A 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], 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, 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

Key Points:

- Standard Linear Inverse Models (LIMs) do not correctly simulate 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.

Corresponding author: Cristian Martinez-Villalobos, `cristian.martinez.v@uai.cl`

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 La Niñas and extreme Eastern Pacific 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 effectively simulates key aspects of 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. By examining NG-LIM simulations, we find that, as observed, El Niños exhibit stronger anomalies and evolve more rapidly than La Niñas. The improved NG-LIM also generates a broad library of synthetic events, which can 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.

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 (McPhaden et al., 2006; Naylor et al., 2007; Anderson et al., 2017; Liu et al., 2023). 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; Taschetto & England, 2009; Deser et al., 2017; Garreaud et al., 2020). A key characteristic of ENSO is its diversity; events can vary greatly in strength, duration, and spatial patterns (Ashok et al., 2007; Karnauskas, 2013; Capotondi et al., 2015; Thomas et al., 2018; Timmermann et al., 2018; Capotondi et al., 2020; Thual & Dewitte, 2023). This ultimately occurs due to the presence of deterministic (e.g., nonlinear oceanic advection, wind stress, thermocline feedback; (Liang et al., 2012; Choi et al., 2013; DiNezio & Deser, 2014; Kim & An, 2020)) and/or stochastic (e.g., westerly wind bursts noise forcing that depends on the state of the ocean; (Levine et al., 2016; Thual et al., 2016; N. Chen & Majda, 2017; Martinez-Villalobos et al., 2019)) ocean-atmosphere feedbacks that operate asymmetrically between warm and cold states. 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 (Kao & Yu, 2009; Takahashi et al., 2011; Vimont et al., 2014; Dewitte & Takahashi, 2019). Furthermore, ENSO exhibits asymmetries, notably El Niño events tend to be stronger and shorter-lived than La Niña events (Okumura & Deser, 2010; Ohba & Watanabe, 2012; Martinez-Villalobos et al., 2019; Jin et al., 2020), leading to varied global impacts and teleconnections (Wallace et al., 1998; Alexander et al., 2002; McPhaden et al., 2006).

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 dataset 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 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 (Penland & Sardeshmukh, 1995; Vimont et al., 2014, 2022; Capotondi & Sardeshmukh, 2015; Lou et al., 2021), assessing ENSO predictability (Penland & Magorian, 1993; Flügel et al., 2004; Newman & Sardeshmukh, 2017), and understanding ENSO's underlying dynamics (Penland & Sardeshmukh, 1995; Newman, Alexander, & Scott, 2011), 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 (Flügel et al., 2004; Berner et al., 2018) (cf. (An et al., 2020)). These models are trained on observed data, allowing for the simulation of statistical properties that are consistent with the available dataset. While the statistics generated are inherently constrained by the fixed-length 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 most commonly used version of LIM fail to accurately capture the asymmetry and diversity of ENSO events.

Here, we propose a straightforward modification to the traditional Linear Inverse Model, termed the Non-Gaussian Linear Inverse Model (NG-LIM). This new approach enhances the simulation of ENSO asymmetry and diversity, addressing some of the limitations found in standard LIMs.

2 Data and Methods

2.1 Data

We use monthly sea surface temperature (SST) data from the NOAA Extended Reconstruction SST v5 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.

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

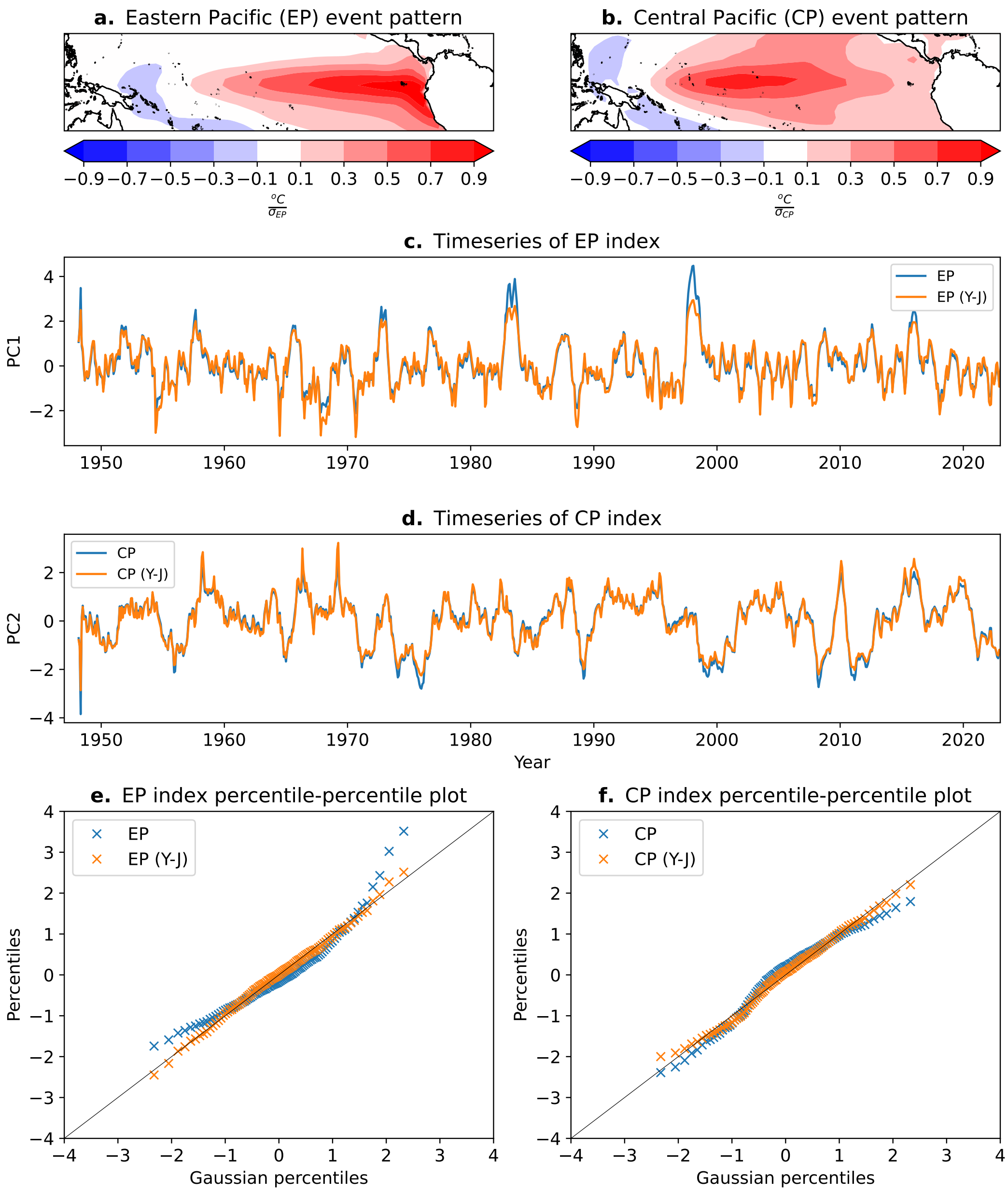

Figure 1. Spatial patterns of **a** Eastern Pacific (EP) and **b** Central Pacific (CP) events based on ERSSTv5. These are calculated as the SSTA regression pattern on the EP and CP index respectively. **c** (**d**) EP (CP) index (blue) and Yeo-Johnson transformed EP (CP) index (orange) monthly time series. **e** (**f**) Percentile-percentile plots of EP (CP) (blue) and Yeo-Johnson transformed EP (CP) (orange) index percentiles vs theoretical Gaussian percentiles

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

Here, $\mathbf{Mx}$ represents the linear approximation to the deterministic dynamics (hence the name "Linear Inverse Model"), $\mathbf{B}$ is a noise amplitude matrix and η is a vector of Gaussian white noise processes. The combination $\mathbf{B}\eta$ 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 Fig. S1, for the first 2 EOFs spatial patterns). In our case $\mathbf{x} = (EP, CP, PC_3, PC_4, ..., PC_{10})$. 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 Central Pacific (CP) events ($EP = \frac{1}{\sqrt{2}}(PC_1 - PC_2)$; $CP = \frac{1}{\sqrt{2}}(PC_1 + PC_2)$) (Takahashi et al., 2011) (Fig. 1a,b). 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 S1). Consistent with previous studies (Penland & Sardeshmukh, 1995; Sardeshmukh & Sura, 2009; Martinez-Villalobos et al., 2019), we find that the statistics generated by the standard LIM are Gaussian, and hence unable to 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 as the Non-Gaussian LIM (NG-LIM). To construct the NG-LIM, we first transform each variable within the state-vector to near Gaussianity using the Yeo-Johnson (YJ) power transformation (Yeo & Johnson, 2000). The hope is that by transforming the state-vector to near Gaussianity, implicitly the asymmetric feedbacks ultimately responsible for ENSO asymmetry and diversity 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. An example of the effect of the YJ transformation on positively and negatively skewed data is provided in Figure S2. Calling the original variable y and the transformed variable y^{YJ}, this transformation is defined as

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

The transformation is performed in Python using the PowerTransformer function from the scikit-learn Preprocessing package. The exponents λ of the transformation are calculated using maximum likelihood estimation, and we estimate uncertainty range (5th-95th percentile) by resampling the original 75yrs of data by randomly picking with replacement 5yr segments. 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, Fig. 1 shows the original and transformed components of the state-vector time series (EP and CP indices; Fig. 1c,d) as well as their corresponding percentile-

percentile plots (Fig. 1e,f). 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 (Fig. 1c,d). Moreover, a comparison between original and transformed variables percentiles and Gaussian theoretical percentiles (Fig. 1e,f) 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 [$y = (\lambda y^{YJ}+1)^{1/\lambda} - 1$ for $y^{YJ} \geq 0$ and $y = 1-((\lambda-2)y^{YJ}+1)^{1/(2-\lambda)}$ for $y^{YJ} < 0$, for $\lambda \neq 0$ and $\lambda \neq 2$] to better preserve the non-Gaussian aspects of the time series.

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.

For our analysis, we generated two 10,000 yr runs (in our case 1yr=360 days), one using the standard LIM and another using the NG-LIM (see eqs. 1 and 2), using the Euler stochastic integration scheme (Ewald & Penland, 2009) with $\Delta t = 3days$. This yields 133 epochs of 75yrs (the length of the observed dataset 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.

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).

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 (e.g., (Karamperidou et al., 2017)), as shown in Fig. 2a, where each point represents the PC1/PC2 value of a given month (cf. Fig. 2 of (Takahashi et al., 2011)). The observed inverted-U shape (Fig. 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 (Dommenget et al., 2013; Karamperidou et al., 2017; Cai et al., 2018; Concha et al., 2024). In observations $\alpha = -0.3$ in all months and $\alpha = -0.29$ also in December (the month when ENSO events usually peak) (Fig. S5a), indicative of an inverted-U shape relationship (or "boomerang" shape (Karamperidou et al., 2017)) with stronger probability of large CP La Niñas and extreme EP El Niños.

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 (Fig. S6), this im-

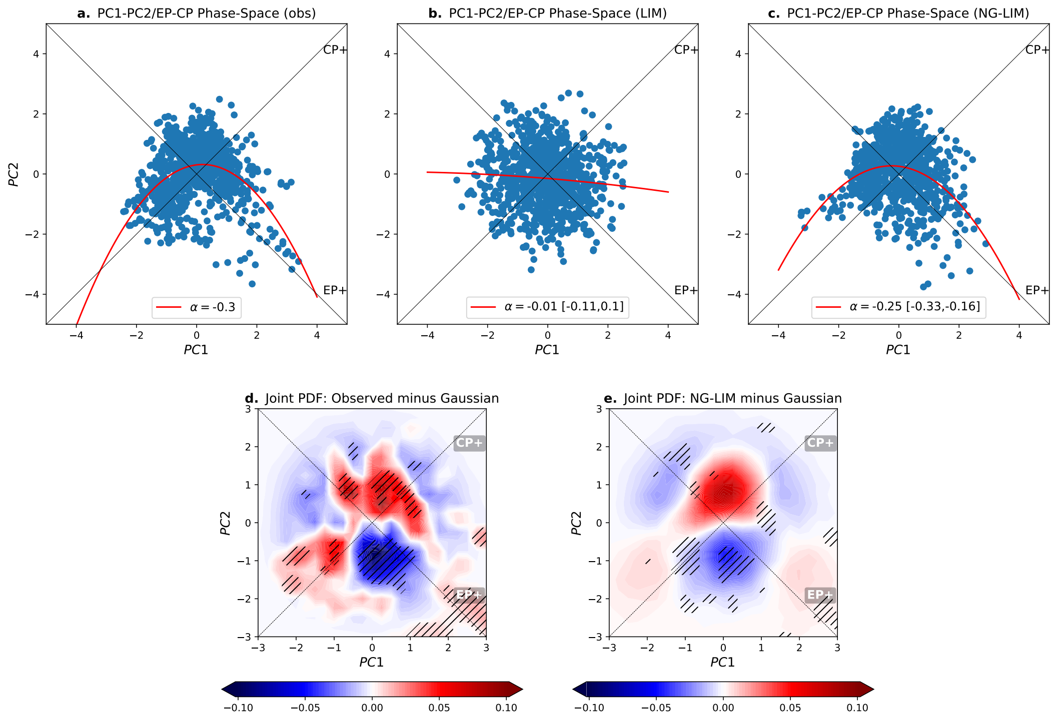

Figure 2. PC1/PC2-EP/CP indices chart in **a.** observations (1948-2022), **b.** generated by the LIM, and **c.** generated by the NG-LIM. In the case of the LIM and NG-LIM we display the epoch of the same length as observations (75 years) corresponding to the median generated α. The distribution of α in the LIM and NG-LIM is shown in Fig. S5. **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. 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.

plies that regardless of the epoch, the NG-LIM simulates a curved relationship with stronger CP La Niñas and EP El Niños (Fig. 2c). When sampling only Decembers (Fig. S5a), the NG-LIM similarly replicates the observed relationship (Fig. S5c), but as expected with a stronger variability in terms of α across epochs (Fig. S6b).

Fig. 2a shows that observed states cluster in a particular way in the PC1-PC2 plane, which is illustrated by calculating the joint probability distributions (Fig. 2d; see also Fig. S7a). Likewise, the NG-LIM successfully puts more probability for strong CP La Niñas and extreme EP El Niños (Fig. 2e; see also Fig. S7c). 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 (Figs. 2d,e). In particular, 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), and which are characterized by low absolute values of PC1 (See Fig. 3e of (Martinez-Villalobos et al., 2024))—, with fewer but more extreme coastal El Niño events, is better simulated by the NG-LIM (Fig. 2d,e). In contrast, the standard LIM has a Gaussian joint-distribution with peak in probability at the origin ($PC1 = 0$, $PC2 = 0$; Fig. S7b). 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 (Fig. 3c-f). In all these cases, observations fall outside the 5th-95th range generated by the standard LIM.

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 (Fig. 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 Central Pacific. As expected, the standard LIM does not simulate a skewness pattern whatsoever (not shown). 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 (Fig. 3b).

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. 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. (Penland & Sardeshmukh, 1995; Zanna & Tziperman, 2005; Vimont et al., 2014; Martinez-Villalobos & Vimont, 2016)) 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 (eq. 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

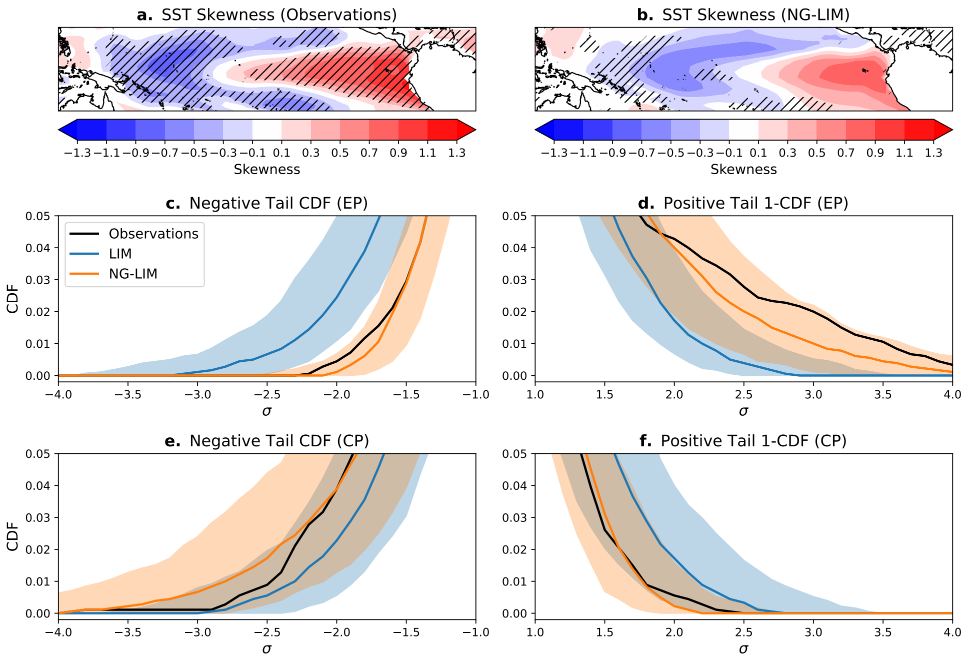

Figure 3. **a.** Spatial map of SST skewness coefficient in observations. **b.** Same as **a.** but for the whole 10,000yr integration of the NG-LIM. The cross-hatching in **a** (**b**) shows regions where the observed SSTA skewness is outside the 5th-95th percentile range of skewness generated by the standard LIM (NG-LIM) across epochs. Figure S8 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 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 CP index. Figure S9 shows the whole EP and CP indices CDFs.

data onto the optimal in Fig. 4d is highly correlated with the Niño 3.4 index evolution months later (Fig. S10).

Fig. 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-months later (Fig. 4b and 4d). However, there are some differences in the evolution; warm optimals tend to develop stronger anomalies in the east and cold optimals stronger anomalies in the west (Fig. 4f), resembling the antisymmetry of El Niño and La Niña events in observations (Fig. 4e). As expected, the standard LIM does not generate this difference, instead evolving cold and warm optimals symmetrically (Fig. S11). The 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 observations (Okumura & Deser, 2010) (Fig. 4g).

4 Summary and Discussion

The bulk of ENSO modeling is done in the forward sense, i.e., one starts from fundamental equations derived from the system's physics (plus parametrizations), hoping to yield insight into the real phenomenon. An example of this, are coupled global climate models, such as the Community Earth System Model (Hurrell et al., 2013; Danabasoglu et al., 2020). A pragmatic and cost effective alternative is the use of data-driven models such as linear inverse models (LIMs). The starting point for these models is the observed data, from which the task is to reverse-engineer the best model consistent with this data. These approaches, and especially the LIM methodology, have been useful to study many aspects of ENSO behavior, including its irregularity (Penland & Sardeshmukh, 1995; Berner et al., 2018), asymmetry (Martinez-Villalobos et al., 2019), patterns of growth and decay (Penland & Sardeshmukh, 1995; Vimont et al., 2014; Capotondi & Sardeshmukh, 2015; Vimont et al., 2022), predictability (Mason & Mimmack, 2002; Newman & Sardeshmukh, 2017), statistical significance of epochs' changes (Capotondi & Sardeshmukh, 2017; Martinez-Villalobos et al., 2019), and associated impact on marine heatwaves (Capotondi et al., 2022; Gregory et al., 2024). However, LIMs in their standard form do not fully capture ENSO asymmetry nor diversity. Here, we propose an empirical approach to deal with this inherent limitation of the standard LIM. It consists in first transforming the variables comprising the relevant state-vector to near-Gaussianity and then calculating a standard LIM in those transformed variables. 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 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.

The NG-LIM proposed here provides 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. (Newman & Sardeshmukh, 2017; Lenssen et al., 2024; Schlör et al., 2024)). Additionally, it could be worth investigating whether the NG-LIM produces more interdecadal ENSO amplitude modulation and more tropical Pacific decadal variability (TPDV; (Newman et al., 2016; Capotondi et al., 2023)) 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. In addition to that, 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., (Vimont, 2005;

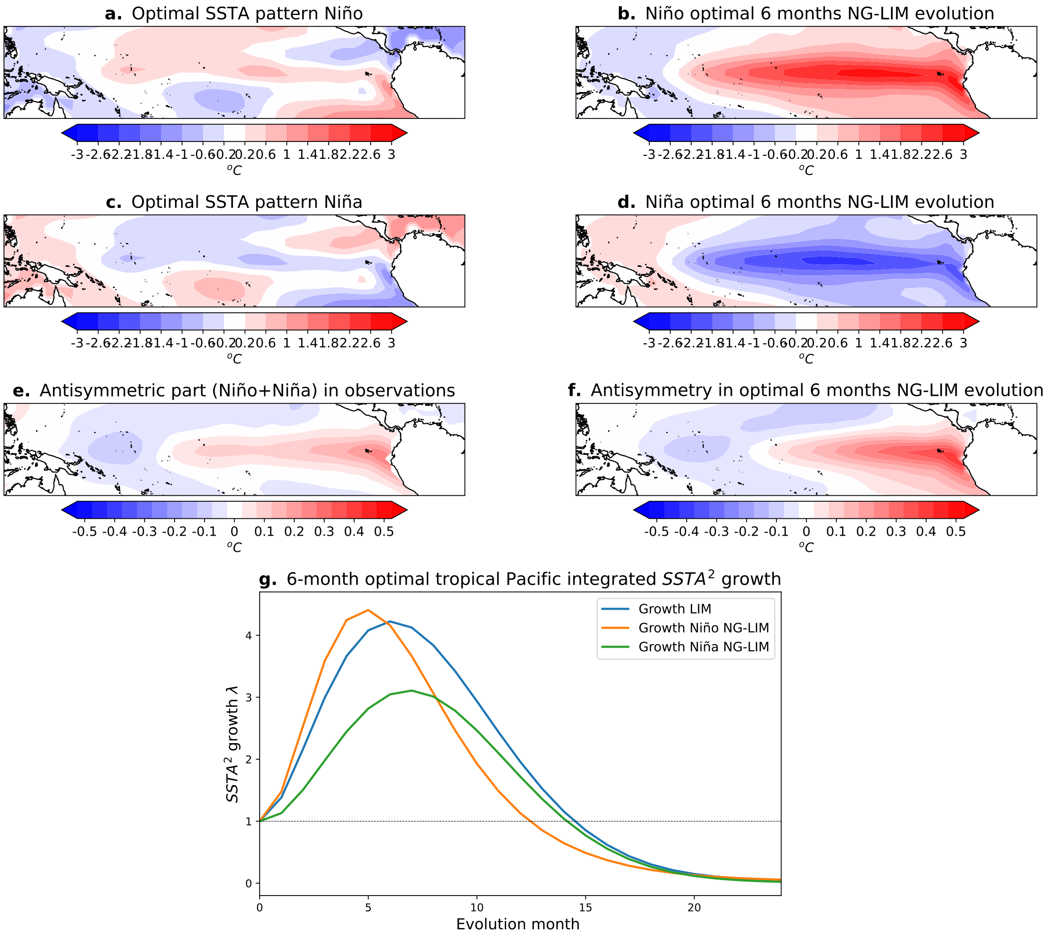

Figure 4. **a.** (**c.**) 6-months optimal initial condition, as calculated by the standard LIM, that evolves into an El Niño (La Niña) event months later. **b.** (**d.**) 6-months evolution of the optimal in **a** (**b**) by the NG-LIM. **e.** Composite of the spatial pattern of the antisymmetric part of ENSO 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^oC (-0.5^oC). **f.** Asymmetry between the NG-LIM evolution of warm and cold optimals. This is calculated as pattern **b** plus pattern **d**. **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).

Ogata et al., 2013; Atwood et al., 2017; Fedorov et al., 2020; Power et al., 2021)). Furthermore, the NG-LIM could be used to provide process-oriented metrics (e.g., (Maloney et al., 2019; Leung et al., 2022)) to evaluate global climate model performance in simulating ENSO events (e.g., (C. Chen et al., 2017; Planton et al., 2021)). For example, the exponents of the transformation could provide such a metric describing the degree of Gaussianity of model variables relative to observations. In addition, a better understanding of the relationship between deterministic and stochastic operators and the quadratic relationship between the two leading PCs may allow to assess why this relationship is not well reproduced in several models (cf. (Karamperidou et al., 2017). In terms of the NG-LIM construction, future improvements could include: i. extending the state-vector to also include a measure of ocean memory (Xue et al., 2000; Newman, Alexander, & Scott, 2011; Capotondi & Sardeshmukh, 2015), such as ocean heat content, that could conceivably extend the horizon of ENSO predictability; ii. calculating a cyclo-stationary version of the NG-LIM (Penland, 1996; OrtizBeviá, 1997; Johnson et al., 2000; Shin et al., 2021; Vimont et al., 2022; Wang et al., 2023), such that the interplay between seasonality and ENSO diversity could be better analyzed; iii. 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 Eastern Pacific events could be better identified and also perhaps predicted (e.g., (Vimont et al., 2014, 2022)).

There is some debate on whether ENSO could be better conceived as a nonlinear deterministic process with or without stochastic forcing (e.g., (Cane & Zebiak, 1985; Schopf & Suarez, 1988; Battisti & Hirst, 1989; Neelin, 1991; Neelin & Jin, 1993; Neelin et al., 1998), 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., (Sardeshmukh & Sura, 2009; Martinez-Villalobos et al., 2019) that also to some extent improve 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, while second and more importantly, given that the number of variables that need to be estimated is the same as in the standard LIM, it requires much less data than the aforementioned models to be fitted (cf. (Martinez-Villalobos et al., 2018, 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 1948 to 2022. An example of EP and CP indices calculated from these integrations are shown in Figs. S12 and S13. These integrations can be used for a variety of purposes, including 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.

As Applicable – Inclusion in Global Research Statement

Acknowledgments

CM acknowledges support from Proyecto ANID Fondecyt Postdoctorado código 3200621, and Data Observatory Foundation ANID Technology Center No. DO210001 and Fondecyt Regular N°1231174. AC was supported by the NOAA Climate Program Office Climate Variability and Predictability Program Award No. NA24OARX431C0024-T1-0, and by DOE Award No. DE-SC0023228. The National Center for Atmospheric Research is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. 1852977. BD acknowledges support from ANID (Concurso de Fortalecimiento al Desarrollo Cientıfico de Centros Regionales 2020-R20F0008-CEAZA, Anillo Eclipse ACT210071, 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. NJH acknowledges funding from the ARC Centre of Excellence for the Weather of the 21st Century (CE230100012) and Australia's National Environmental Science Program Climate Systems Hub.

Open Research Section

The LIM and NG-LIM integrations can be accessed at `https://doi.org/10.5281/zenodo.14775713`.

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.

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.

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

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.

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.

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.

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.

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

Cai, W., Wang, G., Dewitte, B., Wu, L., Santoso, A., Takahashi, K., … McPhaden, M. J. (2018). Increased variability of eastern Pacific El Niño under greenhouse warming. *Nature*, *564*(7735), 201–206.

Cane, M. A., & Zebiak, S. E. (1985). A Theory for El Niño and the Southern Oscillation. *Science*, *228*(4703), 1085–1087.

Capotondi, A., McGregor, S., McPhaden, M. J., Cravatte, S., Holbrook, N. J., Imada, Y., … Xu, T. (2023). Mechanisms of tropical Pacific decadal variability. *Nature Reviews Earth & Environment*, *4*(11), 754–769.

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.

Capotondi, A., & Sardeshmukh, P. D. (2015). Optimal precursors of different types of ENSO events. *Geophysical Research Letters*, *42*(22), 9952–9960.

Capotondi, A., & Sardeshmukh, P. D. (2017). Is El Niño really changing? *Geophysical Research Letters*, *44*(16), 8548–8556.

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., ... Yeh, S.-W. (2015). Understanding ENSO Diversity. *Bulletin of the American Meteorological Society*, *96*(6), 921–938.

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.

Chen, C., Cane, M. A., Wittenberg, A. T., & Chen, D. (2017). ENSO in the CMIP5 Simulations: Life Cycles, Diversity, and Responses to Climate Change. *Journal of Climate*, *30*(2), 775–801.

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.

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.

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.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916.

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.

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

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.

DiNezio, P. N., & Deser, C. (2014). Nonlinear Controls on the Persistence of La Niña. *Journal of Climate*, *27*(19), 7335–7355. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/19/jcli-d-14-00033.1.xml`

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.

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.

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.

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.

Gregory, C. H., Artana, C., Lama, S., León-FonFay, D., Sala, J., Xiao, F., ... Holbrook, N. J. (2024). Global Marine Heatwaves Under Different Flavors of ENSO. *Geophysical Research Letters*, *51*(20), e2024GL110399.

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.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... Zhang, H.-M. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, *30*(20), 8179–8205.

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.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., ... Marshall, S. (2013). The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360.

Jin, F.-F., Chen, H.-C., Zhao, S., Hayashi, M., Karamperidou, C., Stuecker, M. F., ... Geng, L. (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*(18), 3327–3335.

Kao, H.-Y., & Yu, J.-Y. (2009). Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *Journal of Climate*, *22*(3), 615–632.

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.

Karnauskas, K. B. (2013). Can we distinguish canonical El Niño from Modoki? *Geophysical Research Letters*, *40*(19), 5246–5251.

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.

Kondrashov, D., Kravtsov, S., Robertson, A. W., & Ghil, M. (2005). A Hierarchy of Data-Based ENSO Models. *Journal of Climate*, *18*(21), 4425–4444.

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.

Lee, J., Planton, Y. Y., Gleckler, P. J., Sperber, K. R., Guilyardi, E., Wittenberg, A. T., ... Pallotta, G. (2021). Robust Evaluation of ENSO in Climate Models: How Many Ensemble Members Are Needed? *Geophysical Research Letters*, *48*(20), e2021GL095041.

Lenssen, N., DiNezio, P., Goddard, L., Deser, C., Kushnir, Y., Mason, S. J., ... Okumura, Y. (2024). Strong El Niño Events Lead to Robust Multi-Year ENSO Predictability. *Geophysical Research Letters*, *51*(12), e2023GL106988.

Leung, L. R., Boos, W. R., Catto, J. L., DeMott, C. A., Martin, G. M., Neelin, J. D., ... Zhou, Y. (2022). Exploratory Precipitation Metrics: Spatiotemporal

Characteristics, Process-Oriented, and Phenomena-Based. *Journal of Climate*, *35*(12), 3659–3686.

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.

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. (Section: Journal of Climate)

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.

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

Maloney, E. D., Gettelman, A., Ming, Y., Neelin, J. D., Barrie, D., Mariotti, A., . . . Zhao, M. (2019). Process-Oriented Evaluation of Climate and Weather Forecasting Models. *Bulletin of the American Meteorological Society*, *100*(9), 1665–1686.

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.

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.

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.

Mason, S. J., & Mimmack, G. M. (2002). Comparison of Some Statistical Methods of Probabilistic Forecasting of ENSO. *Journal of Climate*, *15*(1), 8–29.

McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2006). ENSO as an Integrating Concept in Earth Science. *Science*, *314*(5806), 1740–1745.

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.

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.

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: Oceans*, *103*(C7), 14261–14290.

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.

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Lorenzo, E. D., . . . Smith, C. A. (2016). The Pacific Decadal Oscillation, Revisited. *Journal of Climate*, *29*(12), 4399–4427.

Newman, M., Alexander, M. A., & Scott, J. D. (2011). An empirical model of tropical ocean dynamics. *Climate Dynamics*, *37*(9), 1823–1841.

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.

Newman, M., Shin, S.-I., & Alexander, M. A. (2011). Natural variation in ENSO flavors. *Geophysical Research Letters*, *38*(14), L14705.

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.

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.

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: Atmospheres*, *102*(D12), 13473–13486.

Penland, C. (1989). Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. *Monthly Weather Review*, *117*(10), 2165–2185.

Penland, C. (1996). A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, *98*(2), 534–558.

Penland, C., & Magorian, T. (1993). Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, *6*(6), 1067–1076.

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.

Penland, C., & Sardeshmukh, P. D. (1995). The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, *8*(8), 1999–2024.

Planton, Y. Y., Guilyardi, E., Wittenberg, A. T., Lee, J., Gleckler, P. J., Bayr, T., ... Voldoire, A. (2021). Evaluating Climate Models with the CLIVAR 2020 ENSO Metrics Package. *Bulletin of the American Meteorological Society*, *102*(2), E193–E217.

Planton, Y. Y., Lee, J., Wittenberg, A. T., Gleckler, P. J., Guilyardi, , McGregor, S., & McPhaden, M. J. (2024). Estimating Uncertainty in Simulated ENSO Statistics. *Journal of Advances in Modeling Earth Systems*, *16*(9), e2023MS004147.

Power, S., Lengaigne, M., Capotondi, A., Khodri, M., Vialard, J., Jebri, B., ... Henley, B. J. (2021). Decadal climate variability in the tropical Pacific: Characteristics, causes, predictability, and prospects. *Science*, *374*(6563), eaay9165.

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.

Sardeshmukh, P. D., & Sura, P. (2009). Reconciling Non-Gaussian Climate Statistics with Linear Dynamics. *Journal of Climate*, *22*(5), 1193–1207.

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` doi: 10.48550/arXiv.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.

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.

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 Re-*

search Letters, *38*(10). doi: 10.1029/2011GL047364

Taschetto, A. S., & England, M. H. (2009). El Niño Modoki Impacts on Australian Rainfall. *Journal of Climate*, *22*(11), 3167–3174.

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.

Thompson, C. J., & Battisti, D. S. (2001). A Linear Stochastic Dynamical Model of ENSO. Part II: Analysis. *Journal of Climate*, *14*(4), 445–466.

Thual, S., & Dewitte, B. (2023). ENSO complexity controlled by zonal shifts in the Walker circulation. *Nature Geoscience*, *16*(4), 328–332.

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.

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., … Zhang, X. (2018). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545.

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.

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.

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.

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: Oceans*, *103*(C7), 14241–14259.

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.

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*(12), L12702.

Xue, Y., Leetmaa, A., & Ji, M. (2000). ENSO Prediction with Markov Models: The Impact of Sea Level. *Journal of Climate*, *13*(4), 849–871.

Yeo, I.-K., & Johnson, R. A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, *87*(4), 954–959.

Zanna, L., & Tziperman, E. (2005). Nonnormal Amplification of the Thermohaline Circulation. *Journal of Physical Oceanography*, *35*(9), 1593–1605.