Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL118649

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, Santiago, Chile, [2]Data Observatory Foundation, ANID Technology Center No. DO210001, Santiago, Chile, [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, 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, Coquimbo, 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, TAS, Australia, [11]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, 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

Abstract Linear Inverse Models (LIMs) are widely used data-driven tools for studying El Niño Southern Oscillation (ENSO). However, standard LIMs struggle to simulate the observed asymmetry and diversity of ENSO events. Observations reveal that strong Central Pacific (CP) La Niñas and extreme Eastern Pacific (EP) El Niños occur more frequently than their counterparts, a feature standard LIMs fail to capture. We introduce a modified model, the Non-Gaussian LIM (NG-LIM), which transforms the LIM variables to better simulate ENSO asymmetry and diversity. Specifically, the NG-LIM reproduces the spatial pattern of sea surface temperature (SST) skewness and the inverted U-shaped relationship between the first two principal components of Tropical Pacific SST anomalies, reflecting more frequent strong CP La Niñas and extreme EP El Niños. NG-LIM simulations also show El Niños that are stronger and evolve more rapidly than La Niñas. This improved inverse model generates synthetic events to supplement the limited observational record.

Plain Language Summary El Niño and La Niña are dominant patterns of climate variability that can have wide-reaching impacts on weather and ecosystems worldwide. Scientists often use mathematical models to study these events, including Linear Inverse Models (LIMs), which analyze past data to make predictions. However, standard LIMs struggle to capture certain asymmetric features of El Niño and La Niña events, like their uneven strength and their spatial footprint. For instance, intense El Niños tend to develop quickly and decay rapidly, while La Niñas often linger longer but are not as extreme. In this study, we introduce a modified model, the Non-Gaussian LIM (NG-LIM), which better represents these asymmetries between El Niño and La Niña. This modified model generates a broader range of synthetic events, providing a valuable tool for understanding these climate patterns.

Key Points:

- Standard Linear Inverse Models (LIMs) do not correctly simulate El Niño Southern Oscillation (ENSO) asymmetry and diversity compared with observations
- We propose a modification to standard LIMs, which realistically replicates the observed ENSO asymmetry and diversity
- This Non-Gaussian LIM (NG-LIM) generates diverse ENSO events, building a synthetic library to supplement limited observational data

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Martinez-Villalobos,
cristian.martinez.v@uai.cl

Citation:

Martinez-Villalobos, C., Capotondi, A., Deser, C., Dewitte, B., Holbrook, N. J., Newman, M., et al. (2025). A low-order data-driven model of ENSO diversity. *Geophysical Research Letters*, *52*, e2025GL118649. https://doi.org/10.1029/2025GL118649

Received 5 AUG 2025
Accepted 3 DEC 2025

1. Introduction

El Niño-Southern Oscillation (ENSO) is the dominant mode of climate variability on interannual timescales, influencing global weather patterns and impacting ecosystems, agriculture, and economies across the world (Anderson et al., 2017; Liu et al., 2023; McPhaden et al., 2006; Naylor et al., 2007). ENSO events manifest primarily as anomalies in sea surface temperatures (SST) in the central and eastern tropical Pacific, which lead to widespread changes in atmospheric circulation, affecting regions far beyond the equatorial Pacific (Ashok & Saji, 2007; Deser et al., 2017; Garreaud et al., 2020; Taschetto & England, 2009). A key characteristic of ENSO is its diversity; events can vary greatly in strength, duration, and spatial patterns (Ashok et al., 2007; Capotondi et al., 2015, 2020; Karnauskas, 2013; Thomas et al., 2018; Thual & Dewitte, 2023; Timmermann et al., 2018). This diversity ultimately arises from deterministic and stochastic ocean-atmosphere feedbacks, including nonlinear advection, wind stress, thermocline effects, and state-dependent westerly wind bursts (N. Chen & Majda, 2017; Choi et al., 2013; DiNezio & Deser, 2014; Kim & An, 2020; Levine et al., 2016; Liang et al., 2012;

Martinez-Villalobos et al., 2019; Thual et al., 2016). El Niño events are typically classified as Eastern Pacific (EP) events, which include strong events, and Central Pacific (CP) events, which are typically weaker and exhibit the largest anomalies in the central equatorial Pacific (Dewitte & Takahashi, 2019; Kao & Yu, 2009; Takahashi et al., 2011; Vimont et al., 2014). Furthermore, ENSO exhibits asymmetries, notably El Niño events tend to be stronger and shorter-lived than La Niña events (Jin et al., 2020; Martinez-Villalobos et al., 2019; Ohba & Watanabe, 2012; Okumura & Deser, 2010), leading to varied global impacts and teleconnections (Alexander et al., 2002; McPhaden et al., 2006; Wallace et al., 1998).

The main characterization of ENSO diversity and asymmetry between El Niño and La Niña has been established using an observational record that is relatively limited (Wittenberg, 2009), with only approximately 30 warm/cold total events over the last century (Okumura & Deser, 2010). This limited data set constrains our ability to understand the full range of variability and the underlying mechanisms of ENSO, especially how the characteristics of ENSO diversity and asymmetry can vary across timescales (Lee et al., 2021; Planton et al., 2024). For example, do decadal or centennial changes in ENSO characteristics necessarily imply changes in ENSO's underlying dynamics? The observed record is too short to be able to address that question with confidence.

One effective approach to address the gaps in our knowledge of the statistical properties of ENSO from the limited observational record is the use of empirically derived inverse models. Among these, Linear Inverse Models (LIMs; (Penland & Sardeshmukh, 1995)) are the most widely utilized. LIMs have been employed for various purposes, including identification of patterns that "optimally" grow into ENSO events (Capotondi & Sardeshmukh, 2015; Lou et al., 2021; Penland & Sardeshmukh, 1995; Vimont et al., 2014, 2022), assessing ENSO predictability (Flügel et al., 2004; Newman & Sardeshmukh, 2017; Penland & Magorian, 1993), and understanding ENSO's underlying dynamics (Newman, Alexander, & Scott, 2011; Penland & Sardeshmukh, 1995), in particular, ENSO-associated stochastic forcing (Penland, 1996; Thomas et al., 2018), diversity (Newman, Shin, & Alexander, 2011), asymmetry (Martinez-Villalobos et al., 2019) and irregularity (Berner et al., 2018; Flügel et al., 2004) (cf. (An et al., 2020)). These models are trained on observed data, allowing for the simulation of features consistent with the statistical properties of the real climate system, as inferred from observations. While the statistics generated are inherently constrained by the length of the observational record, which limits the ability to infer past or future changes in dynamics, LIMs can effectively produce parallel climatic realizations (or "multiverse" "realizations") (Newman, Shin, & Alexander, 2011; Herein et al., 2017; Martinez-Villalobos et al., 2024). Thus, LIMs generate robust statistics by creating simulations of events that are consistent with the observed dynamics but that have not yet been sampled. These simulated events align with some statistical metrics derived from the observed data. However, as we will demonstrate (see also (Martinez-Villalobos et al., 2019)), the standard LIM captures the basic EP/CP patterns but underestimate asymmetry and higher-order aspects of diversity.

Here, we propose a straightforward modification to the traditional LIM, termed the Non-Gaussian LIM (NG-LIM). This new low-order and parsimonious approach enhances the simulation of ENSO asymmetry and diversity by addressing the standard LIM's constraint of producing Gaussian statistics by construction.

2. Data and Methods

2.1. Data

We use monthly sea surface temperature (SST) data from the NOAA Extended Reconstruction SST v5 (ERSSTv5) reanalysis from 1948 to 2022 (B. Huang et al., 2017). We calculate monthly SST anomalies (SSTA) by subtracting the first two Fourier harmonics of the monthly SST climatology, and remove a cubic trend at each point. We also used sea surface temperature data from a selection of 30 CMIP6 models (Eyring et al., 2016) to benchmark the realism of the standard and Non-Gaussian LIM simulations. This comparison is generally favorable for the NG-LIM. Details are provided in Text S1 in Supporting Information S1 and associated figures in the Supporting Information S1.

2.2. Inverse Models

2.2.1. Standard Linear Inverse Model (LIM)

In a standard stationary LIM (Penland, 1989; Penland & Sardeshmukh, 1995), we approximate the evolution of a vector $\mathbf{x}$ representing the state of the tropical Pacific using the following expression

$$\frac{d\mathbf{x}}{dt} = \mathbf{M}\mathbf{x} + \mathbf{B}\boldsymbol{\eta}. \quad (1)$$

Here, **Mx** represents the linear approximation to the deterministic dynamics (hence the name "Linear Inverse Model"), **B** is a noise amplitude matrix and η is a vector of Gaussian white noise processes. The combination $\mathbf{B}\boldsymbol{\eta}$ yields stochastic forcing that is white in time but spatially coherent.

We represent the state of the Tropical Pacific by using a combination of the first 10 standardized principal components of tropical Pacific SSTA in region [20S–20N; 120E−50W] (explaining 90% of variance), corresponding to their respective empirical orthogonal functions (EOFs) (See Figure S1 in Supporting Information S1, for the first 2 EOFs spatial patterns). In our case, $\mathbf{x} = (EP, CP, PC_3, PC_4, \ldots, PC_{10})$. Note that the first two components of the state vector are given by a rotation of the first 2 PCs yielding representations of Eastern (EP) and CP events ($EP = \frac{1}{\sqrt{2}}(PC_1 - PC_2)$; $CP = \frac{1}{\sqrt{2}}(PC_1 + PC_2)$) (Takahashi et al., 2011) (Figures 1a and 1b). We calculate the deterministic operator as in Penland and Sardeshmukh (1995) using a lag of 1 month and the covariance matrix of stochastic forcing using the Fluctuation-Dissipation relationship (Penland & Matrosova, 1994) (See Text S2 in Supporting Information S1). Consistent with previous studies (Martinez-Villalobos et al., 2019; Penland & Sardeshmukh, 1995; Sardeshmukh & Sura, 2009), we find that the statistics generated by the standard LIM are Gaussian, and hence unable to accurately reproduce some statistical properties of observed ENSO, including the probability of extreme events, even when accounting for sampling variability.

2.2.2. Non-Gaussian Linear Inverse Model (NG-LIM)

Here, we introduce a simple modification to the LIM that allows a better representation of non-Gaussian features of ENSO, which we refer to as the Non-Gaussian LIM (NG-LIM). To construct the NG-LIM, we first transform all 10 variables within the state-vector to near Gaussianity using the Yeo-Johnson (YJ) power transformation (Yeo & Johnson, 2000). By transforming the state-vector to near-gaussianity, we expect that the asymmetric feedbacks ultimately responsible for ENSO asymmetry and diversity may also become more symmetric in the transformed variables. Unlike the more widely used Box-Cox transformation, which is only defined for positive values (Box & Cox, 1964; P. Huang et al., 2024), the YJ transformation is well defined for both positive and negative values, so it can be applied directly to anomalies. Calling the original variable y and the transformed variable y^{YJ}, this transformation is defined as

$$y^{YJ} = \begin{cases} \dfrac{\left[(y+1)^{\lambda} - 1\right]}{\lambda}, & \text{if } \lambda \neq 0,\ y \geq 0 \\ \ln(y+1), & \text{if } \lambda = 0,\ y \geq 0 \\ -\dfrac{\left[(-y+1)^{2-\lambda} - 1\right]}{2-\lambda}, & \text{if } \lambda \neq 2,\ y < 0 \\ -\ln(-y+1), & \text{if } \lambda = 2,\ y < 0, \end{cases} \quad (2)$$

where $\ln(\cdot)$ denotes the natural logarithm.

The transformation exponent λ governs how the transformation rescales the tails of a distribution to reduce skewness. As shown in Figure S2a and b in Supporting Information S1, smaller λ values (<1) progressively symmetrize positively skewed variables by compressing the upper tail, whereas larger λ values (>1) reduce negative skewness by compressing the lower tail (see also Figure S2c and d in Supporting Information S1 to see how the exponent of the transformation relates to EP/CP indices skewness in CMIP6 models). From a physical standpoint, this corresponds to rescaling the amplitude of extreme anomalies relative to moderate ones, effectively modulating the strength of nonlinear feedbacks implicit in the data. The YJ mapping thus introduces nonlinearity only through this monotonic, invertible change of variables, yielding an approximate linear evolution in the transformed space while restoring the observed non-Gaussian features when mapped back to physical variables.

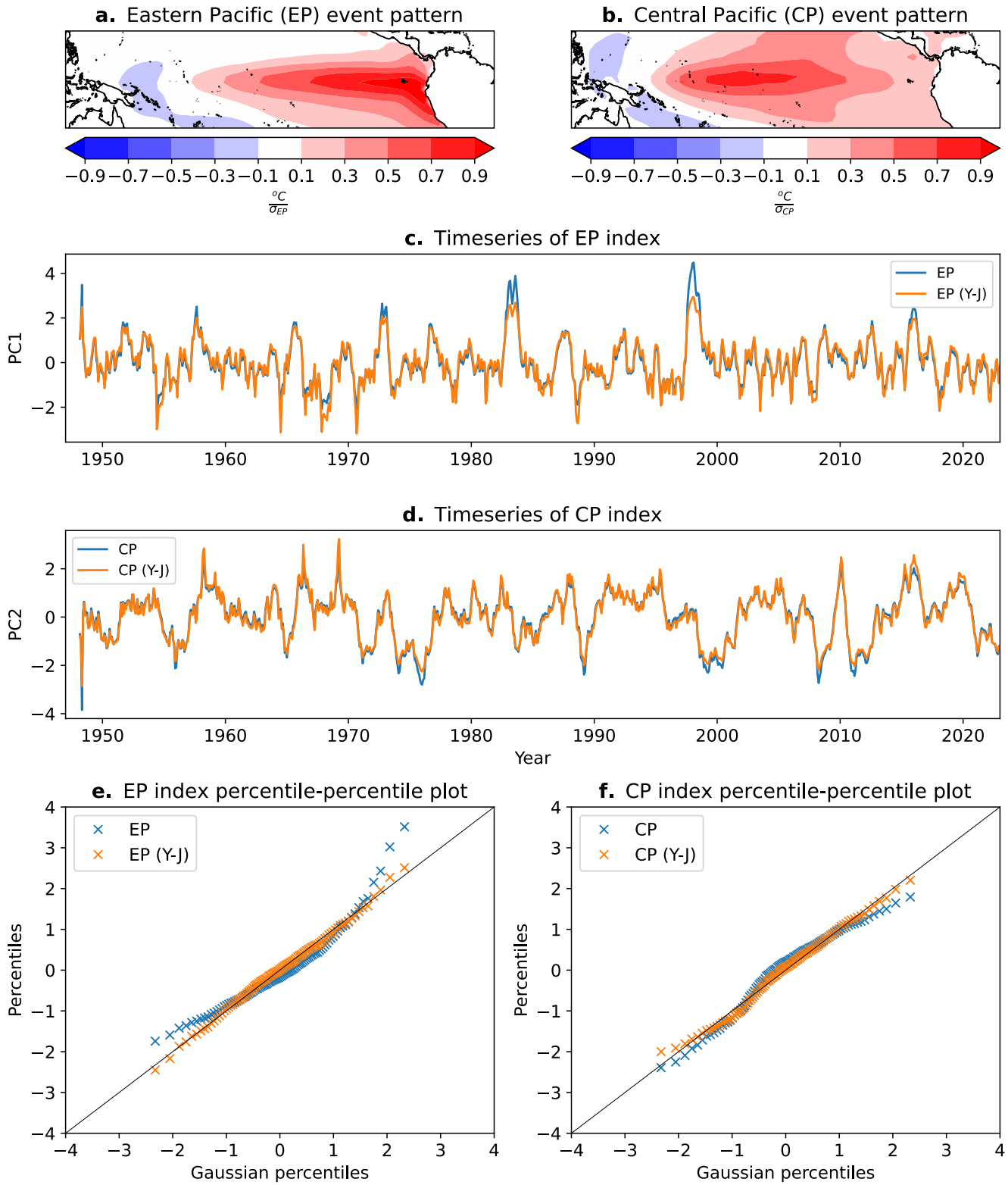

Figure 1.

The transformation is performed in Python using the Power Transformer function from the scikit-learn Preprocessing package. The exponents λ of the transformation are estimated using maximum likelihood, which is equivalent to selecting the λ that minimizes the Kullback-Leibler divergence between the transformed and a Gaussian distribution (Yeo & Johnson, 2000), thereby making the transformed data as close to normal as possible. The uncertainty in λ (5th–95th percentiles) is assessed via block-bootstrap resampling of the 75 years record using 5 years segments with replacement. The values of the exponents for EP and CP indices are $\lambda = 0.45$ (0.30–0.64) and $\lambda = 1.39$ (1.24–1.50), respectively.

As an example, Figure 1 shows the original and transformed components of the state-vector time series (EP and CP indices; Figures 1c and 1d) as well as their corresponding percentile-percentile plots (Figures 1e and 1f). We observe that the original EP (CP) index time series reaches more extreme positive (negative) values compared to their negative (positive) counterparts. The transformed indices, while generally following the original indices, appear more symmetrically distributed between positive and negative values (Figures 1c and 1d). Moreover, a comparison between original and transformed variables percentiles and Gaussian theoretical percentiles (Figures 1e and 1f) shows that the transformed indices are closer to being Gaussian-distributed. The main difference between a standard LIM and the NG-LIM is that in the NG-LIM we construct a LIM using a state vector that has been previously transformed to near Gaussianity and then after the calculations are performed we take the inverse transformation [$y = \left(\lambda y^{YJ} + 1\right)^{1/\lambda} - 1$ for $y^{YJ} \geq 0$ and $y = 1 - \left((\lambda - 2)y^{YJ} + 1\right)^{1/(2-\lambda)}$ for $y^{YJ} < 0$, for $\lambda \neq 0$ and $\lambda \neq 2$] to better preserve the non-Gaussian aspects of the time series. Because this mapping is explicit and invertible, the approach remains within the classical LIM framework (a linear model in a reparameterized coordinate system).

As the transformation is univariate, and applied to a multivariate problem, it matters which variable the transformation is applied to. We also tried applying the transformation to PC1 and PC2 directly, instead of EP and CP, with little improvement compared with the standard LIM (not shown), indicating that EP and CP indices may be more appropriate ENSO variables for this purpose. It might be possible that another combination of PC1 and PC2 may yield better results, although we do not explore that here. Ideally, one would like to transform the state vector so that it becomes approximately *jointly* Gaussian, allowing a LIM to represent non-Gaussian dynamics in the original variables through linear stochastic evolution in transformed space. In general, learning a transformation that achieves joint Gaussianity is a high-dimensional problem and may require flexible, data-driven mappings (e.g., normalizing flows (Papamakarios et al., 2021)). Such approaches, while powerful, are beyond the scope of the present study. Here we take a simpler, fully explicit approach: each component of the state vector is transformed individually using the YJ transformation to make its marginal distribution approximately Gaussian. This preserves full invertibility and interpretability, while providing a first-order correction for non-Gaussianity. Although this procedure does not enforce full joint Gaussianity, previous analyses show that ENSO-related states occupy a low-dimensional manifold dominated by two principal directions (EP-like and CP-like variability) (Takahashi et al., 2011). In such a setting, marginal Gaussianization of the dominant components can still be effective: by symmetrizing the leading axes of variability, it reduces tail imbalances in the occupied region of the phase space, yielding a smoother and more isotropic distribution where linear approximations become more valid. Future work will explore extensions that learn multivariate or flow-based transformations directly from the data to capture higher-order dependencies in a more systematic way.

For our analysis, we generated two 10,000 years runs (in our case 1 year = 360 days), one using the standard LIM and another using the NG-LIM (see Equations 1 and 2), using the Euler stochastic integration scheme (Ewald & Penland, 2009) with $\Delta t = 3 days$. This yields 133 epochs of 75 years (the length of the observed data set used to calculate both LIMs). We use the 5th-and 95th percentiles across these epochs to provide a measure of the spread of the simulated statistics.

Figure 1. Spatial patterns of (a) Eastern Pacific (EP) and (b) Central Pacific (CP) events based on ERSSTv5. These are calculated as the SSTA regression pattern on the EP and CP index respectively. The corresponding EP and CP spatial patterns generated by the standard Linear Inverse Model and NG-LIM are shown in Figure S20e–f in Supporting Information S1. To provide a baseline, Figure S20g and h in Supporting Information S1 also show the CMIP6 multi-model mean EP and CP patterns (see Text S1 in Supporting Information S1). (c), (d) EP (CP) index (blue) and Yeo-Johnson (YJ) transformed EP (CP) index (orange) monthly time series. (e), (f) Percentile-percentile plots of EP (CP) (blue) and YJ transformed EP (CP) (orange) index percentiles versus theoretical Gaussian percentiles.

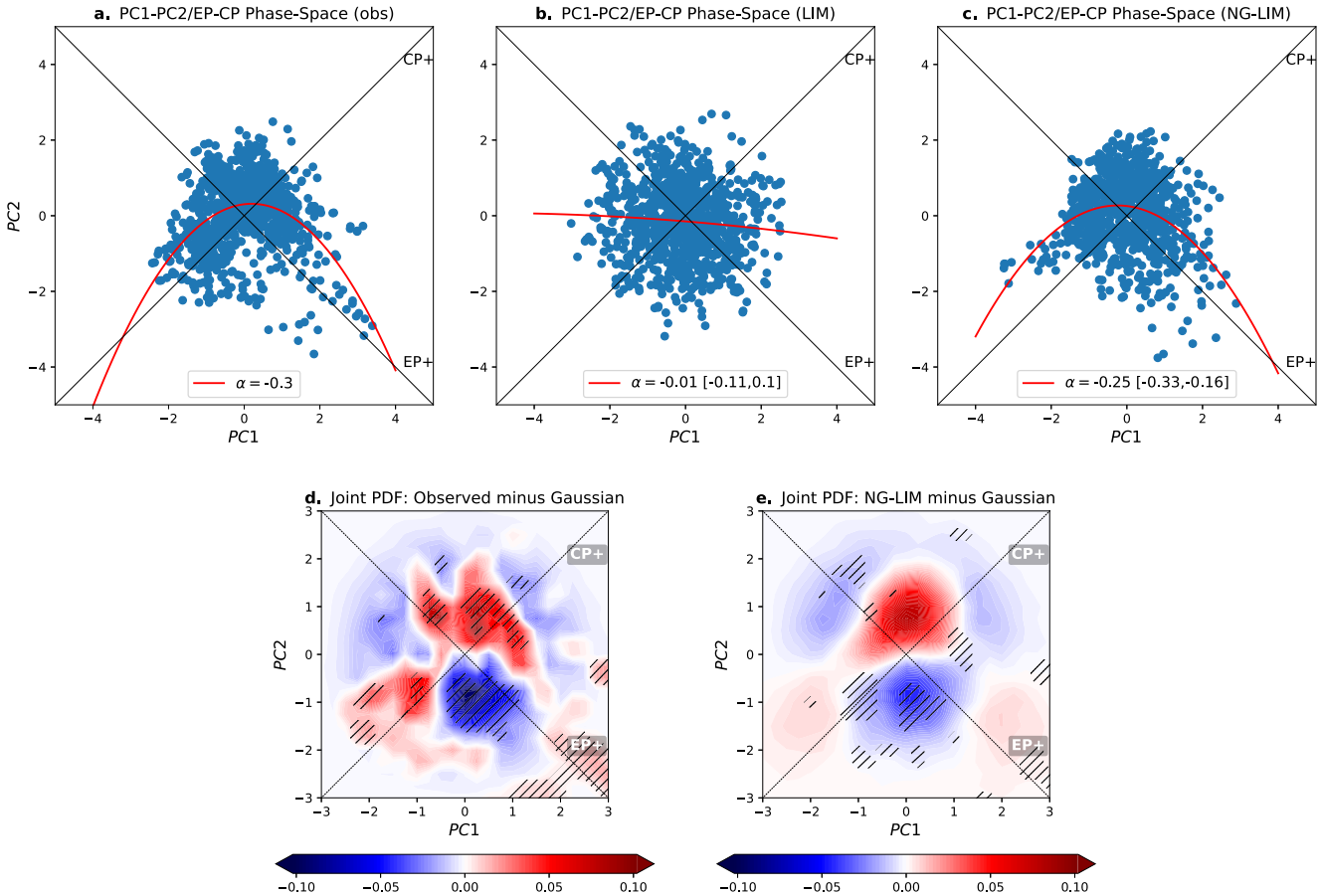

Figure 2. PC1/PC2-EP/Central Pacific (CP) indices chart in (a) Observations (1948–2022), (b) generated by the Linear Inverse Model (LIM), and (c) generated by the NG-LIM. In the case of the LIM and NG-LIM we display the epoch of the same length as observations (75 years) corresponding to the median generated α. The distribution of α in the LIM and NG-LIM is shown in Figure S5 in Supporting Information S1. (d) Observed minus Gaussian joint PDFs. (e) NG-LIM minus Gaussian joint PDFs. In d and e red and blue display regions with higher and lower probability than a Gaussian joint PDF respectively. Starting from the top left ($PC1 < 0$, $PC2 > 0$) and clockwise, the diagonals in the phase-space represent Eastern Pacific (EP) La Niñas, CP El Niños ($PC1 > 0$, $PC2 > 0$), EP El Niños ($PC1 > 0$, $PC2 < 0$) and CP La Niñas ($PC1 < 0$, $PC2 < 0$). Note that the standard LIM generates a Gaussian joint PDF, so a similar plot (standard LIM minus Gaussian) would look white. The cross-hatching in d (e) shows regions where the observed joint PDF is outside the 5th–95th percentile range of joint PDFs generated by the standard LIM (NG-LIM) across epochs.

3. Results

We first verify that the standard LIM and the NG-LIM are capable of simulating the observed autocorrelation functions and spatial patterns of SST variance and lag-variance. Both inverse models provide an accurate simulation of these observed features (Figures S3 and S4 in Supporting Information S1).

Having made those basic checks, next we examine how well the NG-LIM can simulate the joint probability distribution of the PC1-PC2 indices. In observations, the phase-space between PC1 and PC2 (or, with a rotation, EP and CP; see (Takahashi et al., 2011)) indices is not symmetric between the variables, but rather follows a characteristic inverted-U shape (or "boomerang" shape (Karamperidou et al., 2017)), as shown in Figure 2a, where each point represents the PC1/PC2 value in a given month (cf. Figure 2 of (Takahashi et al., 2011)). The observed inverted-U shape (Figure 2a) implies stronger probability of CP La Niñas and EP El Niños than their respective counterparts. This relationship can be simply represented using a quadratic fit between PC2 and PC1

$$PC2 = \alpha PC1^2 + \beta PC1 + \gamma, \tag{3}$$

with the quadratic coefficient of the fit widely used as a compact metric to represent this diversity (Cai et al., 2018; Concha et al., 2024; D. Dommenget et al., 2013; Karamperidou et al., 2017). In observations $\alpha = -0.3$ in all

months and $\alpha = -0.29$ also in December (the month when ENSO events usually peak) (Figure S5a in Supporting Information S1), indicative of this inverted-U shape relationship.

The NG-LIM successfully simulates this inverse U relationship between PC2 and PC1. When comparing across epochs, the α value calculated ranges between -0.33 and -0.16 (5th–95th percentile), thus encompassing the observed value of -0.3, with a mean value of -0.25. While there is considerable variability across epochs (Figure S6 in Supporting Information S1), this implies that regardless of the epoch, the NG-LIM simulates a curved relationship with stronger CP La Niñas and EP El Niños (Figure 2c). When sampling only Decembers (Figure S5a in Supporting Information S1), the NG-LIM similarly replicates the observed relationship (Figure S5c in Supporting Information S1), but as expected with a stronger α variability across epochs (Figure S6b in Supporting Information S1).

Figure 2a shows that observed states cluster in a particular way in the PC1-PC2 plane, which is illustrated by calculating the joint probability distributions (Figure 2d; see also Figure S7a in Supporting Information S1). Likewise, the NG-LIM successfully puts more probability for strong CP La Niñas and extreme EP El Niños (Figure 2e; see also Figure S7c in Supporting Information S1). In contrast, the standard LIM exhibits a nearly Gaussian joint distribution centered at the origin ($PC1 = 0, PC2 = 0$; Figure S7b in Supporting Information S1) and fails to reproduce the observed probability structure. Additionally, the standard LIM does not simulate a realistic probability distribution of EP and CP indices, with extreme CP La Niñas and EP El Niños being less frequent than observed (Figures 3c–3f). In all these cases, observations fall outside the 5th–95th range generated by the standard LIM. Taken together, these differences indicate that the NG-LIM produces a more realistic joint-PDF structure and overall event frequency than the standard LIM, consistent with the non-Gaussian nature of observed variability.

When calculating deviations from Gaussianity in the PC1/PC2 space, we observe that, similar to observations, the NG-LIM not only correctly simulates the excess/lack of probability at the extremes (compared to Gaussian) but also simulates well the lack/excess of probability for moderate events (Figures 2d and 2e). In particular, the NG-LIM better captures the asymmetry between coastal El Niño and La Niña events—warming/cooling events in the far-eastern Pacific not associated with basin-wide warming/cooling (Deser & Wallace, 1987; Garreaud, 2018; Rodríguez-Morata et al., 2019; Takahashi & Martínez, 2019). These coastal events constitute another aspect of ENSO diversity (Karamperidou & DiNezio, 2022). While local air-sea coupled processes may contribute to their growth, here we focus on the large-scale manifestation of coastal events, which is well represented by leading tropical Pacific modes and their interactions with meridional modes (Martinez-Villalobos et al., 2024). Consistent with this view, coastal events are characterized by low absolute values of PC1 (See Figure 3e of (Martinez-Villalobos et al., 2024)), and the model successfully reproduces the observed tendency for fewer but more extreme coastal El Niño events (Figures 2d and 2e; see also Figures S8–S10 in Supporting Information S1 for coastal event frequencies and composites, and Text S1 in Supporting Information S1 for comparison against CMIP6).

The correct simulation of the curved relationship between PC2 and PC1 also translates into how well the warm/cold asymmetries are represented spatially. We measure the asymmetry using the coefficient of skewness S, defined as $S(x) = \frac{\langle x^3 \rangle}{\langle x^2 \rangle^{3/2}}$. A positive S implies greater probability of extreme warm anomalies, and a negative S implies greater probability of extreme cold anomalies. The observed pattern of skewness (Figure 3a) shows that there are stronger warm events (i.e., super El Niños) in the Niño 3.4, Niño 3 and especially Niño 1 + 2 regions, whereas there are stronger cold events in the Niño 4 region and the poleward flanks of the CP. As expected, the standard LIM yields negligible skewness across the domain, consistent with its Gaussian structure (Figure S7b in Supporting Information S1). While there are some differences in the central-western Pacific, the simulated skewness pattern by the NG-LIM shares the observed features, with stronger Niñas in the west and stronger Niños in the east (Figure 3b). Related asymmetries in the temporal evolution and skewness of the Niño 1 + 2, Niño 3, and Niño 4 indices are also captured by the NG-LIM but not by the standard LIM (Figure S11 in Supporting Information S1), and are generally better reproduced than in the CMIP6 ensemble (Figure S12 in Supporting Information S1).

We have shown that the NG-LIM represents a distinct and meaningful improvement over the standard LIM in aspects related to ENSO asymmetry and diversity. Importantly, despite its simplicity, NG-LIM performs comparably to-if not better than-current coupled dynamical models (e.g., CMIP6) in representing these features

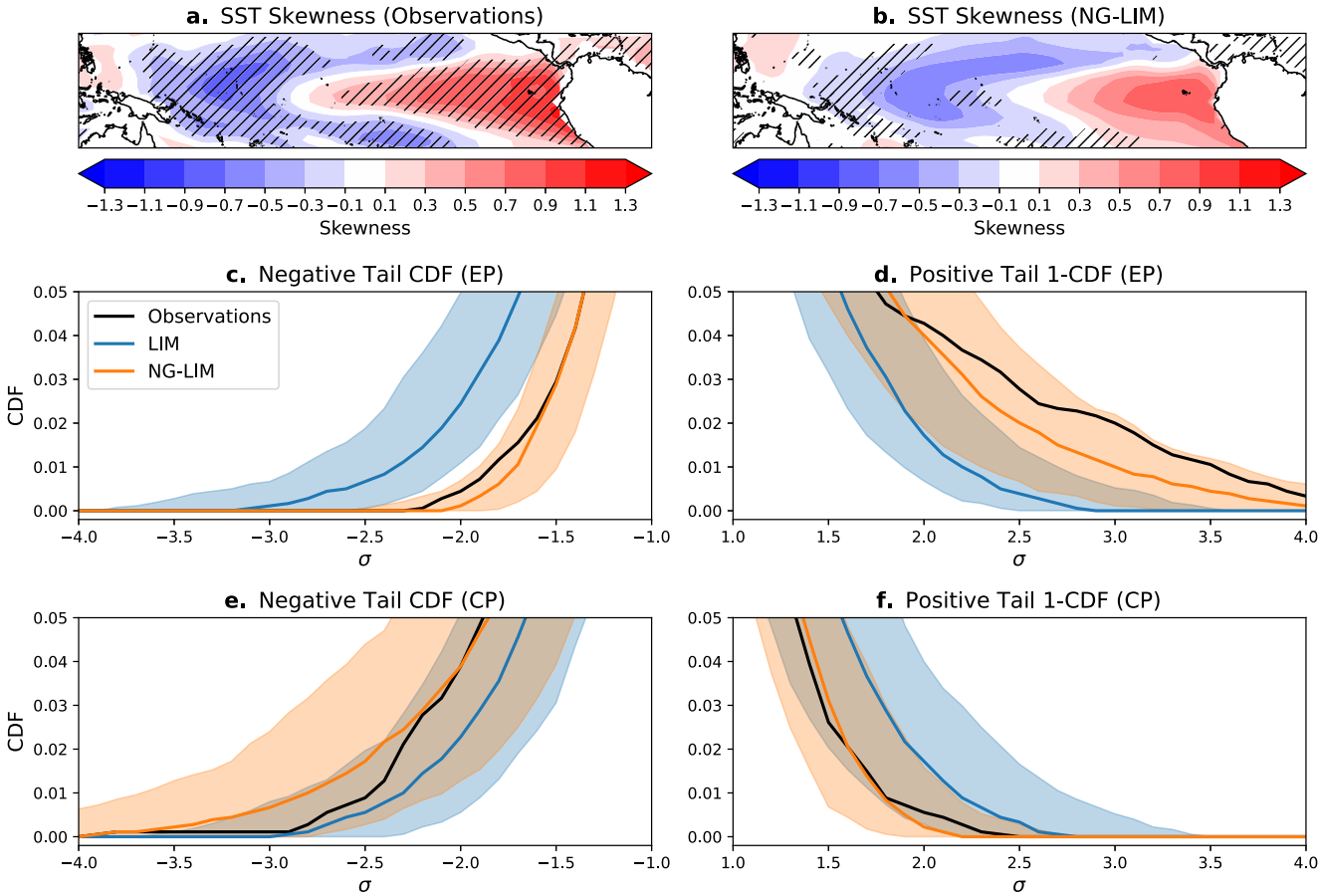

Figure 3. (a) Spatial map of SST skewness coefficient in observations. (b) Same as (a) but for the whole 10,000 years integration of the NG-LIM. The cross-hatching in a (b) shows regions where the observed SSTA skewness is outside the 5th–95th percentile range of skewness generated by the standard Linear Inverse Model (LIM) (NG-LIM) across epochs. Figure S13 in Supporting Information S1 shows a similar comparison as panels a and b but for kurtosis. (c), (d) Negative (positive) tail of the cumulative distribution function (CDF) of the Eastern Pacific (EP) 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. Panel d shows the exceedance (1-CDF). To emphasize the extremes, the panels only show the CDF or exceedance for anomalies below $-\sigma_{EP}$ or above σ_{EP} as appropriate. (e), (f) Same as (c), (d) but for the Central Pacific (CP) index. Figure S14 in Supporting Information S1 shows the whole EP and CP indices CDFs.

(see Text S1 in Supporting Information S1). Having established these improvements, we may also ask how the evolution of events differs between cold and warm phases. We illustrate this by showing the difference in evolution of cold and warm optimal patterns —patterns that through deterministic dynamics optimally grow into Niña and Niño events a number of months later (Penland & Sardeshmukh, 1995; Vimont et al., 2014). The τ-months optimal pattern (i.e., the initial conditions that maximize growth of domain-integrated SSTA variance over τ months; cf. (Martinez-Villalobos & Vimont, 2016; Penland & Sardeshmukh, 1995; Vimont et al., 2014; Zanna & Tziperman, 2005) is found as the leading eigenvector of $\mathbf{G}^T(\tau)\mathbf{G}(\tau)$, where $\mathbf{G}(\tau) = \exp(\mathbf{M}^*\tau)$, and $\mathbf{M}^*$ is analogous to the dynamical deterministic operator $\mathbf{M}$ of Equation 1 but now calculated using a state vector comprising the first 10 non-standardized PCs. It is important to note that the optimal pattern is calculated using the standard LIM with the aforementioned state vector, so that the optimal maximizes the growth of the tropical Pacific squared SSTA by the standard LIM, which is proportional to the sum of the amplitude squared of the PCs. Once we have the optimal pattern calculated, we evolve it using the deterministic part of the standard LIM (Equation 1) and the NG-LIM to display the differences in evolution. We note that these optimals are precursors of ENSO events; the projection of the observed data onto the optimal in Figure 4d is highly correlated with the Niño 3.4 index evolution months later (Figure S15 in Supporting Information S1).

Figure 4a (4c) shows a warm (cold) version of the 6-month optimal pattern. These optimals evolve into El Niño and La Niña events respectively 6-month later (Figures 4b and 4d). However, there are some differences in the

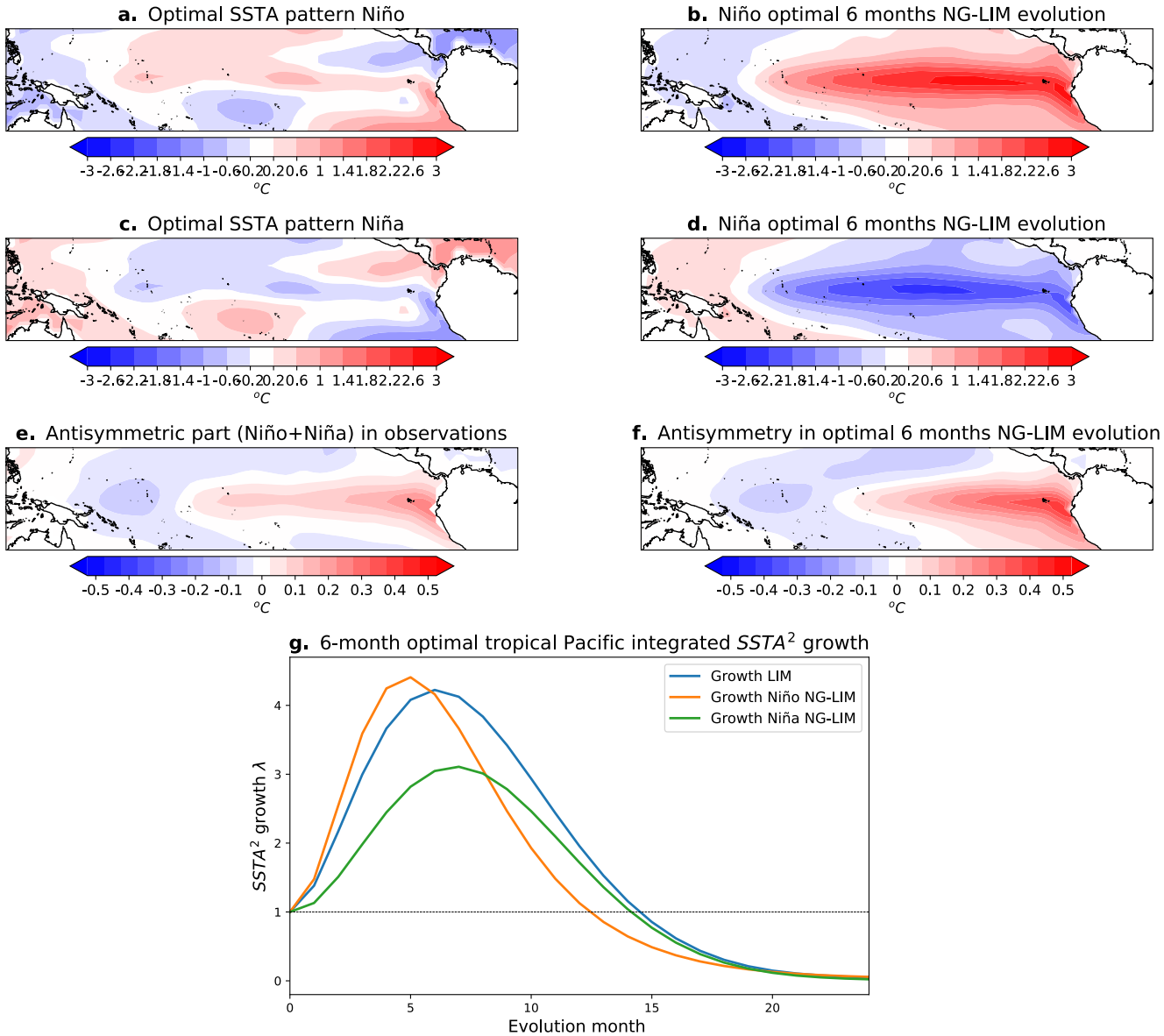

Figure 4. (a), (c) 6-month optimal initial condition, as calculated by the standard Linear Inverse Model (LIM), that evolves into an El Niño (La Niña) event months later. (b), (d) 6-month evolution of the optimal in a (b) by the NG-LIM. (e) Composite of the spatial pattern of the antisymmetric part of El Niño Southern Oscillation in observations. This is calculated as the addition between the composite of El Niño and La Niña events. For the composite, Niño (Niña) events are defined as months where the Niño 3.4 index is above (below) 0.5°C (−0.5°C). (f) Asymmetry between the NG-LIM evolution of warm and cold optimals. This is calculated as pattern b plus pattern d. Note that patterns 4e and 4f are not directly comparable; see Figure S17 in Supporting Information S1 for a comparison of the analogous to the antisymmetry pattern in observations (Panel 4e) in LIM, NG-LIM, and also the CMIP6 ensemble. (g) Ratio between the tropical Pacific SSTA amplitude squared of the optimal as it evolves normalized by the SSTA amplitude squared at the initial condition. This is shown for three cases: i. Optimal initial condition evolved by the deterministic part of the standard LIM (blue). In this case, the curves are identical for warm and cold optimals. ii. Cold optimal evolved by the NG-LIM (evolve into a La Niña event; green). iii. warm optimal evolved by the NG-LIM (evolve into an El Niño event; orange).

evolution; warm optimals tend to develop stronger anomalies in the east and cold optimals stronger anomalies in the west (Figure 4f), resembling the antisymmetry of El Niño and La Niña events in observations (Figure 4e). As expected, the standard LIM does not generate this difference, instead evolving cold and warm optimals symmetrically (Figure S16 in Supporting Information S1). This contrast further illustrates NG-LIM's ability to capture antisymmetric ENSO dynamics, particularly the tendency for warm events to grow more strongly in the EP direction, while cold events preferentially amplify in the CP direction (see also (Yu et al., 2023)). Moreover, the deterministic growth of warm optimals tends to be more rapid and generates stronger anomalies than the cold

optimals, which tend to grow and decay more slowly, much like the general tendency of events in observations (Okumura & Deser, 2010; Yu et al., 2023; B. D. Dommenget & Al-Ansari, 2023) (Figure 4g).

4. Summary and Discussion

Most ENSO modeling is conducted in the forward sense, that is, starting from fundamental equations derived from physical principles (often with additional parametrizations), in hopes of gaining insight into the real-world system. A prime example of this approach is the use of coupled global climate models, such as the Community Earth System Model (Danabasoglu et al., 2020; Hurrell et al., 2013). A pragmatic and cost effective alternative is the use of data-driven models such as LIMs. The starting point for these models is the observed data, from which the task is to reverse-engineer the best model consistent with these data. These approaches, and especially the LIM methodology, have been useful to study many aspects of ENSO behavior, including its irregularity (Berner et al., 2018; Penland & Sardeshmukh, 1995), asymmetry (Martinez-Villalobos et al., 2019), patterns of growth and decay (Capotondi & Sardeshmukh, 2015; Penland & Sardeshmukh, 1995; Vimont et al., 2014, 2022), predictability (Mason & Mimmack, 2002; Newman & Sardeshmukh, 2017), statistical significance of epochs' changes (Capotondi & Sardeshmukh, 2017; Martinez-Villalobos et al., 2019), and modulation of marine heatwaves (Capotondi et al., 2022; Gregory et al., 2024). However, by construction standard LIMs are constrained to generate Gaussian statistics, which are unable to fully characterize the asymmetry of diverse ENSO events. Here, we propose an empirical approach to deal with this inherent limitation of the standard LIM. It consists of first transforming the variables comprising the relevant state-vector to near-Gaussianity (by maximizing likelihood) and then calculating a standard LIM in those transformed variables. The robustness of the method is validated a posteriori by comparison with observed statistics. This transformed space more closely satisfies the assumptions behind the LIM, namely that the system behaves approximately linearly and follows a near-Gaussian distribution, which enables the use of a standard LIM formulation before systematically inverse-transforming back to physical space. This transformation-based approach has conceptual parallels with Koopman operator theory (Nathaniel et al., 2025; Navarra et al., 2021; Penland & Sardeshmukh, 2024; Sánchez et al., 2024), which underpins many recent efforts to linearize nonlinear dynamics in a suitable feature space; we elaborate on this connection in Text S3 in Supporting Information S1.

Applying this framework yields several tangible improvements. The Non-Gaussian LIM (NG-LIM) generates symmetric cold and warm events in the transformed variables, with the asymmetry being introduced when transforming back to the original variables. With this relatively simple modification, the NG-LIM better replicates the spatial asymmetry between El Niño and La Niña, the joint probability distribution of PC1/PC2-EP/CP, — including the inverted-U relationship between PC2 and PC1 — and crucially the probability of warm and cold extremes. Moreover, it is capable of simulating the differences in evolution between warm and cold events, with El Niño events that deterministically grow stronger and decay faster than La Niña events from a given initial optimal condition. A concise summary of these improvements relative to the standard LIM is provided in Table S1 in Supporting Information S1.

The intent of the present study is to propose a modification to a standard LIM that more realistically captures ENSO characteristics; as such, the treatment herein should be considered a starting point for future research avenues. For example, it would be valuable to evaluate to what extent, if any, the NG-LIM could improve on the standard LIM in terms of predictive skill of strong events (e.g., Super El Niños; cf. (Lenssen et al., 2024; Newman & Sardeshmukh, 2017; Schlör et al., 2024)). Additionally, it could be worth investigating whether the NG-LIM produces more interdecadal ENSO amplitude modulation and a larger level of tropical Pacific decadal variability (TPDV; (Capotondi et al., 2023; Di Lorenzo et al., 2023; Newman et al., 2016)) than a standard LIM, by better representing the ENSO asymmetry. A standard LIM can produce ENSO-related TPDV mainly by randomly generating epochs of strong El Niños or La Niñas. Additionally, the NG-LIM could also produce TPDV through amplitude modulation, in which the strong-ENSO decades are more warm-skewed than the weak decades, leading to a decadal residual of warmer east and colder west during the strong-ENSO epochs (e.g. (Atwood et al., 2017; Fedorov et al., 2020; Ogata et al., 2013; Power et al., 2021; Vimont, 2005)).

In terms of the NG-LIM construction, future improvements could include: (a) Extending the state-vector to also incorporate measures of ocean memory (Capotondi & Sardeshmukh, 2015; Newman, Alexander, & Scott, 2011; Xue et al., 2000), such as ocean heat content, that could conceivably extend the horizon of ENSO predictability; (b) A cyclostationary NG-LIM with month-dependent operators is a natural extension (Johnson et al., 2000; Lien

et al., 2025; OrtizBeviá, 1997; Penland, 1996; Shin et al., 2021; Vimont et al., 2022; Wang et al., 2023). Estimating these by calendar month and integrating with time-varying operators would allow the model to better capture the observed seasonal modulation in variance and skewness, and to examine how seasonality interacts with ENSO diversity; (c) investigating the optimal evolution of events, not only in the L^2 sense, but also specifically in the EP and CP direction, such that the asymmetry in the evolution of Central and EP events could be better identified and also perhaps predicted (e.g. (Vimont et al., 2014; Vimont et al., 2022),). Beyond ENSO itself, NG-LIM's improved ability to capture the full range of ocean temperature extremes, including not only Super El Niños but also marine heatwaves and cold spells, makes it a valuable tool for characterizing prolonged, high-impact climate events in the tropical Pacific.

There is some debate on whether ENSO could be better conceived as a nonlinear deterministic process with or without stochastic forcing (e.g (Battisti & Hirst, 1989; Cane & Zebiak, 1985; Neelin, 1991; Neelin et al., 1998; Neelin & Jin, 1993; Schopf & Suarez, 1988)), or more parsimoniously as an approximate linear damped deterministic process energized by stochastic forcing (e.g., (Penland & Sardeshmukh, 1995; Thompson & Battisti, 2001)). In this latter view, asymmetries arise due to the interaction between rapid variations at the synoptic scale and the more slowly evolving state of the system (Martinez-Villalobos et al., 2019). In that sense, there are versions of inverse models that consider nonlinear deterministic dynamics (e.g. (Kondrashov et al., 2005; Kravtsov et al., 2005; C. Chen et al., 2016; Martinez-Villalobos et al., 2024), and others whose deterministic dynamics remain linear but whose stochastic forcing depends on the state of the system (e.g (Martinez-Villalobos et al., 2019; Sardeshmukh & Sura, 2009). These nonlinear versions can also improve, to some extent, on the standard LIM in terms of their simulation of ENSO diversity and asymmetry. The NG-LIM sidesteps this debate and its motivations are more practical. First, it is simpler than the inverse models previously described. Second, and more importantly, because it avoids fitting nonlinear interaction terms or state-dependent noise structures, it requires significantly less data than the aforementioned models to be fitted (cf. (Martinez-Villalobos et al., 2018; Martinez-Villalobos et al., 2024)).

Given the improved representation of ENSO diversity and asymmetry provided by the NG-LIM, we have made available to the research community two long integrations of the state of the Tropical Pacific (one standard LIM, and one NG-LIM) consistent with the observed statistics of ERSSTv5 over 1948 to 2022. An example of EP and CP indices calculated from these integrations are shown in Figures S18 and S19 in Supporting Information S1. These integrations can be used for a variety of purposes, including the assessment of coupled GCMs and as a null-hypothesis for apparent changes in the characteristics of El Niño and La Niña events due to sampling fluctuations, as opposed to external forcing. In addition, we make available the NG-LIM emulation of 30 CMIP6 models (see Text S1 in Supporting Information S1 for details). Several key features of ENSO diversity and asymmetry, such as east–west structural imbalances and coastal event contrasts, are often misrepresented or muted in coupled GCMs (e.g (Karamperidou et al., 2017); see also Text S1 in Supporting Information S1). The NG-LIM, despite its simplicity, recovers many of these features (Figures S20–S23 in Supporting Information S1), highlighting its potential as a diagnostic tool for evaluating ENSO-related processes in these complex models.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The LIM and NG-LIM integrations, together with the emulation of 30 CMIP6 models, can be accessed at Zenodo (Martinez-Villalobos, 2025).

Acknowledgments

CM acknowledges support from Proyecto ANID Fondecyt Iniciación código 11250471, Data Observatory Foundation ANID Technology Center No. DO210001 and Fondecyt Regular N1231174. AC was supported by the NOAA Climate Program Office Climate Variability and Predictability Program Award No. NA24OARX431C0024-T1-01, and by DOE Award No. DE-SC0023228. The National Center for Atmospheric Research is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. BD acknowledges support from ANID (Concurso de Fortalecimiento al Desarrollo Científico de Centros Regionales 2020-R20F0008-CEAZA, Centro de Investigación Oceanográfica en el Pacífico Sur-Oriental COPAS COASTAL FB210021, Fondecyt Regular 1231174) and ANR (Agence Nationale de la Recherche) through the Templex project (ANR-23-CE56-0002). NJH acknowledges funding from the ARC Centre of Excellence for the Weather of the 21st Century (CE230100012). We thank John Albers for comments on an earlier version of this paper.

References

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., & Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *Journal of Climate*, *15*(16), 2205–2231. https://doi.org/10.1175/1520-0442(2002)015<2205:tabtio>2.0.co;2

An, S.-I., Tziperman, E., Okumura, Y. M., & Li, T. (2020). ENSO irregularity and asymmetry. In *El niño Southern oscillation in a changing climate* (pp. 153–172). American Geophysical Union (AGU).

Anderson, W., Seager, R., Baethgen, W., & Cane, M. (2017). Crop production variability in North and South America forced by life-cycles of the El Niño Southern Oscillation. *Agricultural and Forest Meteorology*, *239*, 151–165. https://doi.org/10.1016/j.agrformet.2017.03.008

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research*, *112*(C11), C11007.

Ashok, K., & Saji, N. H. (2007). On the impacts of ENSO and Indian Ocean dipole events on sub-regional Indian summer monsoon rainfall. *Natural Hazards*, *42*(2), 273–285. https://doi.org/10.1007/s11069-006-9091-0

Atwood, A. R., Battisti, D. S., Wittenberg, A. T., Roberts, W. H. G., & Vimont, D. J. (2017). Characterizing unforced multi-decadal variability of ENSO: A case study with the GFDL CM2.1 coupled GCM. *Climate Dynamics*, *49*(7), 2845–2862. https://doi.org/10.1007/s00382-016-3477-9

Battisti, D. S., & Hirst, A. C. (1989). Interannual variability in a tropical atmosphere–ocean model: Influence of the basic State, Ocean geometry and nonlinearity. *Journal of the Atmospheric Sciences*, *46*(12), 1687–1712. https://doi.org/10.1175/1520-0469(1989)046<1687:iviata>2.0.co;2

Berner, J., Sardeshmukh, P. D., & Christensen, H. M. (2018). On the dynamical mechanisms governing El niño–southern oscillation irregularity. *Journal of Climate*, *31*(20), 8401–8419. https://doi.org/10.1175/jcli-d-18-0243.1

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, *26*(2), 211–243.

Cai, W., Wang, G., Dewitte, B., Wu, L., Santoso, A., Takahashi, K., et al. (2018). Increased variability of eastern Pacific El Niño under greenhouse warming. *Nature*, *564*(7735), 201–206. https://doi.org/10.1038/s41586-018-0776-9

Cane, M. A., & Zebiak, S. E. (1985). A theory for El Niño and the Southern oscillation. *Science*, *228*(4703), 1085–1087. https://doi.org/10.1126/science.228.4703.1085

Capotondi, A., McGregor, S., McPhaden, M. J., Cravatte, S., Holbrook, N. J., Imada, Y., et al. (2023). Mechanisms of tropical Pacific decadal variability. *Nature Reviews Earth and Environment*, *4*(11), 754–769. https://doi.org/10.1038/s43017-023-00486-x

Capotondi, A., Newman, M., Xu, T., & Di Lorenzo, E. (2022). An optimal precursor of Northeast Pacific marine heatwaves and central Pacific El niño events. *Geophysical Research Letters*, *49*(5), e2021GL097350. https://doi.org/10.1029/2021gl097350

Capotondi, A., & Sardeshmukh, P. D. (2015). Optimal precursors of different types of ENSO events. *Geophysical Research Letters*, *42*(22), 9952–9960. https://doi.org/10.1002/2015gl066171

Capotondi, A., & Sardeshmukh, P. D. (2017). Is El Niño really changing? *Geophysical Research Letters*, *44*(16), 8548–8556. https://doi.org/10.1002/2017gl074515

Capotondi, A., Wittenberg, A. T., Kug, J.-S., Takahashi, K., & McPhaden, M. J. (2020). ENSO diversity. In *El niño Southern oscillation in a changing climate* (pp. 65–86). American Geophysical Union (AGU).

Capotondi, A., Wittenberg, A. T., Newman, M., Lorenzo, E. D., Yu, J.-Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, *96*(6), 921–938. https://doi.org/10.1175/bams-d-13-00117.1

Chen, C., Cane, M. A., Henderson, N., Lee, D. E., Chapman, D., Kondrashov, D., & Chekroun, M. D. (2016). Diversity, nonlinearity, seasonality, and memory effect in ENSO simulation and prediction using empirical model reduction. *Journal of Climate*, *29*(5), 1809–1830. https://doi.org/10.1175/jcli-d-15-0372.1

Chen, N., & Majda, A. J. (2017). Simple stochastic dynamical models capturing the statistical diversity of El Niño Southern Oscillation. *Proceedings of the National Academy of Sciences*, *114*(7), 1468–1473. https://doi.org/10.1073/pnas.1620766114

Choi, K.-Y., Vecchi, G. A., & Wittenberg, A. T. (2013). ENSO transition, duration, and amplitude asymmetries: Role of the nonlinear wind stress coupling in a conceptual model. *Journal of Climate*, *26*(23), 9462–9476. https://doi.org/10.1175/jcli-d-13-00045.1

Concha, E., Dewitte, B., Martinez-Villalobos, C., Solmon, F., & Sanchez-Gomez, E. (2024). Chile Niño/Niña in the coupled model intercomparison project phases 5 and 6. *Climate Dynamics*, *62*(11), 10049–10066. https://doi.org/10.1007/s00382-024-07434-5

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. https://doi.org/10.1029/2019ms001916

Deser, C., Simpson, I. R., McKinnon, K. A., & Phillips, A. S. (2017). The Northern hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, *30*(13), 5059–5082. https://doi.org/10.1175/jcli-d-16-0844.1

Deser, C., & Wallace, J. M. (1987). El Niño events and their relation to the Southern Oscillation: 1925–1986. *Journal of Geophysical Research*, *92*(C13), 14189–14196. https://doi.org/10.1029/jc092ic13p14189

Dewitte, B., & Takahashi, K. (2019). Diversity of moderate El Niño events evolution: Role of air–sea interactions in the eastern tropical Pacific. *Climate Dynamics*, *52*(12), 7455–7476. https://doi.org/10.1007/s00382-017-4051-9

Di Lorenzo, E., Xu, T., Zhao, Y., Newman, M., Capotondi, A., Stevenson, S., et al. (2023). Modes and mechanisms of Pacific decadal-scale variability. *Annual Review of Marine Science*, *15*(15), 249–275. https://doi.org/10.1146/annurev-marine-040422-084555

DiNezio, P. N., & Deser, C. (2014). Nonlinear Controls on the Persistence of La Niña. *Journal of Climate*, *27*(19), 7335–7355. https://doi.org/10.1175/jcli-d-14-00033.1

Dommenget, B. D., & Al-Ansari, M. (2023). Asymmetries in the ENSO phase space. *Climate Dynamics*, *60*(7), 2147–2166. https://doi.org/10.1007/s00382-022-06392-0

Dommenget, D., Bayr, T., & Frauen, C. (2013). Analysis of the non-linearity in the pattern and time evolution of El Niño southern oscillation. *Climate Dynamics*, *40*(11), 2825–2847. https://doi.org/10.1007/s00382-012-1475-0

Ewald, B., & Penland, C. (2009). Numerical generation of stochastic differential equations in climate models. In R. M. Temam & J. J. Tribbia (Eds.), *Handbook of numerical analysis* (Vol. 14, pp. 279–306). Elsevier. https://doi.org/10.1016/s1570-8659(08)00206-8

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fedorov, A. V., Hu, S., Wittenberg, A. T., Levine, A. F. Z., & Deser, C. (2020). ENSO low-frequency modulation and mean State interactions. In *El niño Southern oscillation in a changing climate* (pp. 173–198). American Geophysical Union (AGU).

Flügel, M., Chang, P., & Penland, C. (2004). The role of stochastic forcing in modulating ENSO predictability. *Journal of Climate*, *17*(16), 3125–3140. https://doi.org/10.1175/1520-0442(2004)017<3125:trosfi>2.0.co;2

Garreaud, R. D. (2018). A plausible atmospheric trigger for the 2017 coastal El Niño. *International Journal of Climatology*, *38*(S1), e1296–e1302.

Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., & Veloso-Aguila, D. (2020). The Central Chile Mega Drought (2010–2018): A climate dynamics perspective. *International Journal of Climatology*, *40*(1), 421–439. https://doi.org/10.1002/joc.6219

Gregory, C. H., Artana, C., Lama, S., León-FonFay, D., Sala, J., Xiao, F., et al. (2024). Global marine heatwaves under different flavors of ENSO. *Geophysical Research Letters*, *51*(20), e2024GL110399. https://doi.org/10.1029/2024gl110399

Herein, M., Drótos, G., Haszpra, T., Márfy, J., & Tél, T. (2017). The theory of parallel climate realizations as a new framework for teleconnection analysis. *Scientific Reports*, *7*(1), 44529. https://doi.org/10.1038/srep44529

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed Sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/jcli-d-16-0836.1

Huang, P., Chen, Y., Li, J., & Yan, H. (2024). Redefined background state in the tropical Pacific resolves the entanglement between the background state and ENSO. *npj Climate and Atmospheric Science*, *7*(1), 1–12. https://doi.org/10.1038/s41612-024-00695-1

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community Earth System model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/bams-d-12-00121.1

Jin, F.-F., Chen, H.-C., Zhao, S., Hayashi, M., Karamperidou, C., Stuecker, M. F., et al. (2020). Simple ENSO models. In *El niño Southern oscillation in a changing climate* (pp. 119–151). American Geophysical Union (AGU).

Johnson, S. D., Battisti, D. S., & Sarachik, E. S. (2000). Seasonality in an empirically derived markov model of tropical Pacific Sea surface temperature anomalies. *Journal of Climate*, *13*(18), 3327–3335. https://doi.org/10.1175/1520-0442(2000)013<3327:siaedm>2.0.co;2

Kao, H.-Y., & Yu, J.-Y. (2009). Contrasting eastern-pacific and central-pacific types of ENSO. *Journal of Climate*, *22*(3), 615–632. https://doi.org/10.1175/2008jcli2309.1

Karamperidou, C., & DiNezio, P. N. (2022). Holocene hydroclimatic variability in the tropical Pacific explained by changing ENSO diversity. *Nature Communications*, *13*(1), 7244. https://doi.org/10.1038/s41467-022-34880-8

Karamperidou, C., Jin, F.-F., & Conroy, J. L. (2017). The importance of ENSO nonlinearities in tropical pacific response to external forcing. *Climate Dynamics*, *49*(7), 2695–2704. https://doi.org/10.1007/s00382-016-3475-y

Karnauskas, K. B. (2013). Can we distinguish canonical El Niño from Modoki? *Geophysical Research Letters*, *40*(19), 5246–5251. https://doi.org/10.1002/grl.51007

Kim, S.-K., & An, S.-I. (2020). Untangling El Niño-La niña asymmetries using a nonlinear coupled dynamic index. *Geophysical Research Letters*, *47*(4), e2019GL085881. https://doi.org/10.1029/2019gl085881

Kondrashov, D., Kravtsov, S., Robertson, A. W., & Ghil, M. (2005). A hierarchy of data-based ENSO models. *Journal of Climate*, *18*(21), 4425–4444. https://doi.org/10.1175/jcli3567.1

Kravtsov, S., Kondrashov, D., & Ghil, M. (2005). Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, *18*(21), 4404–4424. https://doi.org/10.1175/jcli3544.1

Lee, J., Planton, Y. Y., Gleckler, P. J., Sperber, K. R., Guilyardi, E., Wittenberg, A. T., et al. (2021). Robust evaluation of ENSO in climate models: How many ensemble members are needed? *Geophysical Research Letters*, *48*(20), e2021GL095041. https://doi.org/10.1029/2021gl095041

Lenssen, N., DiNezio, P., Goddard, L., Deser, C., Kushnir, Y., Mason, S. J., et al. (2024). Strong El niño events lead to robust multi-year ENSO predictability. *Geophysical Research Letters*, *51*(12), e2023GL106988. https://doi.org/10.1029/2023gl106988

Levine, A., Jin, F. F., & McPhaden, M. J. (2016). Extreme noise–extreme El niño: How state-dependent noise forcing creates El Niño–La niña asymmetry. *Journal of Climate*, *29*(15), 5483–5499. https://doi.org/10.1175/jcli-d-16-0091.1

Liang, J., Yang, X.-Q., & Sun, D.-Z. (2012). The effect of ENSO events on the tropical Pacific mean climate: Insights from an analytical model. *Journal of Climate*, *25*(21), 7590–7606. https://doi.org/10.1175/jcli-d-11-00490.1

Lien, J., Kuo, Y.-N., Ando, H., & Kido, S. (2025). On cyclostationary linear inverse models: A mathematical insight and implication. *Frontiers in Complex Systems*, *3*, 1563687. https://doi.org/10.3389/fcpxs.2025.1563687

Liu, Y., Cai, W., Lin, X., Li, Z., & Zhang, Y. (2023). Nonlinear El Niño impacts on the global economy under climate change. *Nature Communications*, *14*(1), 5887. https://doi.org/10.1038/s41467-023-41551-9

Lou, J., O'Kane, T. J., & Holbrook, N. J. (2021). Linking the atmospheric Pacific-South American mode with oceanic variability and predictability. *Communications Earth and Environment*, *2*(1), 1–8.

Martinez-Villalobos, C. (2025). Tropical Pacific SST anomalies generated by a Non-Gaussian Linear Inverse Model (NG-LIM) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.14775713

Martinez-Villalobos, C., Dewitte, B., Garreaud, R. D., & 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), 1–14.

Martinez-Villalobos, C., Newman, M., Vimont, D. J., Penland, C., & David Neelin, J. (2019). Observed El Niño-La niña asymmetry in a linear model. *Geophysical Research Letters*, *46*(16), 9909–9919. https://doi.org/10.1029/2019gl082922

Martinez-Villalobos, C., & Vimont, D. J. (2016). The role of the mean State in meridional mode structure and growth. *Journal of Climate*, *29*(10), 3907–3921. https://doi.org/10.1175/jcli-d-15-0542.1

Martinez-Villalobos, C., Vimont, D. J., Penland, C., Newman, M., & Neelin, J. D. (2018). Calculating state-dependent noise in a linear inverse model framework. *Journal of the Atmospheric Sciences*, *75*(2), 479–496. https://doi.org/10.1175/jas-d-17-0235.1

Mason, S. J., & Mimmack, G. M. (2002). Comparison of some statistical methods of probabilistic forecasting of ENSO. *Journal of Climate*, *15*(1), 8–29. https://doi.org/10.1175/1520-0442(2002)015<0008:cossmo>2.0.co;2

McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2006). ENSO as an integrating concept in Earth science. *Science*, *314*(5806), 1740–1745. https://doi.org/10.1126/science.1132588

Nathaniel, J., Roesch, C., Buch, J., DeSantis, D., Rupe, A., Lamb, K., & Gentine, P. (2025). Deep Koopman operator framework for causal discovery in nonlinear dynamical systems. *arXiv*. arXiv:2505.14828. [cs]). https://doi.org/10.48550/arXiv.2505.14828

Navarra, A., Tribbia, J., & Klus, S. (2021). Estimation of Koopman transfer operators for the equatorial Pacific SST. *Journal of the Atmospheric Sciences*, *78*(4), 1227–1244. https://doi.org/10.1175/jas-d-20-0136.1

Naylor, R. L., Battisti, D. S., Vimont, D. J., Falcon, W. P., & Burke, M. B. (2007). Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proceedings of the National Academy of Sciences*, *104*(19), 7752–7757. https://doi.org/10.1073/pnas.0701825104

Neelin, J. D. (1991). The slow Sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model. *Journal of the Atmospheric Sciences*, *48*(4), 584–606. https://doi.org/10.1175/1520-0469(1991)048<0584:tssstm>2.0.co;2

Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., & Zebiak, S. E. (1998). ENSO theory. *Journal of Geophysical Research*, *103*(C7), 14261–14290. https://doi.org/10.1029/97jc03424

Neelin, J. D., & Jin, F.-F. (1993). Modes of interannual tropical ocean–atmosphere Interaction—A unified view. Part II: Analytical results in the weak-coupling limit. *Journal of the Atmospheric Sciences*, *50*(21), 3504–3522. https://doi.org/10.1175/1520-0469(1993)050<3504:moitoi>2.0.co;2

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Lorenzo, E. D., et al. (2016). The Pacific decadal oscillation, revisited. *Journal of Climate*, *29*(12), 4399–4427. https://doi.org/10.1175/jcli-d-15-0508.1

Newman, M., Alexander, M. A., & Scott, J. D. (2011). An empirical model of tropical ocean dynamics. *Climate Dynamics*, *37*(9), 1823–1841. https://doi.org/10.1007/s00382-011-1034-0

Newman, M., & Sardeshmukh, P. D. (2017). Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*, *44*(16), 8520–8529. https://doi.org/10.1002/2017gl074088

Newman, M., Shin, S.-I., & Alexander, M. A. (2011). Natural variation in ENSO flavors. *Geophysical Research Letters*, *38*(14), L14705. https://doi.org/10.1029/2011gl047658

Ogata, T., Xie, S.-P., Wittenberg, A., & Sun, D.-Z. (2013). Interdecadal amplitude modulation of El niño–southern oscillation and its impact on tropical Pacific decadal variability. *Journal of Climate*, *26*(18), 7280–7297. https://doi.org/10.1175/jcli-d-12-00415.1

Ohba, M., & Watanabe, M. (2012). Role of the Indo-Pacific interbasin coupling in predicting asymmetric ENSO transition and duration. *Journal of Climate*, *25*(9), 3321–3335. https://doi.org/10.1175/jcli-d-11-00409.1

Okumura, Y. M., & Deser, C. (2010). Asymmetry in the Duration of El Niño and La Niña. *Journal of Climate*, *23*(21), 5826–5843.

OrtizBeviá, M. J. (1997). Estimation of the cyclostationary dependence in geophysical data fields. *Journal of Geophysical Research*, *102*(D12), 13473–13486. https://doi.org/10.1029/97jd00243

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*(1), 57-2617–2680.

Penland, C. (1989). Random forcing and forecasting using principal oscillation pattern analysis. *Monthly Weather Review*, *117*(10), 2165–2185. https://doi.org/10.1175/1520-0493(1989)117<2165:rfafup>2.0.co;2

Penland, C. (1996). A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, *98*(2), 534–558. https://doi.org/10.1016/0167-2789(96)00124-8

Penland, C., & Magorian, T. (1993). Prediction of niño 3 Sea surface temperatures using Linear inverse modeling. *Journal of Climate*, *6*(6), 1067–1076. https://doi.org/10.1175/1520-0442(1993)006<1067:ponsst>2.0.co;2

Penland, C., & 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. https://doi.org/10.1175/1520-0442(1994)007<1352:abcfsn>2.0.co;2

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical Sea surface temperature anomalies. *Journal of Climate*, *8*(8), 1999–2024. https://doi.org/10.1175/1520-0442(1995)008<1999:togots>2.0.co;2

Penland, C., & Sardeshmukh, P. D. (2024). Connections between classical mathematics and machine learning. Retrieved from https://ui.adsabs.harvard.edu/abs/2024AGUFMNG23A..06P(ADSBibcode:2024AGUFMNG23A..06P)

Planton, Y. Y., Lee, J., Wittenberg, A. T., Gleckler, P. J., Guilyardi, r., McGregor, S., & McPhaden, M. J. (2024). Estimating uncertainty in simulated ENSO statistics. *Journal of Advances in Modeling Earth Systems*, *16*(9), e2023MS004147. https://doi.org/10.1029/2023ms004147

Power, S., Lengaigne, M., Capotondi, A., Khodri, M., Vialard, J., Jebri, B., et al. (2021). Decadal climate variability in the tropical Pacific: Characteristics, causes, predictability, and prospects. *Science*, *374*(6563), eaay9165. https://doi.org/10.1126/science.aay9165

Rodríguez-Morata, C., Díaz, H. F., Ballesteros-Canovas, J. A., Rohrer, M., & Stoffel, M. (2019). The anomalous 2017 coastal El Niño event in Peru. *Climate Dynamics*, *52*(9), 5605–5622. https://doi.org/10.1007/s00382-018-4466-y

Sánchez, P. L., Newman, M., Navarra, A., Albers, J., & Subramanian, A. (2024). Koopman operator theory for enhanced Pacific SST forecasting. Tech. Rep. No. EGU24-2554. *Copernicus Meetings*. Retrieved 2025-06-25, from https://meetingorganizer.copernicus.org/EGU24/EGU24-2554.html

Sardeshmukh, P. D., & Sura, P. (2009). Reconciling non-gaussian climate statistics with Linear dynamics. *Journal of Climate*, *22*(5), 1193–1207. https://doi.org/10.1175/2008jcli2358.1

Schlör, J., Newman, M., Thuemmel, J., Capotondi, A., & Goswami, B. (2024). A hybrid deep-learning model for El Niño Southern oscillation in the low-data regime. *arXiv*. Retrieved from http://arxiv.org/abs/2412.03743

Schopf, P. S., & Suarez, M. J. (1988). Vacillations in a coupled ocean–atmosphere model. *Journal of the Atmospheric Sciences*, *45*(3), 549–566. https://doi.org/10.1175/1520-0469(1988)045<0549:viacom>2.0.co;2

Shin, S.-I., Sardeshmukh, P. D., Newman, M., Penland, C., & Alexander, M. A. (2021). Impact of annual cycle on ENSO variability and predictability. *Journal of Climate*, *34*(1), 171–193. https://doi.org/10.1175/jcli-d-20-0291.1

Takahashi, K., & Martínez, A. G. (2019). The very strong coastal El Niño in 1925 in the far-eastern Pacific. *Climate Dynamics*, *52*(12), 7389–7415.

Takahashi, K., Montecinos, A., Goubanova, K., & Dewitte, B. (2011). ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters*, *38*(10), L10704. https://doi.org/10.1029/2011GL047364

Taschetto, A. S., & England, M. H. (2009). El niño modoki impacts on Australian rainfall. *Journal of Climate*, *22*(11), 3167–3174. https://doi.org/10.1175/2008jcli2589.1

Thomas, E. E., Vimont, D. J., Newman, M., Penland, C., & Martínez-Villalobos, C. (2018). The role of stochastic forcing in generating ENSO diversity. *Journal of Climate*, *31*(22), 9125–9150. https://doi.org/10.1175/jcli-d-17-0582.1

Thompson, C. J., & Battisti, D. S. (2001). A Linear stochastic dynamical model of ENSO. Part II: Analysis. *Journal of Climate*, *14*(4), 445–466. https://doi.org/10.1175/1520-0442(2001)014<0445:alsdmo>2.0.co;2

Thual, S., & Dewitte, B. (2023). ENSO complexity controlled by zonal shifts in the Walker circulation. *Nature Geoscience*, *16*(4), 328–332. https://doi.org/10.1038/s41561-023-01154-x

Thual, S., Majda, A. J., Chen, N., & Stechmann, S. N. (2016). Simple stochastic model for El Niño with westerly wind bursts. *Proceedings of the National Academy of Sciences*, *113*(37), 10245–10250. https://doi.org/10.1073/pnas.1612002113

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., et al. (2018). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. https://doi.org/10.1038/s41586-018-0252-6

Vimont, D. J. (2005). The contribution of the interannual ENSO cycle to the spatial pattern of decadal ENSO-Like variability. *Journal of Climate*, *18*(12), 2080–2092. https://doi.org/10.1175/jcli3365.1

Vimont, D. J., Alexander, M. A., & Newman, M. (2014). Optimal growth of Central and East Pacific ENSO events. *Geophysical Research Letters*, *41*(11), 4027–4034. https://doi.org/10.1002/2014gl059997

Vimont, D. J., Newman, M., Battisti, D. S., & Shin, S.-I. (2022). The role of seasonality and the ENSO mode in central and East Pacific ENSO growth and evolution. *Journal of Climate*, *35*(11), 3195–3209. https://doi.org/10.1175/jcli-d-21-0599.1

Wallace, J. M., Rasmusson, E. M., Mitchell, T. P., Kousky, V. E., Sarachik, E. S., & von Storch, H. (1998). On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA. *Journal of Geophysical Research*, *103*(C7), 14241–14259. https://doi.org/10.1029/97jc02905

Wang, Y., Holbrook, N. J., & Kajtar, J. B. (2023). Predictability of marine heatwaves off Western Australia using a linear inverse model. *Journal of Climate*, *36*(18), 6177–6193. https://doi.org/10.1175/jcli-d-22-0692.1

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*(12), L12702. https://doi.org/10.1029/2009gl038710

Xue, Y., Leetmaa, A., & Ji, M. (2000). ENSO prediction with markov models: The impact of Sea level. *Journal of Climate*, *13*(4), 849–871. https://doi.org/10.1175/1520-0442(2000)013<0849:epwmmt>2.0.co;2

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954

Yu, J., Li, T., & Jiang, L. (2023). Why Does a Stronger El Niño Favor Developing towards the Eastern Pacific while a Stronger La Niña Favors Developing towards the Central Pacific? *Atmosphere*, *14*(7), 1185. https://doi.org/10.3390/atmos14071185

Zanna, L., & Tziperman, E. (2005). Nonnormal amplification of the thermohaline circulation. *Journal of Physical Oceanography*, *35*(9), 1593–1605. https://doi.org/10.1175/jpo2777.1

References From the Supporting Information

Goldenson, N., Thackeray, C. W., Hall, A. D., Swain, D. L., & Berg, N. (2021). Using large ensembles to identify regions of systematic biases in moderate-to-heavy daily precipitation. *Geophysical Research Letters*, *48*(9), e2020GL092026. https://doi.org/10.1029/2020gl092026

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., & Noé, F. (2018). Data-Driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, *28*(3), 985–1010. https://doi.org/10.1007/s00332-017-9437-7

Li, Q., Dietrich, F., Bollt, E. M., & 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), 103111. https://doi.org/10.1063/1.4993854

Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, *9*(1), 4950. https://doi.org/10.1038/s41467-018-07210-0

Martinez-Villalobos, C., & Neelin, J. D. (2021). Climate models capture key features of extreme precipitation probabilities across regions. *Environmental Research Letters*, *16*(2), 024017. https://doi.org/10.1088/1748-9326/abd351

Priya, P., & Dommenget, D. (2025). ENSO phase space dynamics in CMIP models. *Climate Dynamics*, *63*(6), 270. https://doi.org/10.1007/s00382-025-07752-2

Richter, I., Stuecker, M. F., Takahashi, N., & Schneider, N. (2022). Disentangling the North Pacific Meridional Mode from tropical Pacific variability. *npj Climate and Atmospheric Science*, *5*(1), 94. https://doi.org/10.1038/s41612-022-00317-8

Takeishi, N., Kawahara, Y., & Yairi, T. (2017). Learning Koopman invariant subspaces for dynamic mode decomposition. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 1130–1140). Curran Associates Inc.