often outperforming the CMIP6 multi-model mean despite using less information. Meanwhile, NG-LIM emulators offer a practical and interpretable tool for diagnosing the behavior of fully coupled models, opening the door to more targeted model evaluation strategies. We have also made available the 10,000yr NG-LIM emulations of the CMIP6 models at `https://doi.org/10.5281/zenodo.14775713`, to facilitate further research and model evaluation.

2 Text S2: Calculation of LIM deterministic and stochastic operators

A LIM can be written as (see equation 1 in the main text):

$$\frac{d\mathbf{x}}{dt} = \mathbf{M}\mathbf{x} + \mathbf{B}\eta. \tag{1}$$

The linear deterministic operator is calculated as:

$$\mathbf{M} = \frac{1}{\tau} log(\mathbf{C}_\tau \mathbf{C}_0^{-1}), \tag{2}$$

where $\mathbf{C}_t = \langle \mathbf{x}(t)\mathbf{x}(0)^T \rangle$ is the lag-covariance matrix at lag t ($\mathbf{C}_0$ is the contemporary covariance matrix). For our calculations we used a lag $\tau = 1$ month.

The noise amplitude matrix $\mathbf{B}$ is determined from the fluctuation-dissipation relationship (Penland and Matrosova 1994):

$$\mathbf{B}\mathbf{B}^T = -\mathbf{M}\mathbf{C}_0 - \mathbf{C}_0\mathbf{M}^T. \tag{3}$$

3 Text S3: Connection to Koopman Operator Theory

Both Koopman theory and the NG-LIM transform the variables of a non-Gaussian, possibly nonlinear, system into representations that can exploit the practical advantages of linear algebra. Koopman theory transforms a nonlinear dynamical system in terms of an infinite-dimensional linear operator acting on functions of the state of the system. Empirically estimated Koopman operators are based on the eigenstructure of a matrix involving a large number of lagged correlations, where "large" is assumed to be big enough to approximate the infinite-dimensional operator, yet small enough to be practical. The evolution of the system can then be described using a series expansion of these eigenvalues and eigenfunctions, with the series coefficients expressed in terms of the initial condition. Importantly, many recent AI-based approaches to nonlinear dynamical systems, including deep learning surrogates, are inspired by Koopman operator theory (Takeishi et al., 2017; Li et al., 2017; Lusch et al., 2018), as they seek to learn feature spaces in which the dynamics are more linear and tractable. However, in most climate applications, empirical Koopman approaches are implemented in a fully deterministic manner (e.g., Navarra et al., 2021; Nathaniel et al., 2025), and are rarely tractable for high-dimensional systems (e.g., Klus et al., 2018), motivating the need for simpler, interpretable alternatives that can accommodate stochastic dynamics.

NG-LIM considers the requirements of standard LIM and then tries to find a transformation of variables such that these requirements are better satisfied. Standard LIM assumes that a time series represents a linear dynamical system driven by Gaussian white noise. If that is the case, then the dynamical system itself obeys Gaussian statistics. Clearly, if the stationary probability of the system is non-Gaussian, the time-dependent transition probability describing the evolution of the system is unlikely to be Gaussian. In geophysical contexts such as ENSO, observed state variables often exhibit non-Gaussian distributions and other signatures of nonlinear dynamics, so a standard LIM may not fully capture their statistical structure or extremes. In NG-LIM, the state variables are subjected to a nonlinear transformation so that the stationary probability density of the transformed variables is Gaussian. The transformation used here (e.g., the Yeo-Johnson transformation) is a statistically principled and widely adopted approach, applied consistently to the full state vector, and is used to improve the Gaussianity of the system in the transformed space. These transformed variables are then subjected to a standard LIM. Then the usual tests for the validity of LIM can be applied in the transformed space. Assuming passage of these tests, the inverse transform can be applied for physical interpretation. This strategy enables the resulting model to better capture nonlinear, asymmetric, and tail behavior and higher-order moments observed in the original data, while preserving the transparency and low dimensionality of the LIM framework. By explicitly incorporating both deterministic and stochastic dynamics, NG-LIM provides a statistically robust and interpretable tool for simulating geophysical variability, offering a complementary alternative to more complex deterministic or AI-based approaches.

Supplementary Table S1. Summary of NG-LIM improvements relative to the standard LIM

Aspect	Limitation in standard LIM	Improvement in NG-LIM	Figures
Joint PDF structure (PC1–PC2 / EP–CP)	Nearly Gaussian, centered distribution; lacks observed "boomerang" (inverted-U) curvature.	Recovers curved joint PDF with clustering of EP El Niño and CP La Niña states; α median ≈ -0.25 (obs $= -0.30$).	Fig. 2a–e; S5–S7
EP/CP index distributions (tails + body)	Underestimates extreme EP warm and CP cold events; CDF deviates in both tails.	Matches observed tail probabilities and CDF shape for both indices.	Fig. 3c–f; S14
Goodness-of-fit (KS distance)	Larger distributional mismatch (KS: EP = 0.082, CP = 0.087).	Best match to observed PDFs (KS: EP = 0.027, CP = 0.046); lower than LIM and CMIP6.	Text S1; Fig. S14
Spatial skewness (ENSO asymmetry)	Warm/cold composites nearly symmetric; weak east–west skew.	Captures observed spatial antisymmetry pattern and regional skewness magnitudes.	Fig. 3a,b
Regional index skewness (Niño 1+2, 3, 4)	Incorrect sign/magnitude of skewness; overly symmetric indices.	Correct sign and magnitude across Niño regions; matches observed values.	Fig. S12
Coastal ENSO diversity and asymmetry	Overly symmetric coastal El Niño/La Niña frequencies.	Improved asymmetric occurrence of coastal ENSO-like and dipole-like events.	Fig. S10
Deterministic evolution asymmetry (optimals)	Warm/cold optimals evolve symmetrically in time and space.	El Niños grow faster and decay sooner than La Niñas; reproduces observed antisymmetric evolution.	Fig. 4a–g; S16–S17
Benchmark vs CMIP6 (diagnostic value)	CMIP6 shows elongated patterns, weaker variance, muted skewness.	NG-LIM reproduces higher-order features more faithfully than CMIP6 multi-model mean; efficient emulator.	Text S1 and figures referred therein

Summary. NG-LIM preserves the standard LIM's strengths on leading patterns/variance while substantially improving higher-order realism: non-Gaussian joint structure, tails, skewness/kurtosis, coastal diversity, and asymmetric evolution.

4 Supplementary Figures

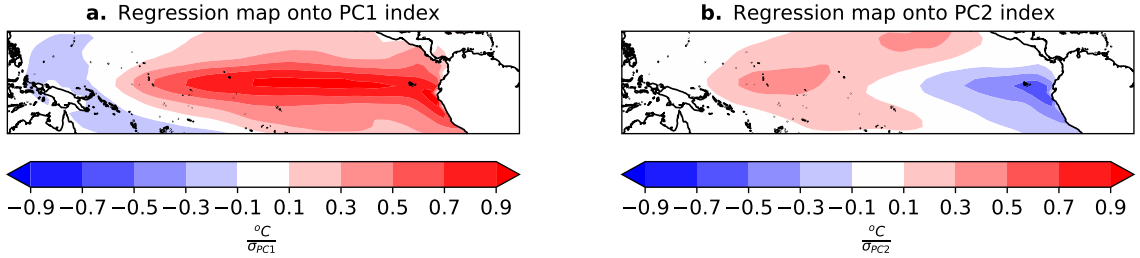

Figure S1: Spatial patterns of **a.** first EOF and **b.** second EOF of tropical Pacific SSTA. These are calculated as the SSTA regression pattern on PC1 and PC2 respectively.

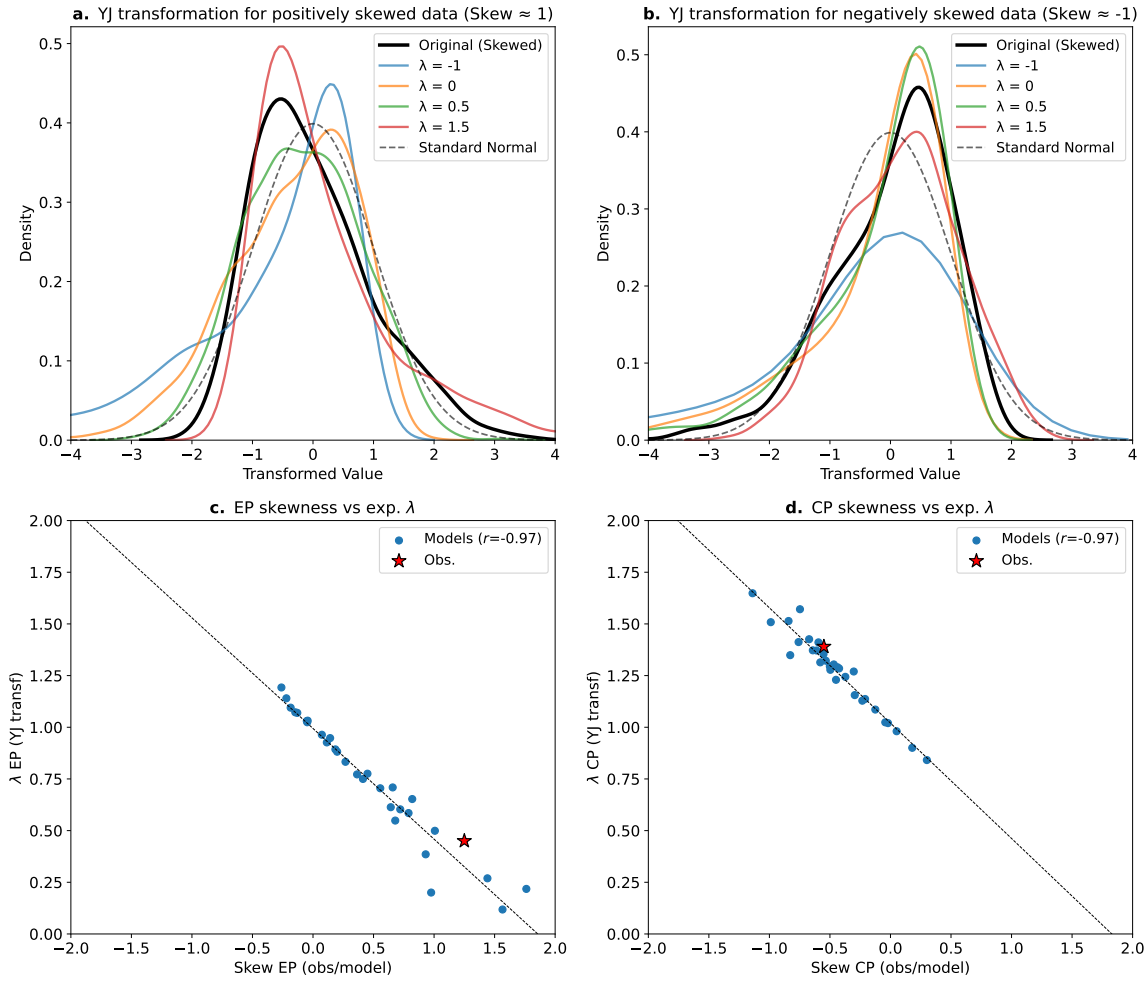

Figure S2: Examples of probability distributions of transformed data when applying the YJ transformation with exponents $\lambda = -1, 0, 0.5, 1.5$ on **a.** positively skewed data ($S \approx 1$) and **b.** negatively skewed data ($S \approx -1$) created synthetically. The original data has mean zero and standard deviation of 1. The dashed line shows a normal distribution for reference. Values $\lambda < 1$ compress the upper tail of positively skewed data, while $\lambda > 1$ compress the lower tail of negatively skewed data. The case $\lambda = 1$, corresponding to the identity transformation, is omitted. Panels **c** and **d** show the relationship between skewness and the optimal YJ exponent λ for the EP and CP indices, respectively, across the CMIP6 ensemble. Each point represents one model, and stars indicate the observed value. A strong negative linear relationship emerges in both indices, with $r \approx -0.97$ in both cases, indicating that models with larger positive skewness require smaller YJ exponents, and models with more negative skewness require larger exponents to normalize their distributions. Observed values fall close to the model-based regression line in each case, showing that the skewness-λ relationship found in the ensemble is consistent with observations.

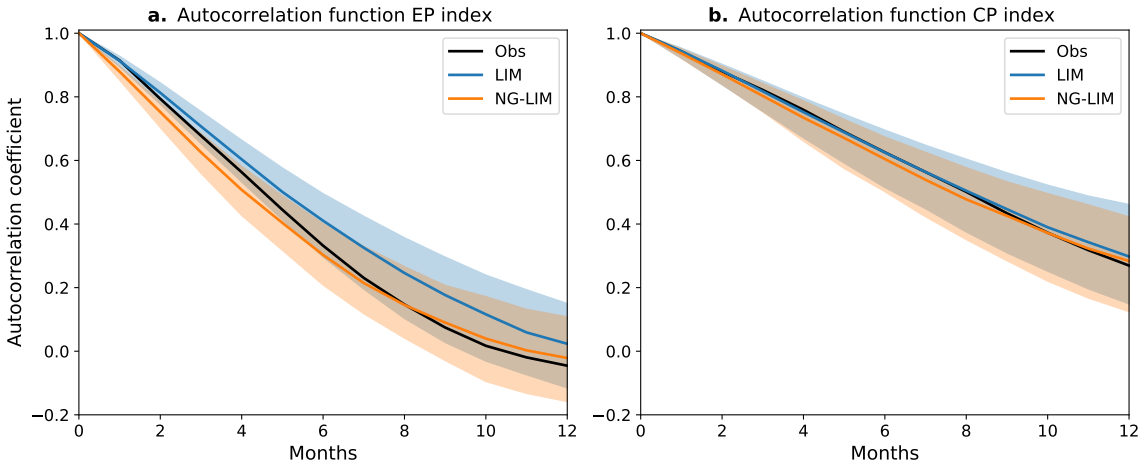

Figure S3: Autocorrelation function of **a.** EP and **b** CP indices in observations (black), median estimation generated by the standard LIM (blue) and generated by the NG-LIM (orange). The shading encompasses the 5th-95th percentiles of the autocorrelation function across epochs of the same length as observations.

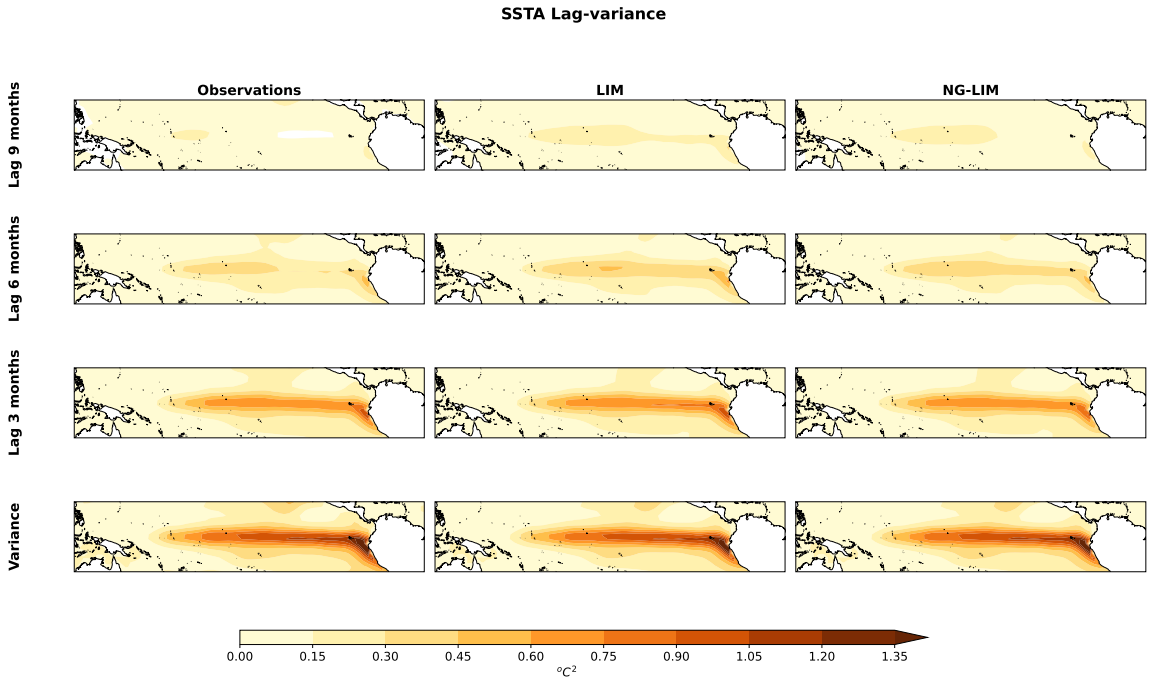

Figure S4: Tropical Pacific SSTA lag-variance spatial pattern in **a.** observations, **b.** generated by the standard LIM, and **c.** generated by the NG-LIM, for lags of 9, 6, 3 and 0 months.

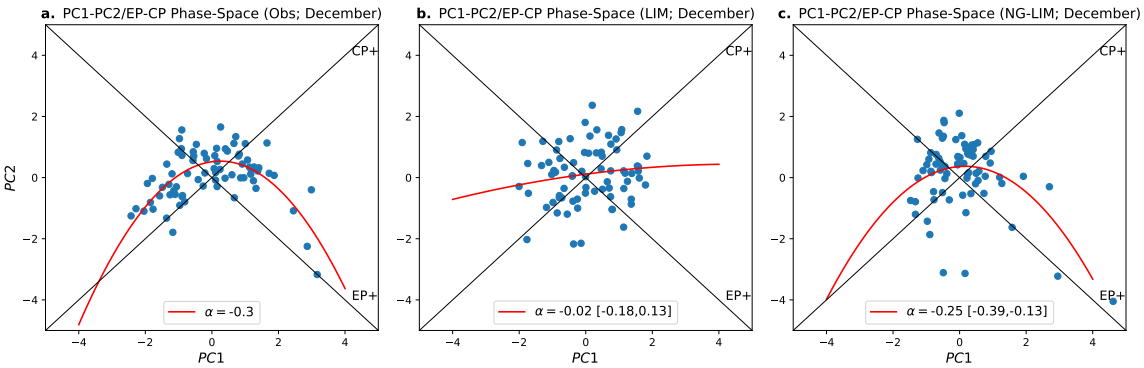

Figure S5: Same as Figs. **2a,b,c** in the main text but only for December.

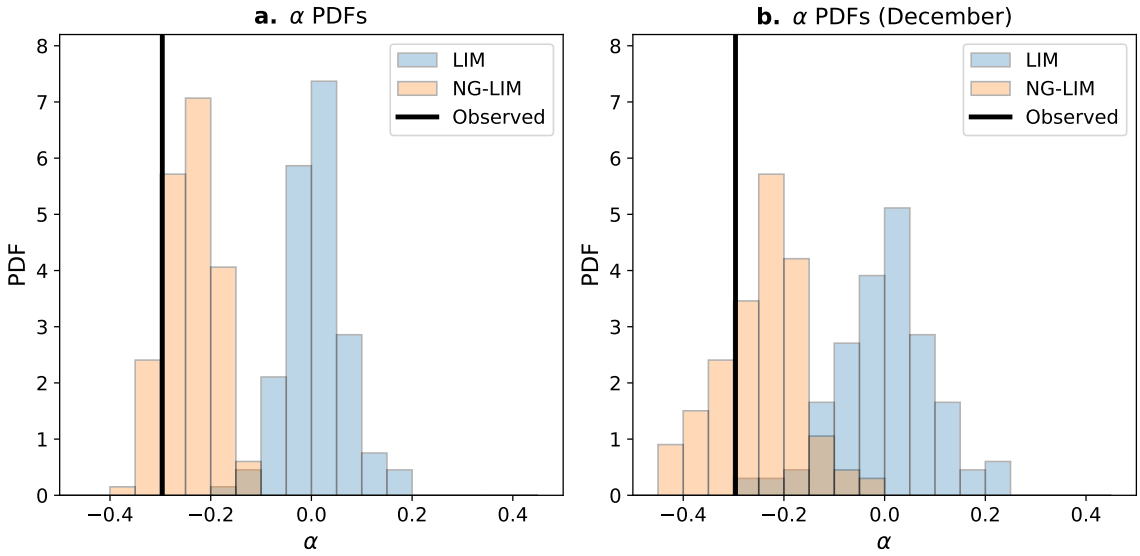

Figure S6: Probability distribution of estimated parameter α across epochs of same length as observations in the standard LIM (blue) and NG-LIM (orange). The black line shows the observed value. The estimations use **a.** annual data, and **b.** only Decembers.

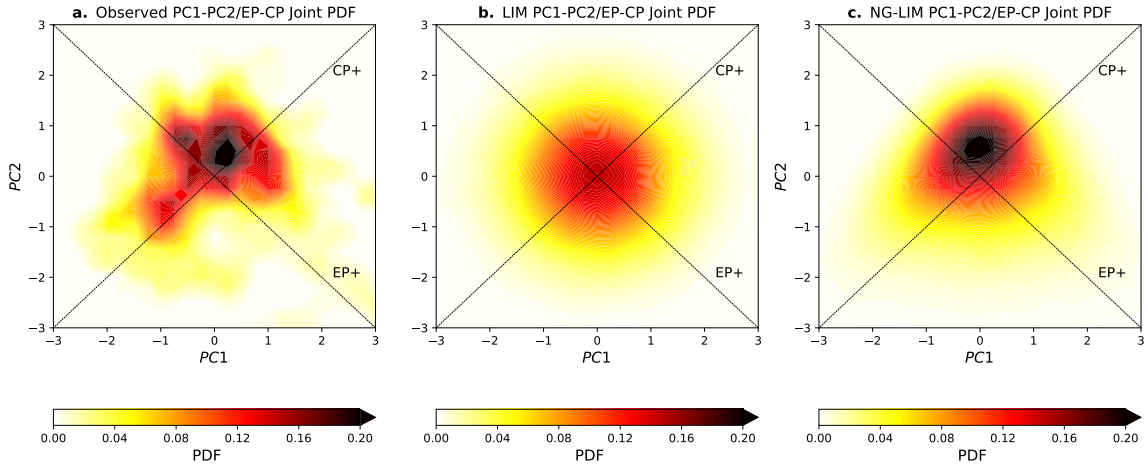

Figure S7: Joint probability distributions of normalized PC1 and PC2 indices in **a.** observations, **b.** standard LIM (whole integration), **c.** NG-LIM (whole integration). All these joint PDFs integrate to 1. The joint EP-CP indices PDF can be seen at the diagonals.

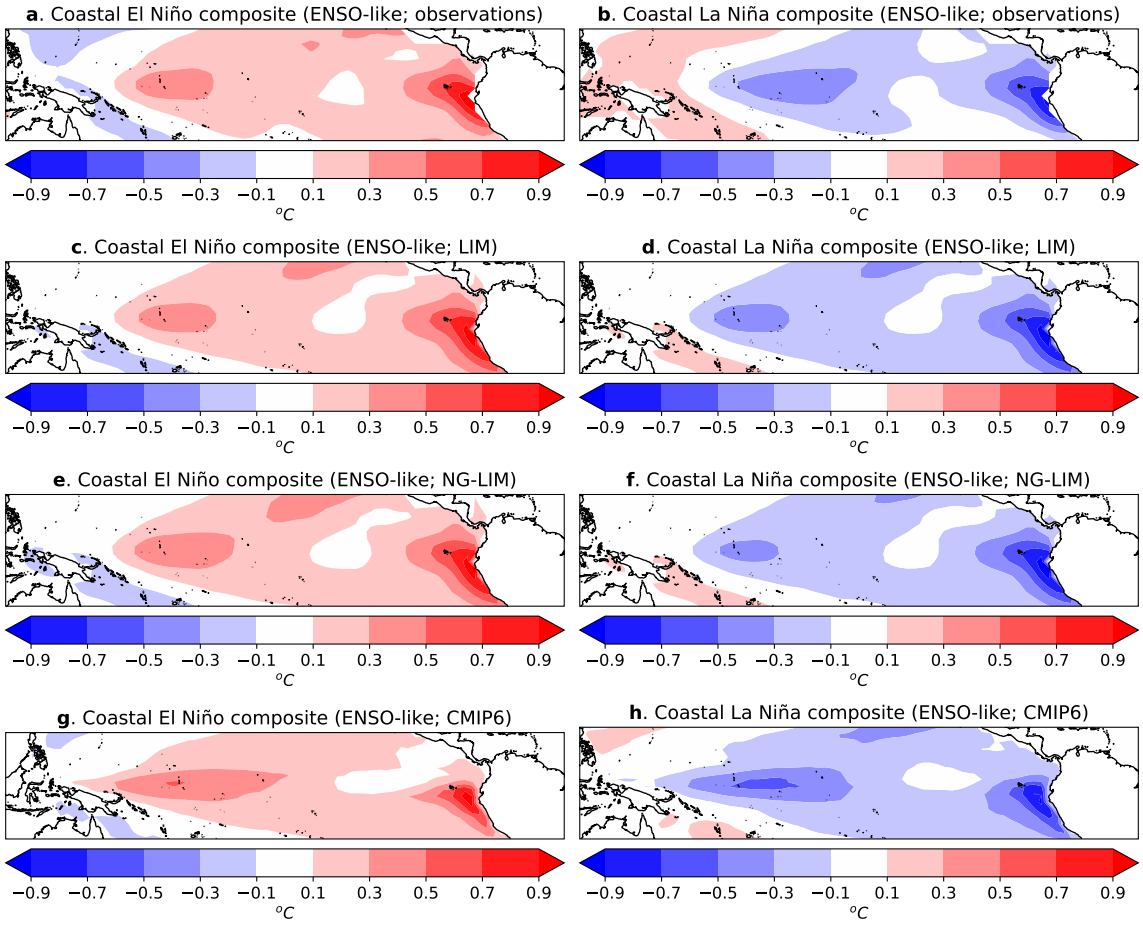

Figure S8: Sea surface temperature anomalies composites for ENSO-like coastal El Niño (left column) and coastal La Niña (right column) months, defined following Martinez-Villalobos et al., 2024, using threshold-based selection in the Niño 1+2 and Niño 3 regions and a PC2-based criterion (see Martinez-Villalobos et al., 2024 section "Coastal El Niño and La Niña events during the satellite era"). Panels show SST anomalies composited across all identified months in **a-b** observations, **c-d** the standard LIM simulation, **e-f** the NG-LIM simulation, and **g-h** the CMIP6 multi-model mean. Anomalies are shown in ^oC. The observed composites (**a-b**) exhibit strong warming or cooling concentrated along the South American coast, with reduced warming extending westward into the western equatorial Pacific, a characteristic of ENSO-like coastal events. The CMIP6 multi-model mean (**g-h**) performs worse than either LIM or NG-LIM, showcasing more concentrated warming in the east and a more extended westward pattern. Overall, NG-LIM slightly improves upon the standard LIM and outperforms CMIP6 in reproducing the observed spatial patterns associated with ENSO-like coastal events.

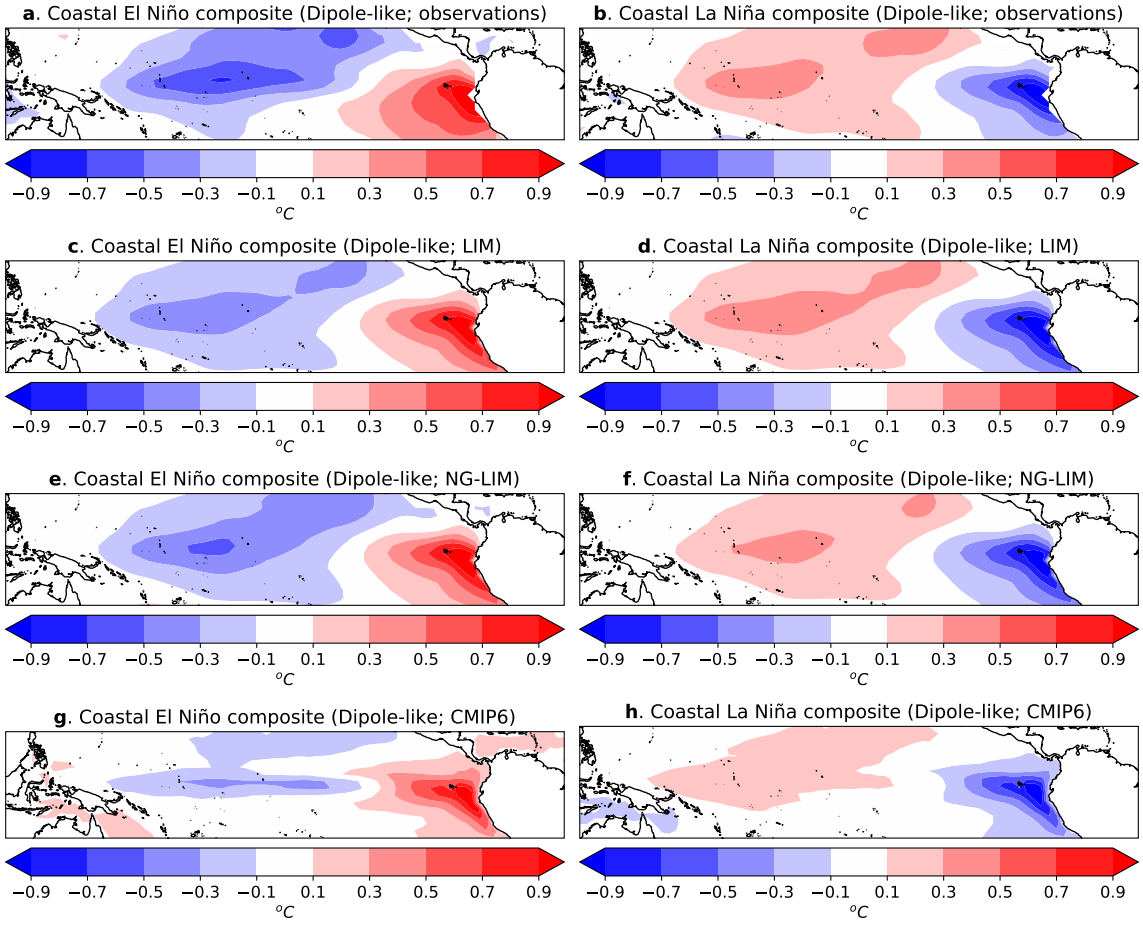

Figure S9: Sea surface temperature anomaly composites for dipole-like coastal El Niño (left column) and coastal La Niña (right column) months, defined following Martinez-Villalobos et al., 2024, using threshold-based selection in the Niño 1+2 and Niño 3 regions and a PC2-based criterion (see Martinez-Villalobos et al., 2024 section "Coastal El Niño and La Niña events during the satellite era"). Panels show SST anomalies composited across all identified months in **a-b** observations, **c-d** the standard LIM simulation, **e-f** the NG-LIM simulation, and **g-h** the CMIP6 multi-model mean. Anomalies are shown in ^oC. Observed dipole-like events (**a-b**) feature a distinct SST dipole: strong warming or cooling along the coast of South America accompanied by anomalies of opposite sign in the central equatorial Pacific. The NG-LIM model (**e-f**) captures this dipolar structure and its asymmetry between El Niño and La Niña phases better than the standard LIM (**c-d**), which generates a symmetric pattern between coastal El Niño and La Niña events. The CMIP6 multi-model mean (**g-h**) performs worse than both the standard LIM and NG-LIM, displaying poorly placed anomalies outside the coastal area. Compared to both LIM and CMIP6, the NG-LIM offers an improvement in capturing the spatial pattern, and asymmetry of dipole-like coastal events as observed

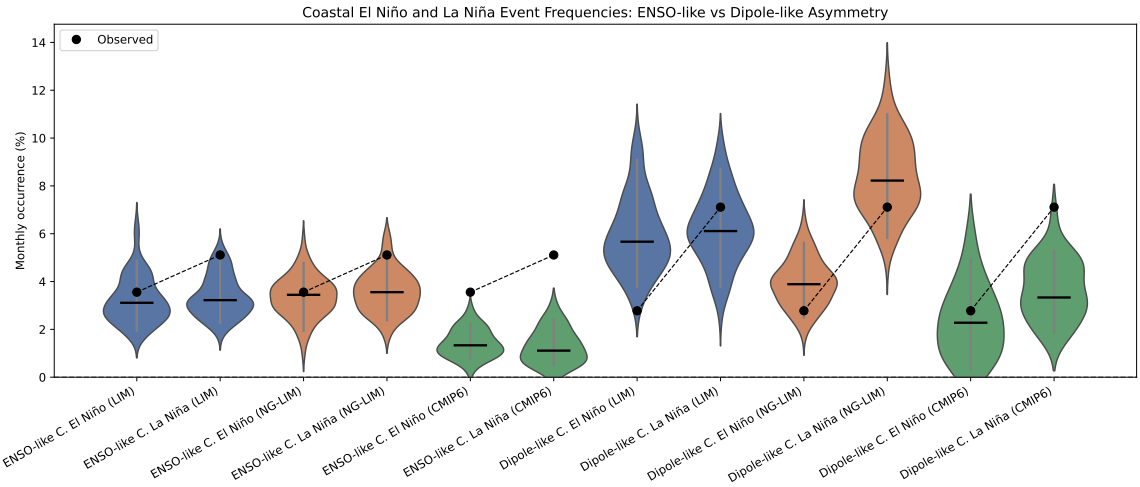

Figure S10: Monthly occurrence frequencies (in %) of coastal El Niño and La Niña events, classified as ENSO-like and dipole-like following Martinez-Villalobos et al., 2024 (see their section "Coastal El Niño and La Niña events during the satellite era"). Violin plots show distributions of frequencies from (1) standard LIM simulations (blue), (2) NG-LIM simulations (orange), and (3) 30 CMIP6 models (green), evaluated over non-overlapping 75-year periods. Observed frequencies are shown as black circles, and dashed lines connect the corresponding coastal El Niño and La Niña values for each type. Vertical gray bars indicate the 5th-95th percentile range, and horizontal black lines mark the medians. Observations show marked asymmetries in the frequency of coastal El Niño and La Niña events, particularly for dipole-like events, which occur more often in their cold (La Niña) phase. LIM simulations do not capture these asymmetries, producing nearly symmetric event frequencies across phases. NG-LIM improved upon LIM by reproducing both the magnitude and direction of the observed asymmetries more accurately, especially for dipole-like events, and consistently encompasses the observed values within its ensemble spread. CMIP6 models not only exhibit large inter-model spread but also tend to simulate fewer coastal events overall than observed, a deficiency also noted in Martinez-Villalobos et al., 2024. These results highlight the ability of NG-LIM to better capture the observed diversity and asymmetry of coastal event frequency, while maintaining the interpretability and efficiency of linear inverse modeling frameworks.

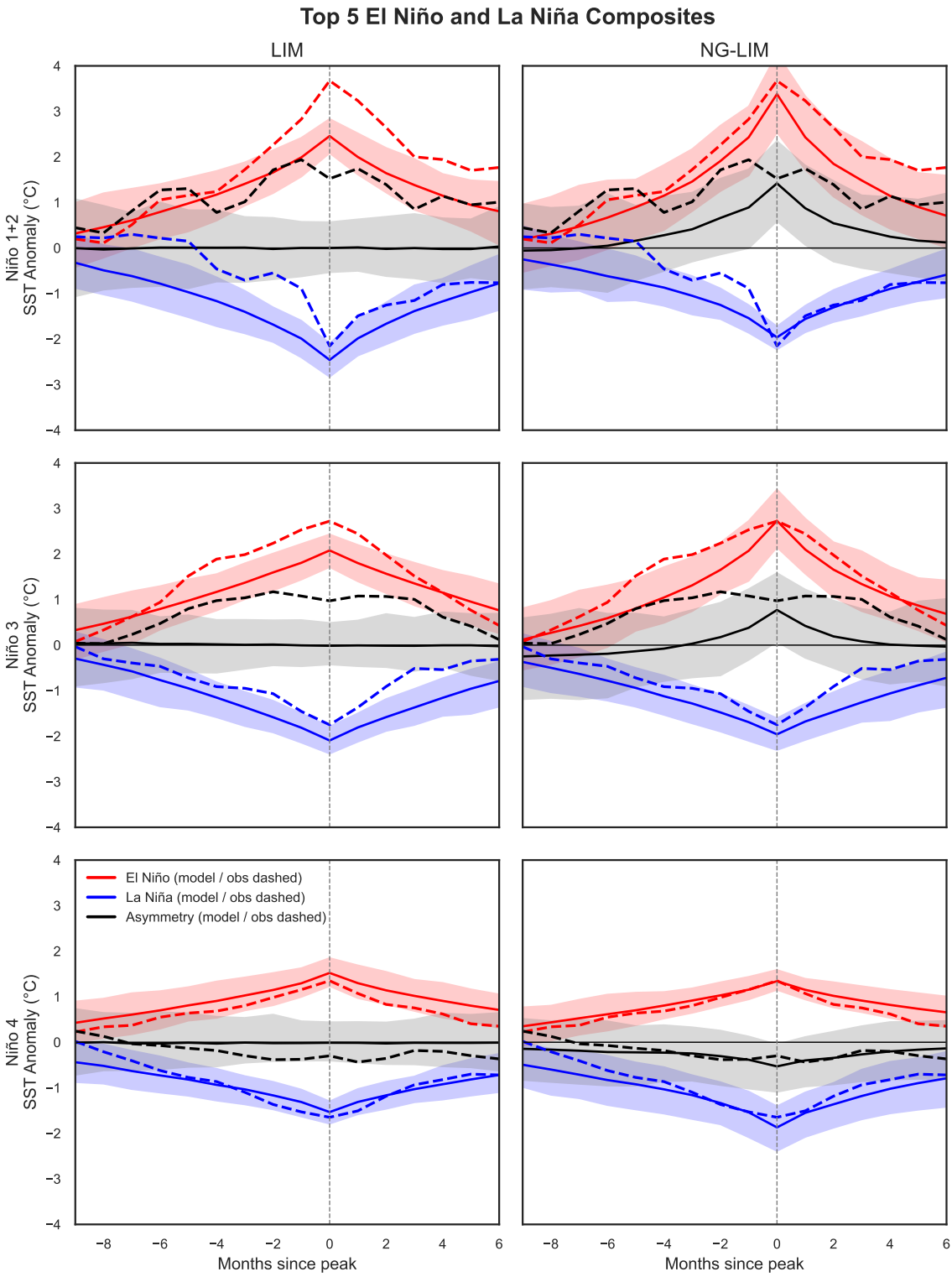

Figure S11: Composite evolution of the top 5 El Niño and La Niña events in Niño 1+2, Niño 3, and Niño 4 regions, based on 75-year epochs from the standard LIM (left) and NG-LIM (right). Composites are peak-anchored by the maximum or minimum monthly value of each Niño index within each epoch. Solid lines show the ensemble mean evolution across epochs for El Niño (red), La Niña (blue), and their sum (black), while dashed lines show the observed composites. Shading indicates the 5th–95th percentile range across epochs for each composite. Observed asymmetries—stronger and faster-growing El Niño events in the east, and stronger La Niña events in Niño 4—are captured by NG-LIM but not by the standard LIM.

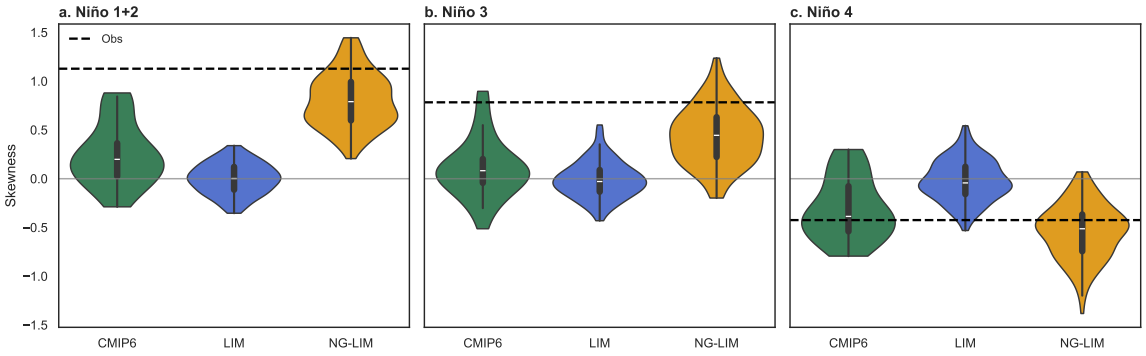

Figure S12: Distribution of Niño 1+2, Niño 3, and Niño 4 index skewness across CMIP6 models (green), the standard LIM (blue), and NG-LIM (orange). Each violin shows the distribution across either ensemble members (LIM and NG-LIM) or models (CMIP6), with the inner box indicating the interquartile range. Dashed black lines mark the observed skewness values in each region. LIM and NG-LIM realizations are based on 133 non-overlapping 75-year epochs, matching the observational record length. NG-LIM more accurately captures the positive skewness in the eastern Pacific and the negative skewness in Niño 4, in contrast to the nearly symmetric output from LIM and the broader spread of CMIP6 models. The CMIP6 multi-model mean also underestimates the magnitude of skewness in all three regions.

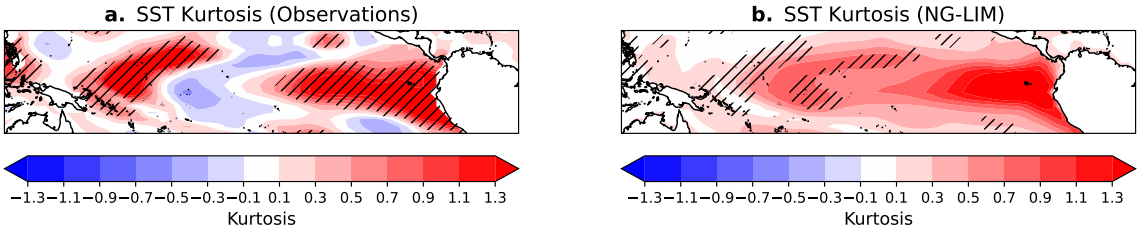

Figure S13: Same as Figures 3a,b in the main text but showing excess kurtosis. Excess kurtosis K is defined as $K(x) = \frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} - 3$. Typically, distributions with $K > 0$ ($K < 0$) have heavier (lighter) tails than Gaussian. The cross-hatching in **a** (**b**) shows regions where the observed SSTA excess kurtosis is outside the 5th-95th percentile range of excess kurtosis generated by the standard LIM (NG-LIM) across epochs

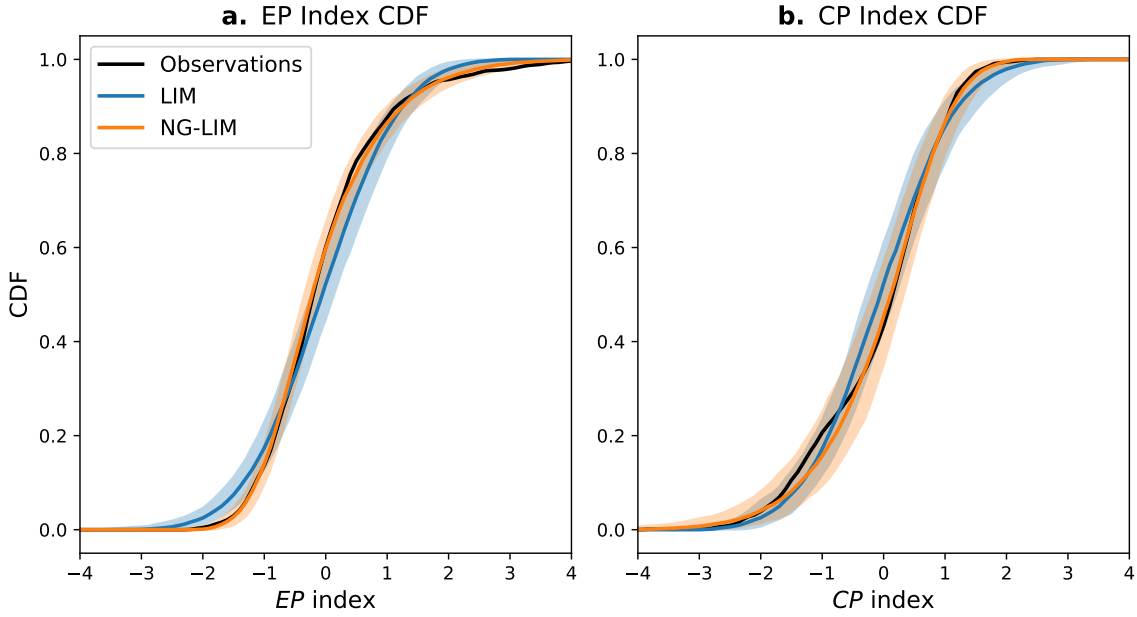

Figure S14: **a.** (**b.**) Cumulative distribution function (CDF) of the EP (CP) index in observations (black), median estimation generated by the standard LIM (blue) and generated by the NG-LIM (orange). The shading encompasses the 5th-95th percentiles of the CDF estimation across epochs of the same length of observations.

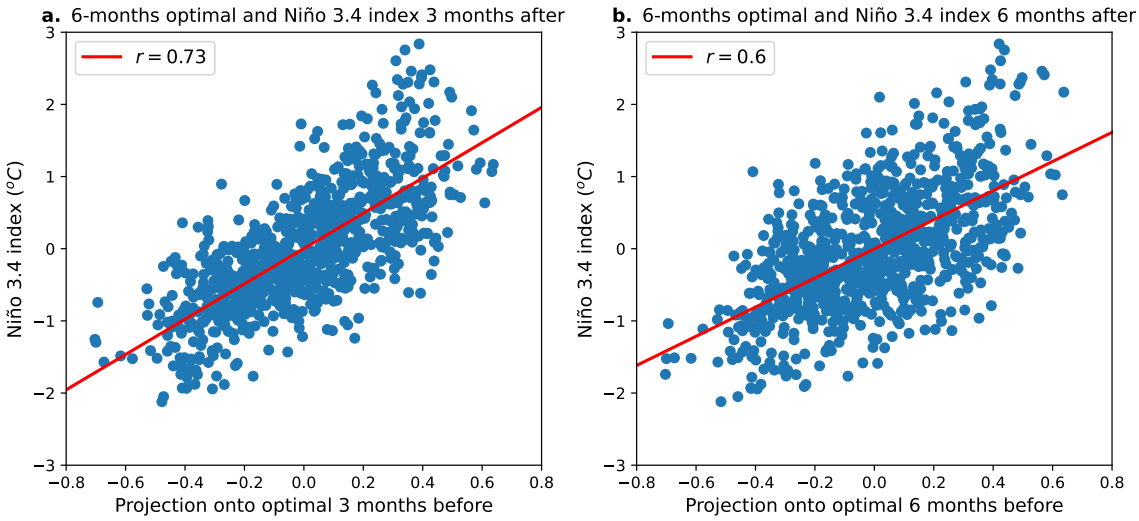

Figure S15: Scatter plot of the projection of SSTA onto the warm optimal pattern and the Niño 3.4 index **a** 3-months later and **b** 6-months later. The optimal projection ranges from -1 to 1, with 1 implying that the SSTA look exactly as in Fig. 4a, up to a rescaling factor.

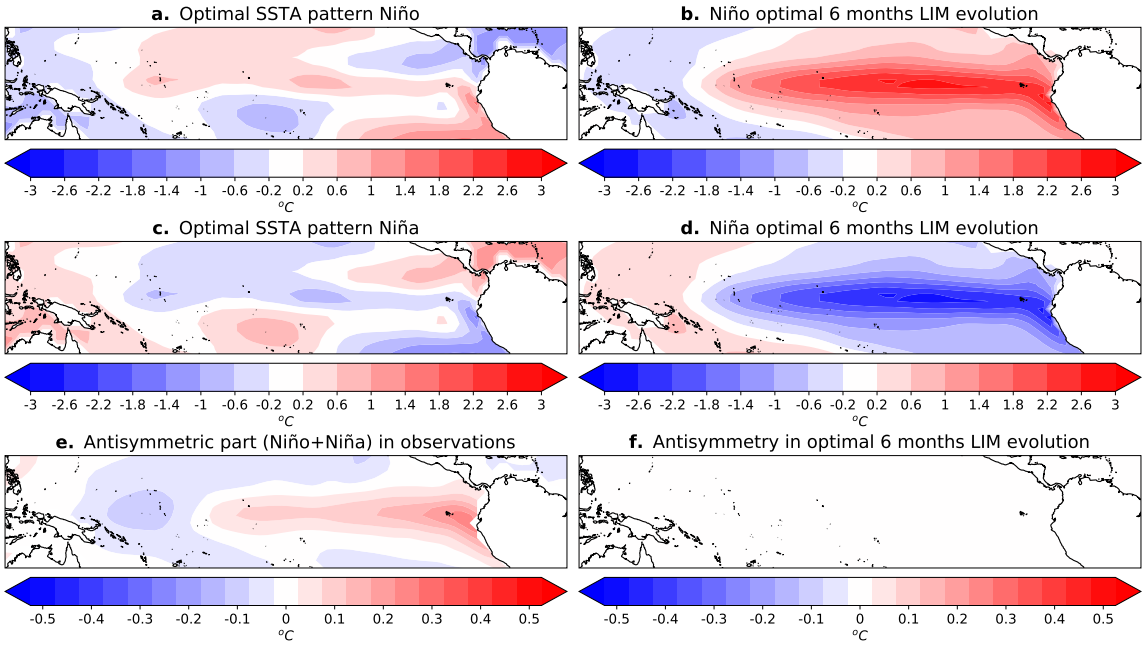

Figure S16: Same as panels 4a to 4f in the main text, but for the standard LIM. Note that as expected, there is no spatial asymmetry in the evolution of the warm and cold optimals in this case (panel f).

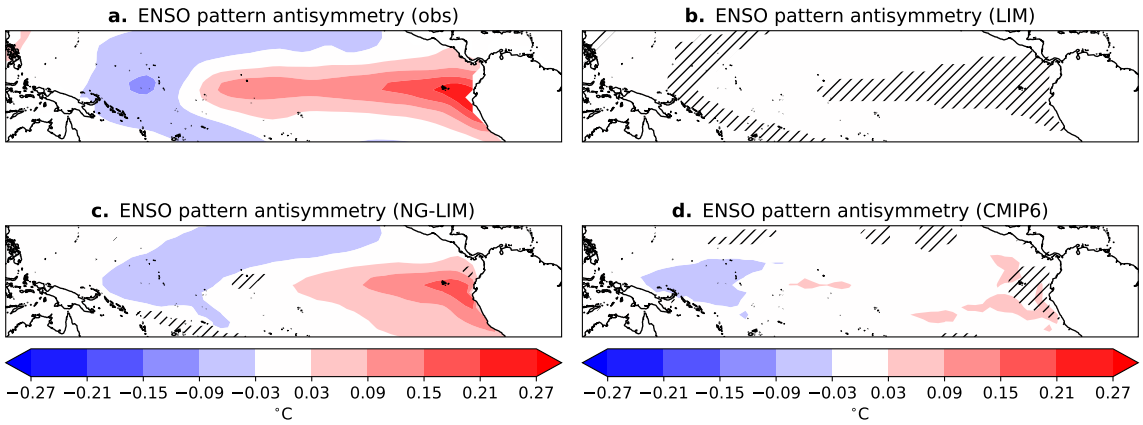

Figure S17: The figure shows the antisymmetric component of ENSO, calculated as the sum of El Niño and La Niña composite SST anomalies, where events are defined using the Niño 3 index ($\geq 0.5^oC$ for El Niño, $\leq -0.5^oC$ for La Niña). This metric emphasizes spatial differences in warm and cold events, including the east–west contrast and stronger coastal anomalies in the east during El Niño. Panel **a** shows the observed antisymmetry pattern. The NG-LIM (**c**) captures both the amplitude and spatial structure of this pattern, including enhanced eastern Pacific warming and westward cold anomalies, more realistically than the standard LIM (**b**), which generates a symmetric pattern by construction. CMIP6 models (multi-model mean; **d**) show some antisymmetric signal, but their ensemble diversity leads to spatial averaging and a weaker, less coherent structure. Cross-hatching in panels (**b-d**) indicates regions where the observed antisymmetry lies outside the 5th–95th percentile range of the ensemble generated by each model (i.e., across epochs of the same length as in observations in LIM and NG-LIM, or group of models in CMIP6). The extensive hatching in the standard LIM confirms its inability to simulate this feature. The NG-LIM shows much less hatching, supporting its improved realism. In contrast, the relatively modest hatching in CMIP6 does not imply better performance, but rather reflects the ensemble's large inter-model spread.

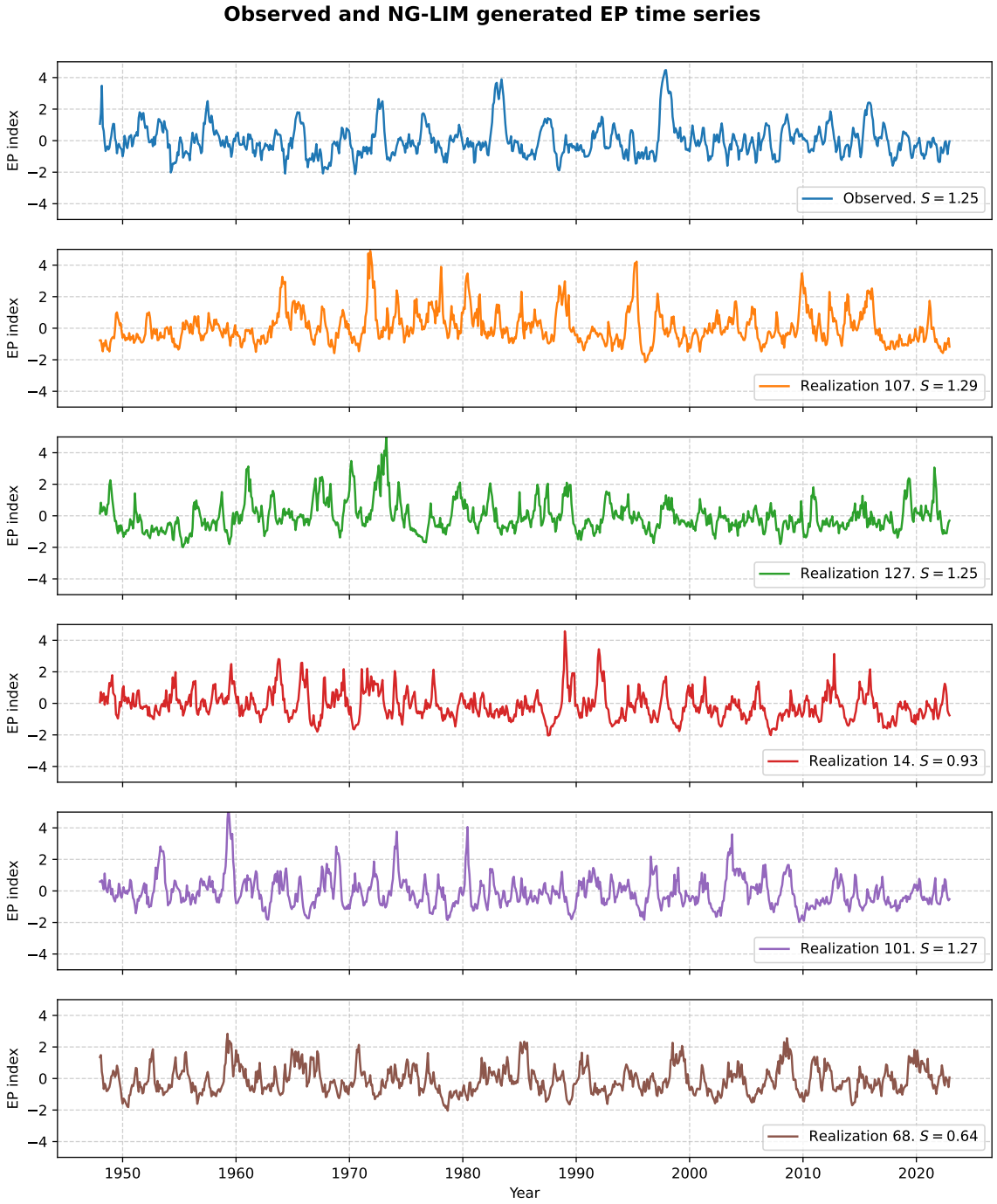

Figure S18: Time series of the observed EP index (top panel) and examples of randomly selected EP realizations generated by the NG-LIM. The legend indicates which realization (out of 133) is shown, along with its corresponding EP skewness coefficient, S

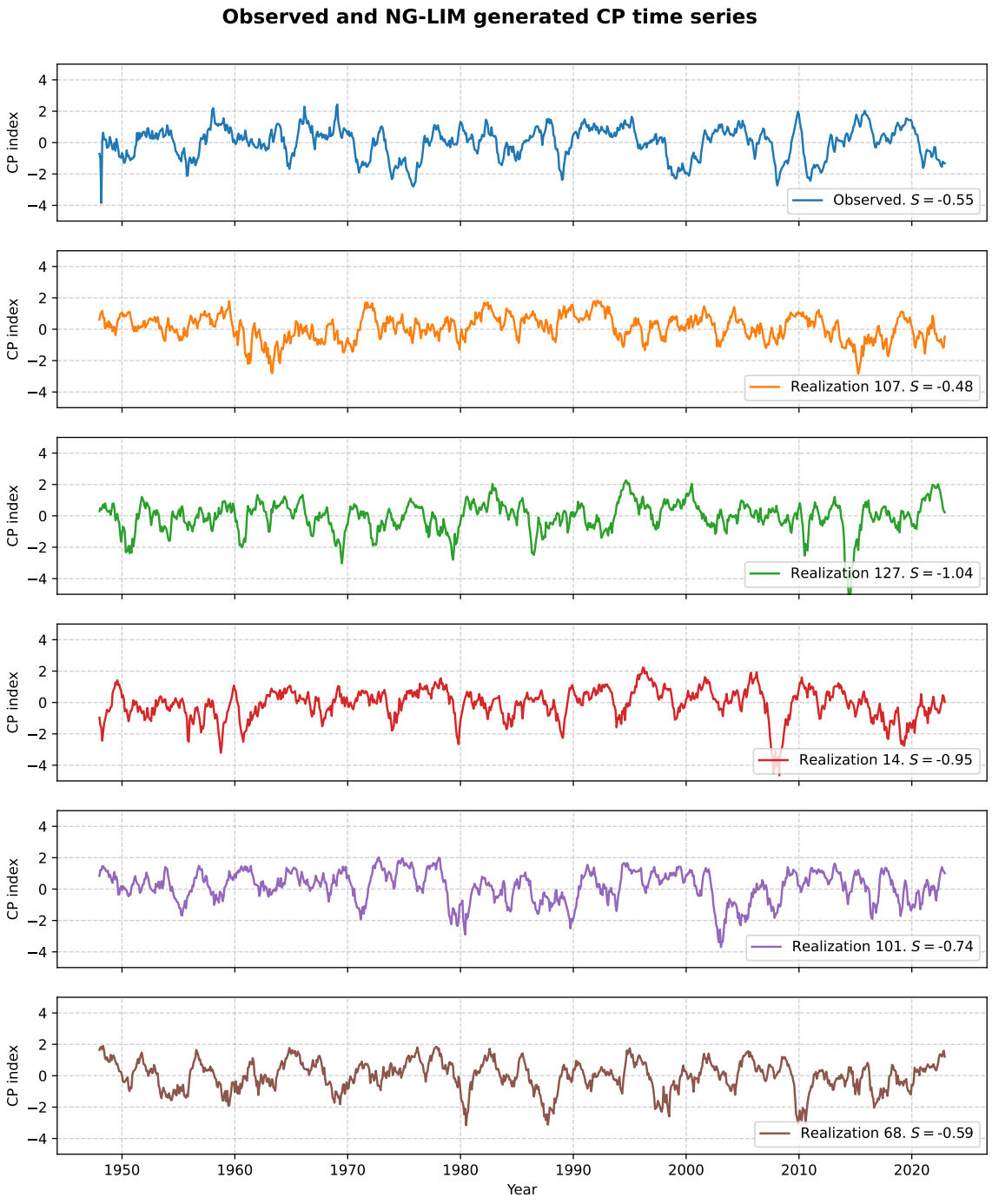

Figure S19: Time series of the observed CP index (top panel) and examples of randomly selected CP realizations generated by the NG-LIM (from same epochs as for the EP index in Fig. S18). The legend indicates which realization (out of 133) is shown, along with its corresponding CP skewness coefficient, S

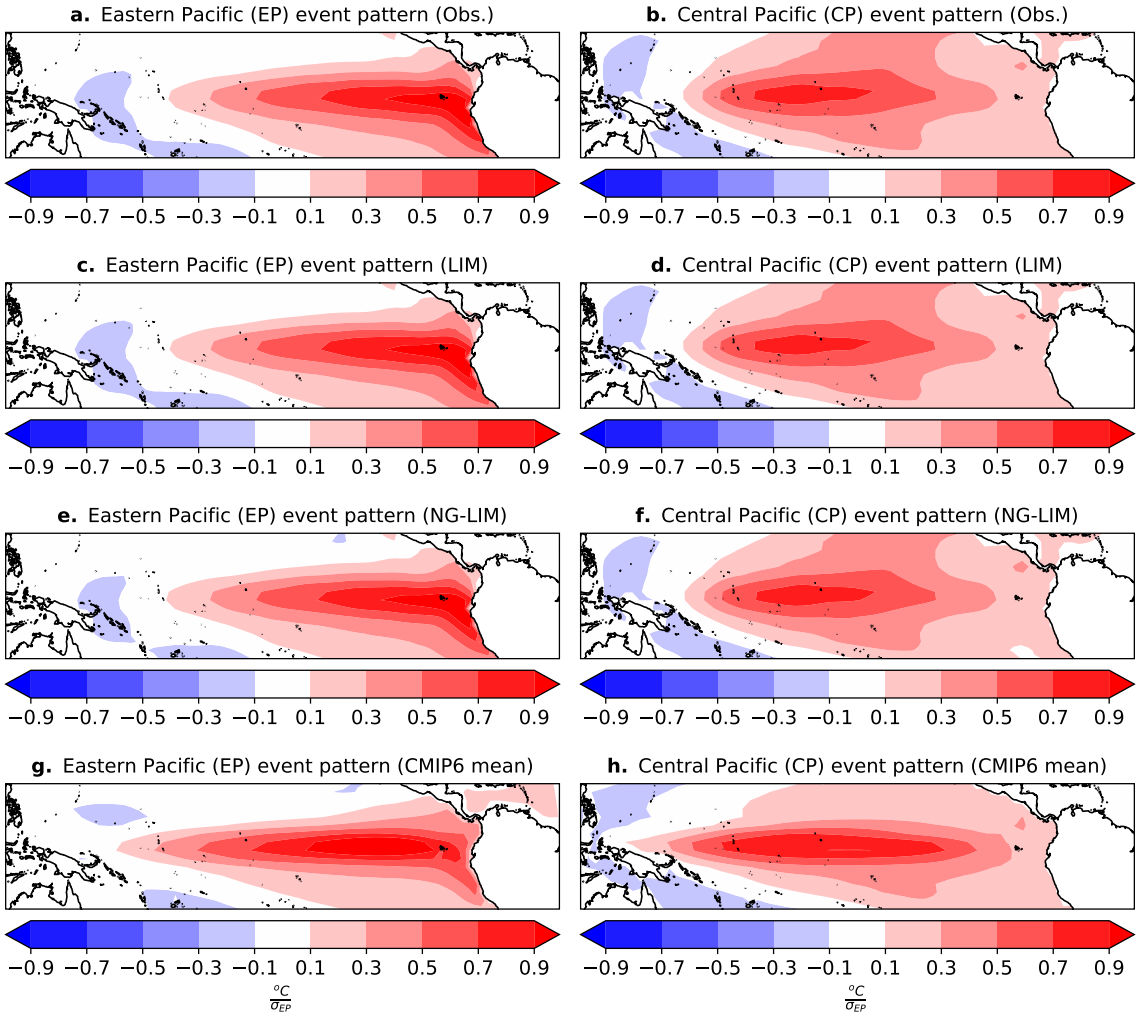

Figure S20: Eastern Pacific (left panels) and Central Pacific (right panels) spatial patterns in observations (NOAA ERSST v5; 1948-2022) (panels **a**, **b**), generated by the standard LIM (panels **c**, **d**), NG-LIM (panels **e**, **f**), as well as the CMIP6 multi-model mean (panels **g**, **h**). In all these cases, the patterns are calculated as the regression map onto the corresponding EP and CP indices. These indices are constructed as a function of the first and second principal components (PC1 and PC2) of tropical Pacific SST anomalies (see section 2.2.1 in the main text). Panels **a** and **b** are exactly the same panels shown in Fig. 1a and 1b in the main text, and are repeated here to facilitate comparison with the corresponding patterns simulated by the LIM, NG-LIM and CMIP6 multi-model mean. We note that both standard LIM and NG-LIM capture the spatial patterns about equally well. In observations, PC1 and PC2, and their associated EOF1 and EOF2 patterns, explain 55.83% and 12.19% of Tropical Pacific SST anomalies variance. In the standard LIM these figures are 55.37% and 12.27%, in the NG-LIM they are 55.55% and 12.19%, and in the CMIP6 multi-model mean they are 51.3% and 7.4% (on average), respectively. We observe that the patterns in CMIP6 are on average more elongated and have a warm pool bias.

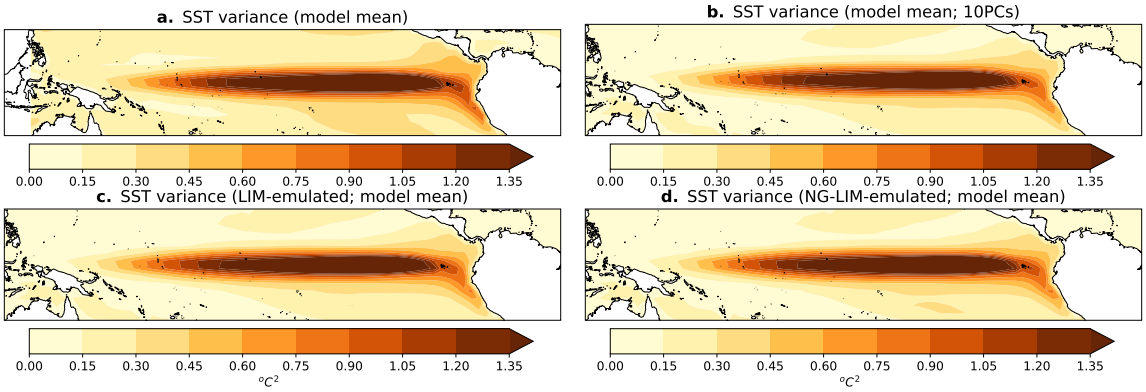

Figure S21: Multi-model mean of simulated **a.** SSTA variance in CMIP6. **b.** Same as **a.** but only taking into account SSTA captured by the first 10 EOFs (multi-model mean of 79% of variance, compared with 90% in ERSST v5). **c.** Same as **b** but for the multi-model mean of CMIP6 models SSTA variance emulated by the standard LIM. **d** Same as **c** but emulated by the NG-LIM. Comparing with Fig. S4 (bottom panels), we can appreciate a warm pool and a cold-tongue bias in CMIP6 simulated variance.

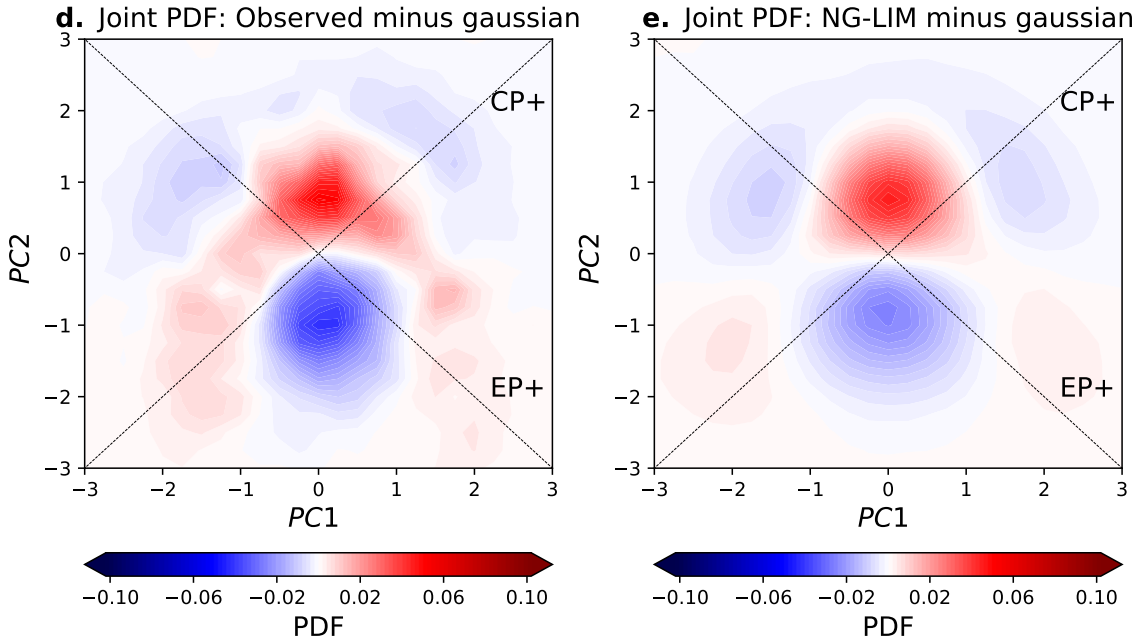

Figure S22: **a.** Multi-model mean of the deviation of joint-Gaussianity of PC1-PC2 (EP-CP) probability distribution simulated by 30 CMIP6 models. This can be compared to Fig. 2d in the main text showing the observed deviation of joint-Gaussianity of PC1 and PC2. **b.** Same as **a** but now showing the multi-model mean deviation of Gaussianity emulated by the NG-LIM applied to CMIP6 models. Important not to confuse with Fig. 2e in the main text, which shows the simulated deviation of gaussianity by the NG-LIM when applied to observations.

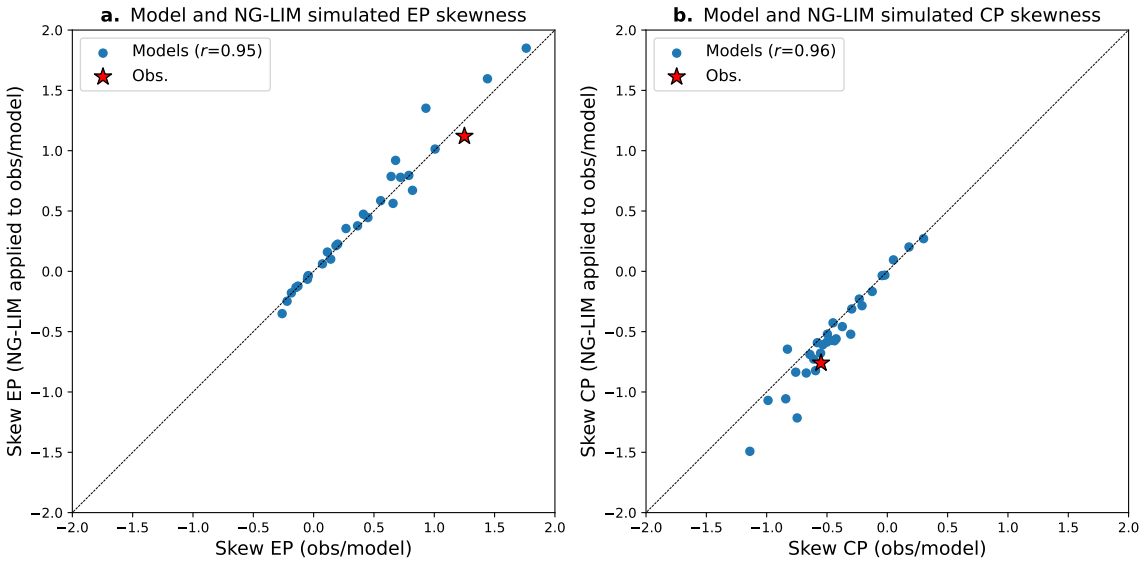

Figure S23: Scatter plot of the skewness coefficient of the **a.** EP and **b.** CP indices, simulated by CMIP6 models and in observations (x-axis) and the NG-LIM applied to each CMIP6 model and to observations (y-axis). Here, each dot corresponds to one model, and the red star marks the observed value. The diagonal is a 1-1 line, and the corresponding legend shows the correlation coefficient between the CMIP6 and NG-LIM emulated EP and CP indices skewness coefficient simulated across models.

5 References

Goldenson, N., Thackeray, C. W., Hall, A. D., Swain, D. L., and Berg, N. (2021). Using large ensembles to identify regions of systematic biases in moderate-to-heavy daily precipitation. Geophysical Research Letters, 48(9), e2020GL092026.

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., and Noé, F. (2018). Data-driven model reduction and transfer operator approximation. Journal of Nonlinear Science, 28(3), 985-1010.

Li, Q., Dietrich, F., Bollt, E. M., and Kevrekidis, I. G. (2017). Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(10).

Lusch, B., Kutz, J. N., and Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. Nature communications, 9(1), 4950.

Martinez-Villalobos, C., and Neelin, J. D. (2021). Climate models capture key features of extreme precipitation probabilities across regions. Environmental Research Letters, 16(2), 024017.

Martinez-Villalobos, C., Dewitte, B., Garreaud, R. D., and Loyola, L. (2024). Extreme coastal El Niño events are tightly linked to the development of the Pacific Meridional Modes. npj Climate and Atmospheric Science, 7(1), 123.

Nathaniel, J., Roesch, C., Buch, J., DeSantis, D., Rupe, A., Lamb, K., and Gentine, P. (2025). Deep Koopman operator framework for causal discovery in nonlinear dynamical systems. arXiv preprint arXiv:2505.14828.

Navarra, A., Tribbia, J., and Klus, S. (2021). Estimation of Koopman transfer operators for the equatorial Pacific SST. Journal of the Atmospheric Sciences, 78(4), 1227-1244.

Penland, C., and Matrosova, L. (1994). A balance condition for stochastic numerical models with application to the El Niño-Southern Oscillation. Journal of climate, 7(9), 1352-1372.

Priya, P., and Dommenget, D. (2025). ENSO phase space dynamics in CMIP models. Climate Dynamics, 63(6), 270.

Richter, I., Stuecker, M. F., Takahashi, N., and Schneider, N. (2022). Disentangling the North Pacific meridional mode from tropical Pacific variability. Npj Climate and Atmospheric Science, 5(1), 94.

Takahashi, K., Montecinos, A., Goubanova, K., and Dewitte, B. (2011). ENSO regimes: Reinterpreting the canonical and Modoki El Niño. Geophysical research letters, 38(10).

Takeishi, N., Kawahara, Y., and Yairi, T. (2017). Learning Koopman invariant subspaces for dynamic mode decomposition. Advances in neural information processing systems, 30.