A Hybrid Deep Learning Model for El Niño Southern Oscillation Dynamics

Jakob Schlör,[a]

Jannik Thümmel,[a] Antonietta Capotondi,[b,c] Matthew Newman,[a] Bedartha Goswami,[a]

[a] *Machine Learning in Climate Science, University of Tübingen, Germany*

[b] *Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO*

[c] *NOAA Physical Sciences Laboratory, Boulder, CO, USA*

Corresponding author: Jakob Schlör, jakob.schloer@uni-tuebingen.de

ABSTRACT: Event-to-event differences of the El Niño Southern Oscillation (ENSO) result in different patterns of extreme climate conditions globally, which requires ENSO forecasts accurately predict both the likelihood and the type of an event. In this study, we examine to which extent ENSO dynamics may be represented by multivariate linear dynamics and, relatedly, whether predictable nonlinearities must be accounted for or may be treated stochastically. By combining Recurrent Neural Networks (RNNs) with the Linear Inverse Model (LIM), we evaluate the role of predictable nonlinearities and non-Markovian dynamics in tropical Pacific sea surface temperature anomalies. Our results indicate that incorporating nonlinearities substantially improves forecast accuracy, especially for the western tropical Pacific with a lag of 9 to 18 months. We identify the main source of nonlinearity to the asymmetry between warm and cold events. Additionally, we find that the predictability of our Hybrid-model can be assessed using LIM's theoretical skill, a feature not achievable with purely deep learning approaches.

SIGNIFICANCE STATEMENT: Enter significance statement here, no more than 120 words. See www.ametsoc.org/index.cfm/ams/publications/author-information/significance-statements/ for details.

1. Introduction

Seasonal forecasting worldwide rests largely on the predictability of tropical sea surface temperature (SST) anomalies. To tackle the challenges of seasonal forecasting, models essentially aim to emulate the climate systems's dynamics that can be separated into two classes. The first tries to represent the complex dynamics of the entire system, as in global climate models (GCM), which offer detailed simulations but at a high computational cost. Alternatively, simplistic models that are tailored to the specific problem and are less resource-intensive are used for forecasting. With the availability of high-quality observational data, a new class of models, known as hybrid models, has emerged. These models enhance accuracy by adjusting for errors in numerical models using data-driven corrections.

In this work, we propose a Hybrid-model designed to forecast the El Niño Southern Oscillation (ENSO). ENSO, the dominant mode of annual climate variability, is a prominent example of seasonal forecasting. Its warm and cold events are characterized by tropical SSTA, which exhibit a rich diversity in their spatial structure, temporal evolution, and impact on extreme weather conditions worldwide (Capotondi et al. 2015; Timmermann et al. 2018). Subsequently early forecasts of not only the ENSO event likelihood but also its spatial structure are of great value for agriculture and society globally (Callahan and Mankin 2023; Strnad et al. 2022).

Despite extensive model development, the tropical SST forecast skill of the operational North American Multi-Model Ensemble (NMME) (Kirtman et al. 2014) - a collective of eight coupled atmosphere-ocean models - remains close to that of a vastly simpler linear inverse model (LIM) derived from observed covariances of SST, sea surface height, and wind fields (Newman and Sardeshmukh 2017). The LIM captures the essence of the predictable SST dynamics with its forecast skill closely matching the NMME and slightly trailing the potential skill estimated using the LIM's forecast signal-to-noise ratios. This suggests that the scope for further skill improvement is small in most regions, except in the western equatorial Pacific where the NMME skill is currently much higher than the LIM skill. Addressing this, our study builds upon the LIM to enhance forecast

accuracy. We combine the LIM with a deep neural network, focusing on learning the remaining predictable nonlinearities in the monthly evolution of tropical SSTA, thus formulating a Hybrid-model.

Introduced by Penland and Sardeshmukh (1995), the LIM represents the slower-varying ocean dynamic as stochastically forced by the rapidly varying atmosphere using a linear stochastic differential equation subject to white noise. The LIM, despite its linear simplicity, demonstrates annual-to-seasonal forecasting skill on par with coupled ocean-atmosphere models for ENSO and other climate processes like the North Atlantic Oscillation (Newman and Sardeshmukh 2017; Albers and Newman 2021). Enhancements to the LIM, including the cyclostationary (CS)-LIM which captures ENSO seasonal phase locking (Shin et al. 2021), ocean memory (Chen et al. 2016), and state-dependent noise (Martinez-Villalobos et al. 2018), have been proposed. While those variants yield improved predictability, their dynamics are still linear. Takahashi et al. (2011) and other (Takahashi and Dewitte 2016; Okumura 2019; Geng and Jin 2022), however, propose a substantial impact of nonlinear processes on the Tropical Pacific, especially given the skewed SSTA distribution with more extreme warm temperatures.

Deep learning (DL) models, capable of emulating intricate nonlinear functions, have been shown to yield skillful ENSO forecasts (Ham et al. 2019; Petersik and Dijkstra 2020; Cachay et al. 2021; Zhou and Zhang 2023). However, these models suffer from a lack of interpretability and reliable predictability assessment. Even probabilistic DL models are generally overconfident with a weak to no spread-to-skill relationship (Guo et al. 2017).

In contrast, hybrid models leverage the strengths of both simple and interpretable base models and DL models, which simplify the learning objective for the DL models, enhancing data efficiency and predictive skill (van Straaten et al. 2023; Rasp and Lerch 2018). In our research, we utilize the LIM as the base model and use a recurrent neural network (RNN), specifically an LSTM, to capture the residuals of the LIM forecast to the target data. The strategy of combining a simple linear empirical model with a DL-model has been previously suggested by Goel et al. (2017). Specifically, they train an RNN upon a vector autoregressive model for time-series forecasting. Similarly, Wang et al. (2021); Zhou and Zhang (2022) integrated the principal oscillation pattern, akin to the LIM, with an RNN for forecasting the Nino3.4 and ONI index, respectively. Our method aligns conceptually with the work by Rodrigues et al. (2021), who proposes a neural network that

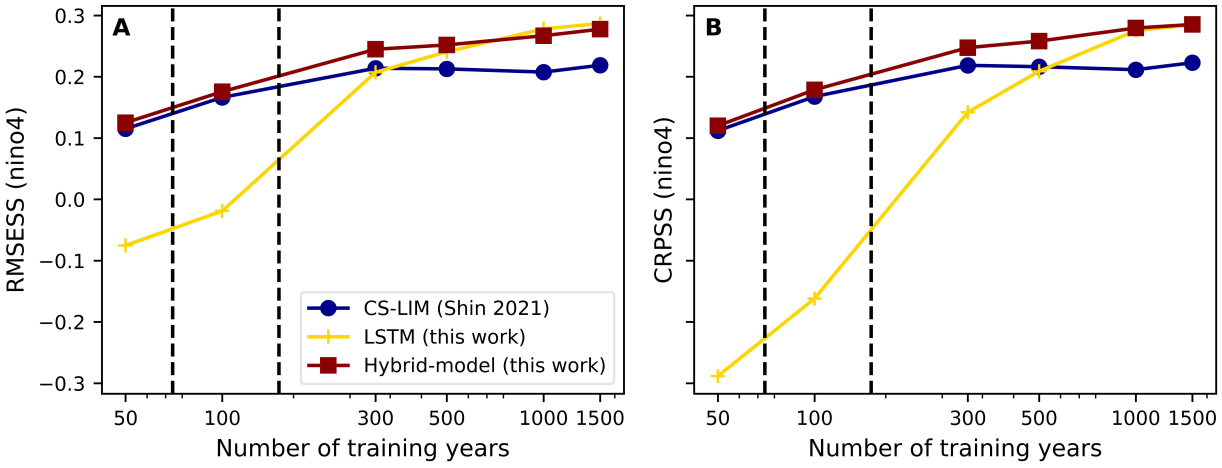


FIG. 1. **Forecast skill over training dataset length.** The average RMSE (A) and CRPS (B) skill scores of Nino4 index for $\tau = 9$-month forecasts from the CS-LIM, LSTM, and Hybrid model, each trained on 50 to 1500 years of monthly SSTA and SSHA data. Monthly climatology serves as the reference for the scores, as described in section 2.3. With only 70 or 150 years of data, akin to the length of ORAS5 and CERA-20C reanalysis products (denoted as dashed lines), the LSTM's forecast skill is inferior to the CS-LIM and Hybrid models. The LSTM attains a skill level comparable to these models only when trained on over 500 years of data, underlining the extensive data requirements for deep learning emulators.

resembles a Resnet block with the difference that, instead of the identity, the bypass is the LIM operator (in their work called dynamical mode decomposition), applied to global SSTA.

However, we diverge from existing methodologies by first employing a seasonal varying linear base model, the CS-LIM, and secondly using an LSTM network, that captures both nonlinear as well as non-markovian dynamics. Additionally, our models generate probabilistic forecasts, where each ensemble member presents a spatio-temporal forecast of the tropical Pacific. We carefully analyze the learned residual dynamics, which points us toward the role of nonlinearities in the tropical Pacific. We, further, demonstrate that optimal initial conditions derived from the CS-LIM enable us to discern between high- and low-skill cases, thus, estimating the Hybrid-models predictability.

Our analysis is conducted on 2000 years of pre-industrial control simulations from the Community Earth System Model version 2 (CESM2) (Danabasoglu et al. 2020), a state-of-the-art coupled global climate model known for its reasonable representation of ENSO diversity. However, it is important

to note that CESM2 exhibits biases in the tropical Pacific, such as the exaggerated variability of SST in the western Pacific. The extensive length of the simulation run enables us to estimate the data requirements for our base model, full DL model, and Hybrid model to accurately forecast Tropical SSTA. Fig. 1 displays the average Root Mean Square Error (RMSE) skill score and Continuous Ranked Probability Score (CRPS) skill score over the test set for a 12-month lag time forecast. This evaluation was performed for the CS-LIM, the LSTM, and the Hybrid model, all of which were trained on 50 to 1500 years of monthly SSTA and SSHA. For computing these skill scores, we used the monthly climatology as the reference model, as described in section e. When the models are fitted with only 70 or 150 years of data, similar to the length of available reanalysis products of ORAS5 and CERA-20C, both the CS-LIM and the Hybrid model forecast skill is significantly larger than the LSTM's skill that is even worse than a climatological forecast. It is only with over 500 years of training data that the LSTM's skill level is on par with the CS-LIM and the Hybrid model, highlighting the large amount of data required to train the full deep learning emulator in comparison to the Hybrid model.

Our paper is organized as follows: Section 2 introduces the CESM2 data and its preprocessing, followed by section 3, which details the LIM, LSTM, and Hybrid model setup and architecture. Section 4 presents our Hybrid model's enhanced skill and predictability, along with an analysis of ENSO asymmetry. The paper concludes in section 5 with a discussion of the results.

2. Data

Training DL models for ENSO prediction with monthly data is limited by the short observational record. To circumvent this, we use the 2000-year CESM2 preindustrial control simulation (Danabasoglu et al. 2020), focusing on monthly SST and SSH data in the tropical Pacific region (130°E - 70°W, 31°S - 32°N), which we interpolate to a resolution of 1°x1°. SSH is a proxy for the upper ocean temperatures and the thermocline depth. Despite the lack of external forcing in the control simulation, we observe a trend in SST data. We linearly detrend the data and remove the seasonal cycle by subtracting the monthly climatology. As SSTA and SSHA differ in units and scales, standardization is performed before model training. Both our LIM and Hybrid-model requires to reduce the dimensionality, which is achieved using EOF analysis. The dataset is divided into

training (75%, year 1-1500), validation (15%, years 1500-1800), and test set (10%, 1800-2000), with the validation set used for refining the hyperparameters of our models.

3. Methods

The objective of our study is to accurately predict SSTA and SSHA fields for a specified forecast time. We define the stacked variable fields at a given time t as $\mathbf{x}(t) = (\mathbf{x}_{\text{SSTA}}, \mathbf{x}_{\text{SSHA}})(t)$, where each field spans the tropical Pacific $\mathbf{x}_{\text{SSTA/SSHA}}(t) \in \mathbb{R}^{N_{lat} \times N_{lon}}$. Our task is to estimate a function f which depends on the learnable parameters θ, which aims to predict the future state $\hat{\mathbf{x}}(t+\tau)$ based on past states $\{\mathbf{x}(t), \mathbf{x}(t-1), ..., \mathbf{x}(t-h)\}$, i.e.

$$\hat{\mathbf{x}}_m(t+\tau) = f_\theta\left(\tau, \mathbf{x}(t), \mathbf{x}(t-1), ..., \mathbf{x}(t-h)\right), \tag{1}$$

where τ represents the forecast lag time in months and h is the number of historical steps considered before initiating the forecast. All our model forecasts consist of m ensemble members, which allows us to access the prediction uncertainty. The dynamics of the tropical Pacific Ocean show a strong seasonal phase locking. We therefore condition f on the month of the year, $m(t)$. The month conditioning depends on the model architecture and will be discussed for each model separately.

a. Empirical Orthogonal Function (EOF)

Estimating the linear operator of the LIM requires a matrix inversion of the number of input dimensions. The matrix inversion is intractable for the full spatial fields of SSTA and SSHA. For this reason, each state $\mathbf{x}(t)$ is transformed into a lower-dimensional state $\mathbf{z}(t) = (\mathbf{z}_{\text{SSTA}}, \mathbf{z}_{\text{SSHA}})$. Dimensionality reduction of the SSTA and SSHA fields in the tropical Pacific is achieved through EOF analysis, utilizing the first 20 Principal Components (PCs) for SSTA and the first 10 PCs for SSHA. Including higher-order PCs does not affect our results. Forecasting in the lower dimensional space is then equivalently conducted on these PCs, as formulated: $\hat{\mathbf{z}}_m(t+\tau) = g_\theta\left(\tau, \mathbf{z}(t), \mathbf{z}(t-1), ..., \mathbf{z}(t-h)\right)$. For analysis and evaluation, we transform our forecast back to grid space. To adequately replicate the high spatial frequencies of the input fields, we add variability by randomly sampling from the higher-order PCs (20-300) for both SSTA and SSHA at

each timestep. These random loadings are then combined with their respective EOFs and added to the forecast fields, ensuring a closer match to the spatial intricacies of the original data.

b. Linear Inverse Model

The LIM describes the dynamic of the tropical Pacific as a multivariate linear system subject to stochastic forcing from the atmosphere. The underlying dynamics of such a system is described by a linear stochastic differential equation,

$$\frac{d\mathbf{z}}{dt} = \mathbf{L}\mathbf{z} + \zeta \tag{2}$$

where $\mathbf{L}$ is the linear operator describing the dynamics of $\mathbf{z}$ and $\zeta \sim \mathcal{N}(0, \mathbf{Q})$ is a noise vector that is white in time but spatially correlated given by the covariance matrix $\mathbf{Q}$. Forecasts of $\mathbf{z}$ for a lag time τ are given by the transition probability as,

$$p\left(\mathbf{z}(t+\tau)|\mathbf{z}(t)\right) = \mathcal{N}\left(\mathbf{z}(t+\tau); \mu_\tau(t), \mathbf{\Sigma}_\tau\right) \tag{3}$$

$$\text{with} \qquad \mu_\tau(t) := e^{\mathbf{L}\tau}\mathbf{z}(t) \qquad \text{and} \qquad \mathbf{\Sigma}_\tau := \int_0^\tau e^{\mathbf{L}s}\mathbf{Q}\mathbf{Q}^T e^{\mathbf{L}^T s} ds, \tag{4}$$

where $\mu_\tau(t)$ is the infinite ensemble mean forecast and $\mathbf{\Sigma}_\tau$ is the forecast covariance matrix.

Penland and Sardeshmukh (1995) show that the linear operator $\mathbf{L}$ and the noise covariance $\mathbf{Q}$ can be estimated from the data under two assumptions. First, the system has to be statistically stationary which allows us to write the Fluctuation-Dissipation relationship as

$$\mathbf{L}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{L}^T + \mathbf{Q} = 0, \tag{5}$$

where $\mathbf{C}(0) = \langle \mathbf{z}(t)\mathbf{z}^T(t) \rangle$ is the spatial covariance matrix. Secondly, the autocorrelation of the system decays with lag time τ which can be expressed using the time-lag covariance matrix $\mathbf{C}(\tau) = \langle \mathbf{z}(t+\tau)\mathbf{z}^T(t) \rangle$ as

$$\lim_{\tau\to\infty} \mathbf{C}(\tau)\mathbf{C}(0) = 0 \Rightarrow \mathbf{C}(\tau) = e^{\mathbf{L}\tau}\mathbf{C}(0), \tag{6}$$

where $\mathbf{G}(\tau) := \exp(\mathbf{L}\tau)$ is the Greens function that must tend to zero for long lag times. Typically, both assumptions hold for detrended anomaly data of a chaotic system.

Once $\mathbf{L}$ and $\mathbf{Q}$ are estimated from the data, we obtain forecast trajectories from an initial time t to $t+T$, by numerically integrating eq. 2 using the forward Euler-method with incremental update steps δ as

$$\mathbf{z}(t+\delta) = \mathbf{z}(t) + \mathbf{L}\mathbf{z}(t)\delta + \zeta\sqrt{\delta}, \tag{7}$$

where $\zeta \sim \mathcal{N}(0, \mathbf{Q})$ is a random sample from the noise distribution. We create n ensemble member trajectories from t to $t+T$ when integrating the system n-times. The infinite ensemble member mean is given by $\mu(t)$ in eq. 3.

In the equatorial Pacific, the variance in SSTA shows a distinct annual pattern with low variance during the boreal spring and high variance in the boreal winter. This peak in winter variance aligns with the occurrence of the most intense warm and cold ENSO events, a phenomenon referred to as "ENSO phase locking" (Rasmusson and Carpenter 1982). Shin et al. (2021) showed that including seasonality in the LIM improves its forecast reliability. Their cyclostationary (CS)-LIM involves estimating unique linear operators and noise covariances for each month, represented as j. The numerical integration of the stationary (ST)-LIM outlined in eq. 7 changes to

$$\mathbf{z}_j(t+\delta) = \mathbf{z}_j(t) + \mathbf{L}_j^{CS}\mathbf{z}_j(t)\delta + \zeta_j^{CS}\sqrt{\delta}, \tag{8}$$

where $\zeta_j^{CS} \sim \mathcal{N}\left(0, \mathbf{Q}_j^{CS}\right)$. Both $\mathbf{L}_j^{CS}$ and $\mathbf{Q}_j^{CS}$ are consant within each month j. The CS-LIM forms the base model of our Hybrid-model outlined in the following.

c. Hybrid-model

We introduce a novel Hybrid-model that combines the LIM with an LSTM network. While the LIM captures the predictable linear dynamics, the LSTM learns the residuals between the LIM predictions and the actual data, thus the nonlinear dynamics. Our methodology is schematically detailed in Fig. 2.

During inference, we project the initial state of the tropical Pacific, $\mathbf{x}(t)$, onto the leading EOFs and employ the LIM to predict future states over $\tau = 1, \ldots, T$ timesteps. For each timestep, $t+\tau$, we predict a correction, $\hat{\mathbf{z}}_{\text{res}}(t+\tau)$, to the LIM forecast, $\hat{\mathbf{z}}_{\text{LIM}}(t+\tau)$. The final forecast is thus defined as:

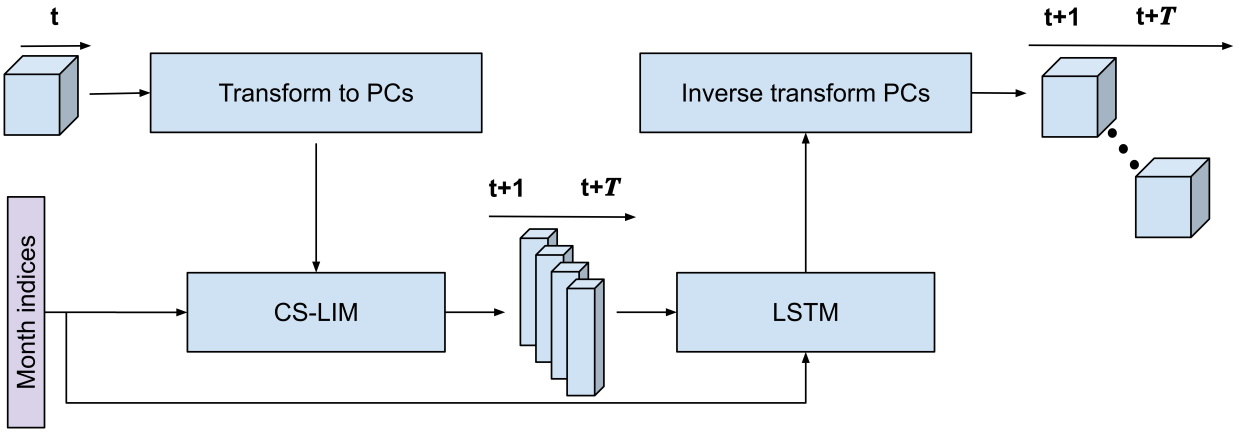


FIG. 2. **Schematic Representation of the Hybrid-Model**. First, the initial state at time t is projected onto the PCs. This is followed by an ensemble forecast using the CS-LIM which is conditioned on the forecast months. Subsequently, the LSTM adjusts each ensemble member of the linear CS-LIM forecast. Finally, the refined forecast is transformed back to grid space by multiplying the PCs with the respective EOF patterns.

$$\hat{\mathbf{z}}(t+\tau) = \hat{\mathbf{z}}_{\mathrm{LIM}}(t+\tau) + \hat{\mathbf{z}}_{\mathrm{res}}(t+\tau). \tag{9}$$

The nonlinear correction is modeled by the LSTM as $\hat{\mathbf{z}}_{\mathrm{res}}(t+\tau) = g_\theta(\tau, \hat{\mathbf{z}}_{\mathrm{LIM}}(t+1), \ldots, \hat{\mathbf{z}}_{\mathrm{LIM}}(t+\tau))$, where θ represents the learnable parameters of the network. The LSTM is selected not only for capturing nonlinear relationships inherent in deep neural networks but also for its ability in learning non-markovian dynamics.

To include seasonality within the LSTM, we introduce an affine transformation to its latent state, $\mathbf{h}$, as follows:

$$\mathbf{h}_{\mathrm{FiLM}} = (1+\alpha)\mathbf{h} + \beta, \tag{10}$$

where α and β represent embeddings for each month, enabling the network to adapt its latent state dynamically to seasonal variations.

The LSTM is trained by first determining the LIM operators using the training dataset, followed by conducting m-ensemble member hindcasts. The LSTM parameters, θ, are then optimized to minimize the continuous ranked probability score (CRPS) between the observed data, $\mathbf{z}(t+\tau)$, and the hindcast, $\hat{\mathbf{z}}(t+\tau)$, across all ensemble members and lag times $\tau \in [1, T]$.

$$l\left(\hat{Z}_t, z_t\right) = \sum_{\tau=1}^{T} \mathrm{CRPS}\left(\mathbf{z}(t+\tau), \hat{Z}(t+\tau)\right) \tag{11}$$

Here, z_t denotes the sequence of ground truth data, and $\hat{Z}_t$ represents the sequence of m-ensemble hindcasts. The CRPS is a probabilistic metric that compares the cumulative probability distribution (CDF) of the forecast to the CDF of the target. In our case, the target, $\mathbf{z}(t)$, is a single observation, thus the CDF is a step function. The CRPS has an analytical solution for parametric distributions, like the Gaussian distribution, as well as for the empirical distribution (Gneiting and Raftery 2007). For our m-ensemble member forecast, $\hat{Z}(t)$, we use the CRPS for emprical distributions, which is defined as,

$$\mathrm{CRPS}(t) = E\left[|\hat{Z}(t) - \mathbf{z}(t)|\right] - \frac{1}{2} \cdot E\left[\left|\hat{Z}(t) - \hat{Z}'(t)\right|\right]. \tag{12}$$

The LSTM component is configured to process the forecast from each CS-LIM ensemble member, represented as $[\hat{\mathbf{z}}(t+1), ..., \hat{\mathbf{z}}(t+T)]$. This input is transformed into a higher-dimensional latent space through a linear layer. Following this, the transformed input is sequentially processed by two LSTM layers. Each timestep's input is combined with the LSTM's hidden state from the previous timestep, allowing for the progressive accumulation of information across the entire forecast sequence. The final step involves projecting the hidden state back into the PC space, utilizing another linear layer. To forecast the fields in grid space, we multiply the output of the LSTM, $\hat{\mathbf{z}}(t+\tau)$, with the respective EOFs, described in sec.a.

d. Deep learning baselines

In this work, we provide a comparison of the Hybrid-model against fully neural network-based approaches. Similar to the structure of our hybrid model, we construct an LSTM that operates in the PC space. Additionally, we explore the application of a Convolutional LSTM (ConvLSTM) architecture, specifically tailored to perform forecasts on spatial fields.

1) PC-LSTM

We implement an Encoder-Decoder LSTM architecture in PC space as a fully neural-network-based model (Sutskever et al. 2014). Unlike the LIM and Hybrid-model, this model incorporates

information from timepoints preceding the initialization time, t, which we term as 'history'. The Encoder network is designed to aggregate this historical information into a latent space. It begins with a linear layer that transforms the history, represented as $[z(t-n),...,z(t)]$, into a higher-dimensional latent space. This transformed input is then processed by two LSTM layers, where the input is added to the hidden state from the previous timestep. The hidden state at time t is passed to the Decoder network. The Decoder, mirroring the structure of the Encoder, consists of two LSTM layers, which are rolled out over the prediction horizon $t+T$, transferring the hidden state sequentially without additional input integration. This is followed by a linear layer that transforms the hidden state back to the PC space. To generate m-ensemble members, we employ separate linear layers for each member. Similar to the Hybrid model, we incorporate seasonal information through an affine transformation in both the Encoder and Decoder networks. The entire network is trained end-to-end, using the CRPS loss function as described in Eq. 11.

2) SwinLSTM

We employ a second DL model, that operates directly on grid space without relying on PC truncation of data. This model, which we refer to as the sliding window (Swin) LSTM, incorporates a ConvLSTM-like architecture with an Encoder-Decoder framework which is similar to the aforementioned LSTM. The encoder in this architecture accumulates historical information in the hidden state, while the decoder iteratively generates the forecast sequence. Unlike the LSTM, the input, latent, and hidden states in our model maintain dimensions of channel, height, and width, albeit of varying sizes. For model input, variables are stacked along the channel dimension. The encoder then first downscale the height and width dimensions, by simultaneously expanding the channel dimension via a convolutional layer. Following this, the transformed input is processed through two layers of a ConvLSTM-like structure, where each layer adds the input to the previous timestep's hidden state. Our approach includes an adaptation of the ConvLSTM layer, where we segregate the spatial and channel mixing, interspersing them with layer normalization, as depicted in Fig. A2. This modification facilitates the conditioning on monthly embeddings, implemented through the affine transformation outlined in Eq. 10. Finally, the decoder network, which also consists of two adapted ConvLSTM layers and a transpose convolution, transforms the aggregated hidden state back into the grid space.

e. Evaluation metrics

Our analysis of the models, all of which generate ensemble member predictions, is based on probabilistic metrics as well as deterministic metrics of their ensemble mean. We evaluate all models on the test set (200 years) using SSTA and SSHA in the tropical Pacific.

(i) Root mean square error skill score (RMSESS) The RMSESS is a deterministic metric that compares the RMSE of the model to the RMSE of a reference forecast. Throughout this work, we choose the climatology as our reference forecast. The RMSESS can be written as

$$\mathrm{RMSESS}(\tau) = 1 - \frac{\mathrm{RMSE}_{\mathrm{model}}(\tau)}{\mathrm{RMSE}_{\mathrm{ref}}(\tau)} = 1 - \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{\mathbf{x}}(t+\tau) - \mathbf{x}(t+\tau)\right)^2}}{\sigma_{\mathbf{x}}} \tag{13}$$

where $\sigma_{\mathbf{x}}$ is the variance of the data. An RMSESS of 1 is a perfect model forecast and 0 is as good as the climatology forecast.

(ii) Pattern correlation The pattern correlation, ρ, is the spatial correlation coefficient between the model forecast and the target at each time point. The pattern correlation is defined as,

$$\rho(t+\tau) = \frac{\mathrm{cov}(\hat{\mathbf{x}}(t+\tau), \mathbf{x}(t+\tau))}{\sigma_{\hat{\mathbf{x}}(t+\tau)}\sigma_{\mathbf{x}(t+\tau)}} \tag{14}$$

where the covariance matrix is defined as $\mathrm{cov}(\hat{\mathbf{x}}, \mathbf{x}) = \langle \hat{\mathbf{x}}, \mathbf{x} \rangle$, with $\langle \cdot, \cdot \rangle$ denotes the scalar product.

Sardeshmukh et al. (2000) derived the theoretically expected pattern correlation for an infinite-member ensemble-mean forecast as

$$\rho_{\infty} = \frac{S^2(t,\tau)}{\left[\left(S^2(t,\tau)+1\right)S^2(t,\tau)\right]^{1/2}} \tag{15}$$

with $S^2(t,\tau)$ being the signal-to-noise ratio of the perfect model forecast. For the LIM, the signal-to-noise ratio is defined as $S^2(t,\tau) = \mathrm{tr}\left(F(t,\tau)\right)/\mathrm{tr}\left(E(\tau)\right)$ with the signal covariance, $F(t,\tau) = \langle \hat{\mathbf{x}}(t+\tau), \hat{\mathbf{x}}(t+\tau) \rangle$, and the time-independent noise covariance $E(t,\tau) = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^T(\tau)$.

(iii) Continuous ranked probability skill score (CRPSS) Equivalently to the skill score of the RMSE, we define the CRPS skill score using Eq. 12 as, $\mathrm{CRPSS}(\tau) = 1 -$

$\mathrm{CRPS}_{\mathrm{model}}(\tau)/\mathrm{CRPS}_{\mathrm{ref}}(\tau)$, where the reference forecast is the climatology. A CRPSS of 1 is a perfect model forecast and 0 is as good as the climatology forecast.

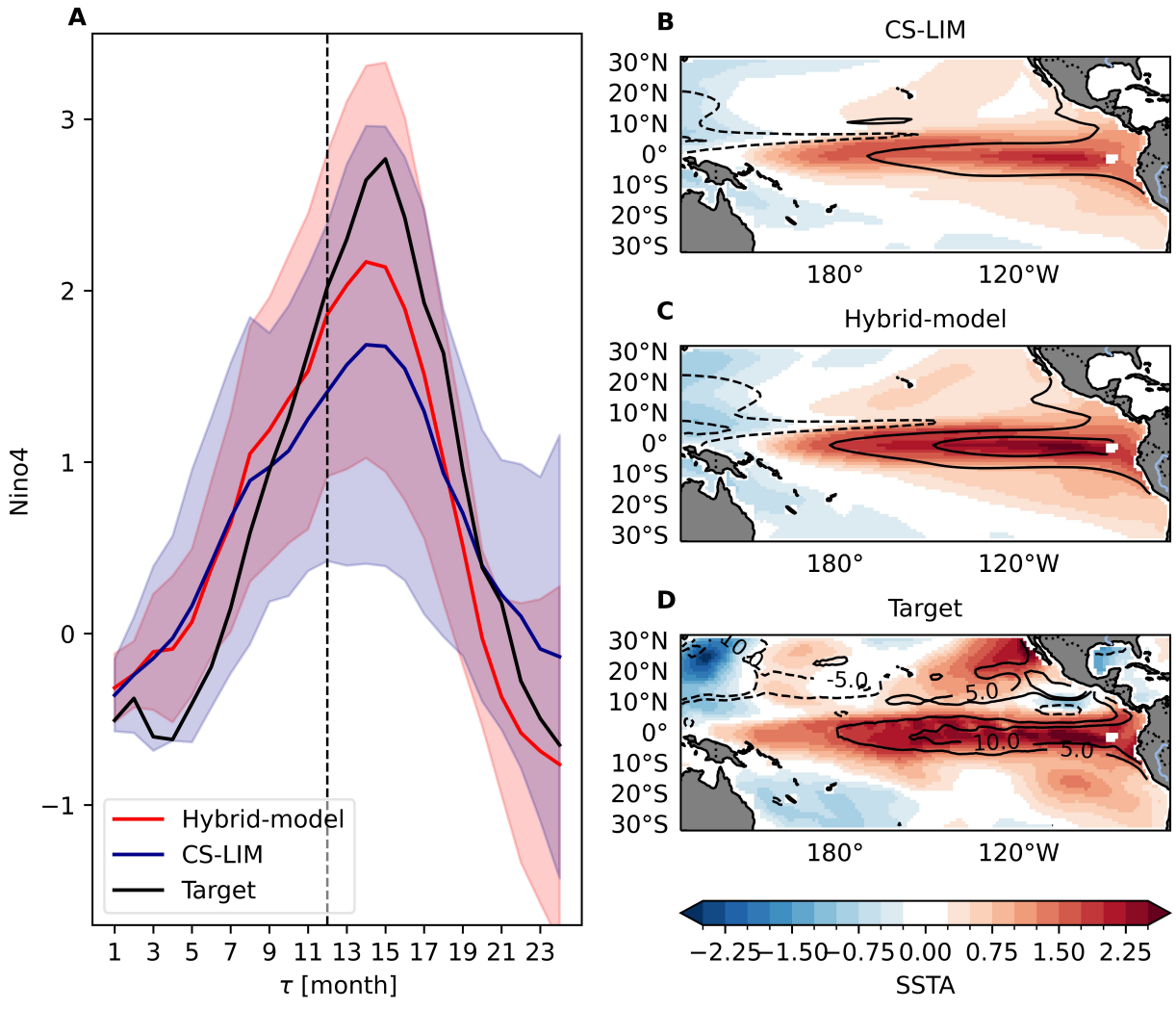
4. Results

a. Improved skill due to nonlinearities

When carefully designed, the LIM captures all predictable linear dynamics of the system. We fit the LIM using the PCs of SSTA and SSHA from CESM2 preindustrial control run in the tropical Pacific, as described in section 3b. Our Hybrid-model is designed to learn the residuals between the LIM's predictions and the target data, as detailed in section 3c. The improvement of the Hybrid-model upon the LIM can thus be attributed to the nonlinearities in the tropical Pacific ocean dynamics.

We present an example 24-month hindcast of the LIM and Hybrid-model (Fig. 3). The forecasts are initialized in December, 12 months before the El Nino event, indicated as a dashed line in Fig. 3 A. El Niño events are identified when the Nino3.4 index exceeds its standard deviation for at least three consecutive months. The evolution of the average SSTA in the Nino4 region of the target is presented as a black line in Fig. 3 A. Throughout this work, we analyze the Nino4 instead of the Nino3.4 region because of the west Pacific bias in CESM2. Both the CS-LIM (red line) and Hybrid-model (blue line) forecasts predict the observed warming, as depicted by the ensemble members' mean, with the shading indicating their respective standard deviations. The SSTA and SSHA field forecasts at $\tau = 12$ months lag time (Fig. 3C,D), exhibit the El Nino warming in the equatorial Pacific but lacks the refined spatial structure evident in the target fields (Fig. 3 A). Notably, the Hybrid-model forecast magnitude is closer to the CESM2 target data than the CS-LIM forecast. The Hybrid-models uncertainty range, given by the standard deviation between 16 ensemble member forecasts, is also narrower than the CS-LIM forecast, yet still encompasses the target data, suggesting a better model calibration.

To move beyond an example, we evaluate the RMSE and CRPS skill scores of the CS-LIM and Hybrid-model on 200 years of test data to quantify its skill improvement. We have to ensure that the LIM captures all predictable linear dynamics. To this end, various LIM variants, including known factors influencing the tropical Pacific, are analyzed.



293 FIG. 3. **Example El Nino hindcast:** Example of an hindcaset initialized 12 months prior to an EN event. The
294 mean (solid line) and standard deviation (shading) over the 16 ensemble members of the CS-LIM and Hybrid
295 model forecast are shown in A. The mean SSTA forecast at $\tau = 12$ months for the CS-LIM (B) and Hybrid-model
296 (C) show the same warming patterns than the target (C).

321 The initial LIM variant, formulated by Penland and Sardeshmukh (1995), is solely fitted to the
322 first 30 PCs of SSTA in the Pacific and is termed stationary (ST)-LIM (ssta). An advancement to
323 this is the CS-LIM (ssta), introduced by Shin et al. (2021), which integrates seasonal variation and
324 substantially surpasses the skill of ST-LIM (ssta), as shown by the RMSE (A) and CRPS (B) skill
325 scores of the Nino4 index in Fig. 4. For reference, we present the skill of the persistent forecast

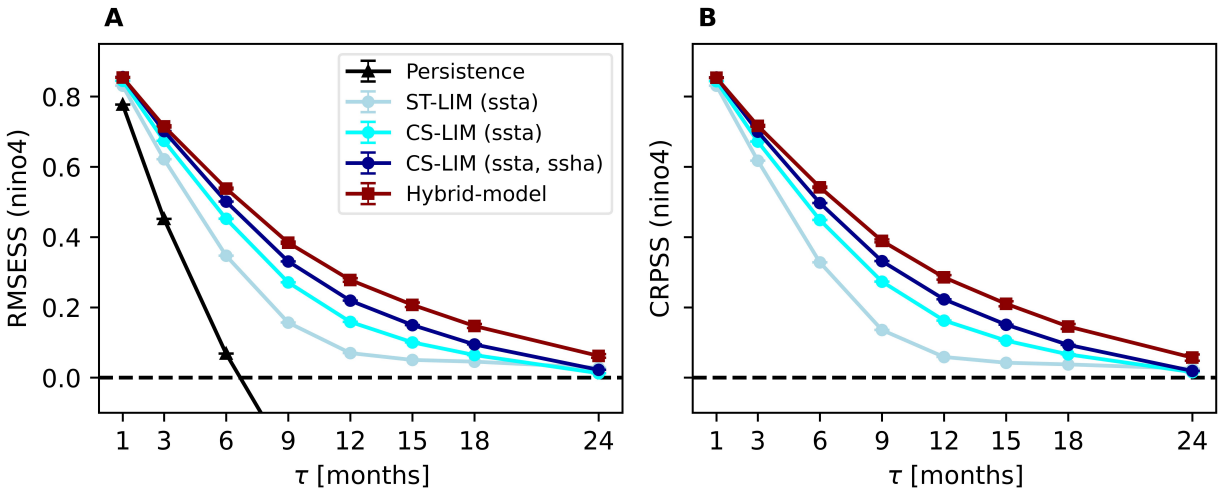


FIG. 4. **RMSE and CRPS skill scores of the LIM versions and our Hybrid-model.** Skill scores for RMSE (A) and CRPS (B) across our LIM versions and the Hybrid model are evaluated over forecast lag time (τ) using the average SSTA in the Nino4 index region on the test set. The progression in LIM versions from the stationary (ST)-LIM, which uses SSTA data, to the more advanced cyclostationary (CS)-LIM incorporating seasonality, and then to the CS-LIM (ssta, ssha) that includes both seasonality and SSH factors is depicted. Enhancing the CS-LIM (ssta, ssha), our Hybrid model utilizes an LSTM to effectively learn and adjust for its residuals.

which is worse than a climatological forecast (dashed line at zero) after $\tau = 6$ months forecast lag time.

In line with Chen et al. (2016)'s insight on the role of ocean variables in forecasting, a third variant, the CS-LIM (ssta, ssha), incorporating the first 10 PCs of SSHA in the tropical Pacific is employed. We select SSH for its model and observational accessibility and its strong correlation with ENSO prediction factors, namely the upper ocean heat budget and thermocline depth. This version exhibits a further skill enhancement over the CS-LIM (ssta). Progressive improvements are evident in each LIM version, with the inclusion of seasonality and ocean variables contributing significantly to enhanced skill(Fig. 4).

For our Hybrid-model, we use the LIM with the highest skill, the CS-LIM (ssta, ssha), and learn its residuals to the data using an LSTM. The LSTM takes 16 ensemble member forecasts of the CS-LIM as input and learns their residuals by minimizing the CRPS loss function detailed in Equation 11. To ensure the robustness of our findings, we employ five separate training sessions with varied weight initialization and data shuffling, which variability is depicted through error bars

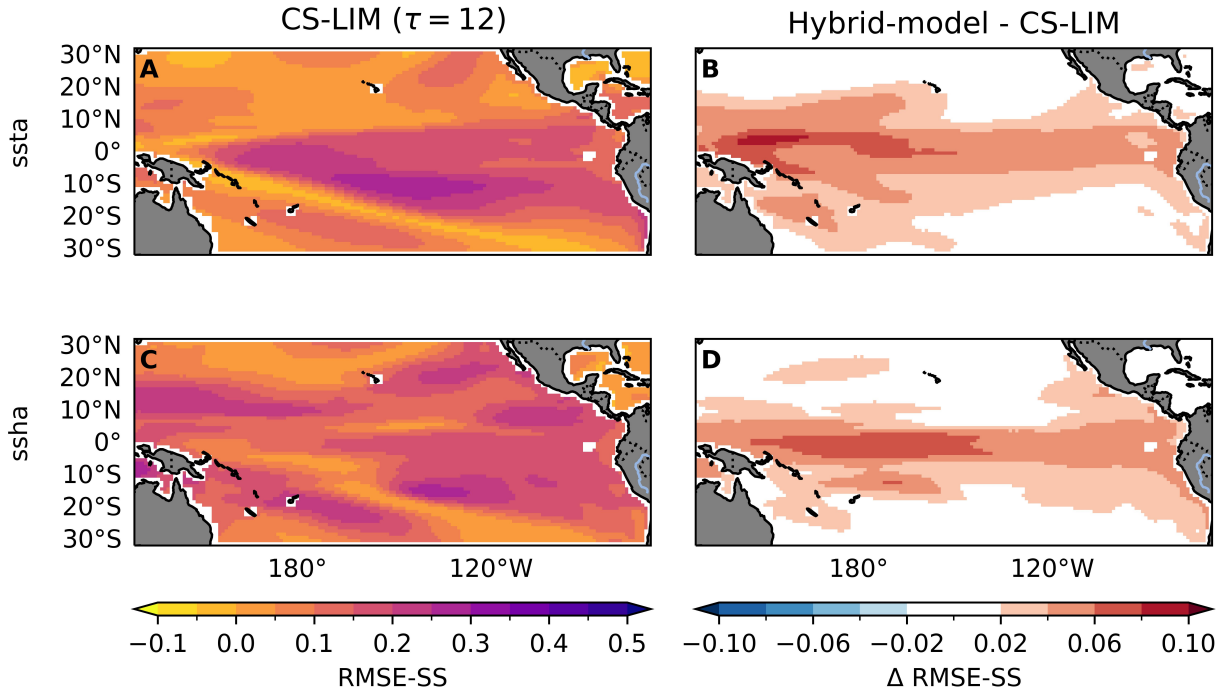


FIG. 5. **Spatial distribution of skill improvement.** RMSE skill score of SSTA and SSHA for the $\tau = 12$ month hindcast of CS-LIM (A, C) and the differences in RMSE skill score to the Hybrid model (B, D). Red colors indicate an improvement in skill while blue colors indicate a decrease in skill. Using a two-sided t-test, we evaluate the significance of the difference between the 1000 randomly bootstrapped means of CS-LIM and the Hybrid-model. Only those values that surpass the 95% confidence interval threshold are shown.

in Fig. 4. The skill improvement of the Hybrid-model upon the CS-LIM is significant at lag times larger than 6 months, which we can attribute to predictable nonlinearities.

We conducted further examination of the spatial distribution of skill for both the LIM and the Hybrid-model. The ensemble mean RMSESS for a 12-month CS-LIM forecast of SSTA exhibits higher skill around the equator compared to the Extratropics (Fig. 5A). In contrast, the RMSESS of SSHA demonstrated high skill in both the Eastern and North Western Pacific (Fig. 5C). Notably, the pronounced skill in SSTA in the far Western Pacific could be attributed to the west Pacific bias inherent in the CESM2.

When comparing the Hybrid-model with the CS-LIM, as depicted in Figure 5, panels B and D, there is a discernible skill improvement around the equator, with the most significant enhancement

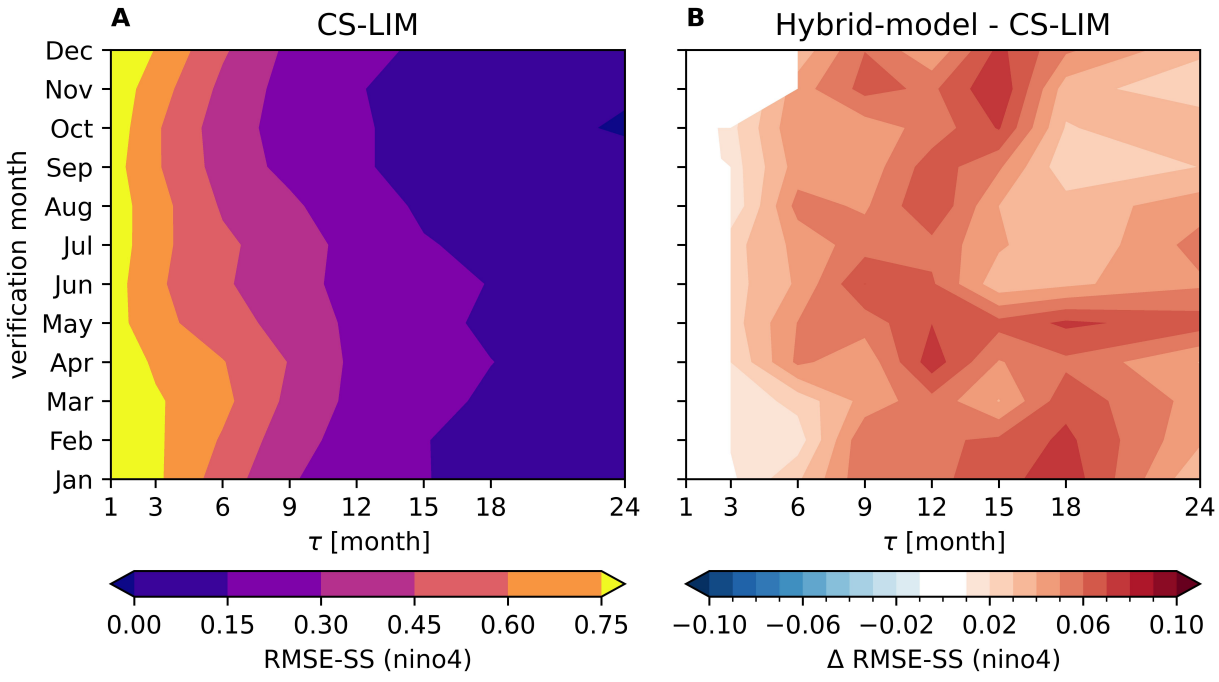


FIG. 6. **Seasonal skill dependency of Nino4.** The average RMSE skill score of the CS-LIM forecast for Nino4 SSTA is analyzed over various lag times and verification months (A). The Hybrid-model forecast significantly improves relative to the CS-LIM, particularly for lag times between 9 and 18 months (B). In both panels, statistical significance is determined using a two-sided t-test on 1000 bootstrapped means of CS-LIM and Hybrid-model forecasts. Results displayed exceed the 95% confidence interval.

observed in the western tropical Pacific. This improvement is consistent with the patterns observed when employing the CRPSS.

In addition to assessing the spatial distribution of skill, we investigate the seasonal skill variation. The average RMSESS of the Nino4 SSTA from the CS-LIM forecast, evaluated over different lag times and verification months, indicates that late winter and spring months are better predicted than the late summer and fall (Fig. 6A). These results are consistent with the findings by Shin et al. (2021), and align with the phenomenon known as the spring predictability barrier, characterized by a notable drop in the autocorrelation of the tropical Pacific SSTA in boreal spring.

For the Hybrid-model forecast, we observe the most significant RMSESS improvements upon the CS-LIM at lag times ranging between 9 and 18 months (Fig. 6B). Specifically, the maximum

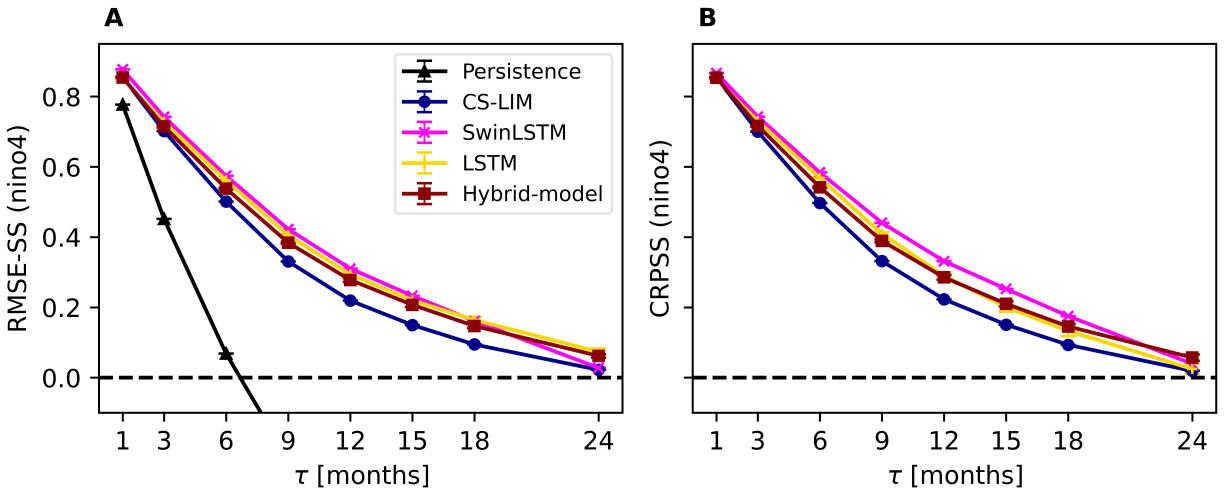


FIG. 7. **Skill of fully deep learning models.** RMSE and CRPS skill score of the Nino4-index for the best LIM (CS-LIM), our Hybrid-model and both deep learning baselines. Both deep learning models (LSTM, SwinLSTM) have comparable skill to the Hybrid-model suggesting that most predictable nonlinearities are captured. While the LSTM model is trained to forecast the PCs, the SwinLSTM forecasts are directly on the SSTA and SSHA fields. Error bars indicate the standard deviation between model training runs with varied weight initialization and data shuffling.

enhancements are seen at 15 and 18 months during the winter months (December to February), while in spring, the peak skill improvement is at a lag time of 12 months.

b. Comparison to deep learning baselines

We observed that the Hybrid-model significantly enhances forecast accuracy over the best CS-LIM which underscores the important role of nonlinearities in tropical Pacific Ocean dynamics. To further validate our findings, we conducted a comparative analysis between the Hybrid-model and two fully nonlinear deep learning models: an Encoder-Decoder LSTM operating in PC-space, simply referred to as LSTM, and a modified ConvLSTM, which we term the sliding window (Swin)LSTM, trained on the SSTA and SSHA in grid space. The latter is employed to evaluate the effects of PC truncation on forecasting skills.

Both LSTM and SwinLSTM, utilizing 16 ensemble members, are probabilistically trained using a CRPS loss function detailed in Equation 11. To ensure the robustness of our findings, each model underwent five separate training sessions with varied weight initialization and data shuffling, the

variability of which is depicted through error bars in Fig. 7. Further details on the architecture and training process are described in section d.

The RMSE and CPRS skill score of these models, alongside the CS-LIM and the Hybrid-model, is presented in Fig. 7. The Hybrid-model has similar skills to the deep learning models, though the SwinLSTM exhibits a slight improvement at the 12-month forecast horizon. This suggests that the Hybrid-model successfully captures most of the predictable dynamics, with the marginal gains of the SwinLSTM likely attributable to the PC truncation. Crucially, the Hybrid-model achieves this level of forecasting skill with significantly fewer parameters compared to the full deep learning models. This aspect is particularly beneficial in scenarios with limited data, as shown in Figure 1. When working with less than 500 years of monthly data, both the LIM and Hybrid-model outperform the LSTM, indicating that the pure deep learning models require a minimum of 600 data points to match the performance of the LIM and Hybrid-mode

c. Predictability assessment in the Hybrid-model

Within the linear framework of the LIM, we are able to calculate the optimal initial condition for a forecast lag time τ. This initial condition, also referred to as optimal precursor, is the singular vector that aligns with the largest singular value of the forecast operator $G(\tau) = \exp(\mathbf{L}\tau)$. Figure 8A depicts the optimal initial condition of the CS-LIM for a 12-month forecast starting in April. The subsequent evolution of this optimal initial pattern after $\tau = 12$ months is shown in Figure 8B. It is important to note that the magnitude of these patterns is arbitrary, a result of the unit norm of the singular vector.

For any given initial forecast state, the CS-LIM's predictability can be estimated by projecting the state onto its optimal initial condition. The strength of this projection, or how well the first singular vector aligns with the initial state – termed as optimal initial growth – is a key determinant of the potential forecast skill of the CS-LIM. To illustrate that, we select initial states from the test set that have either the lowest (0-10%) or highest (90-100%) optimal initial growth. The average pattern correlation skill ρ, illustrated in Figure 8C, is substantially larger for hindcasts initialized from states with highest optimal initial growth than for states with lowest optimal initial growth.

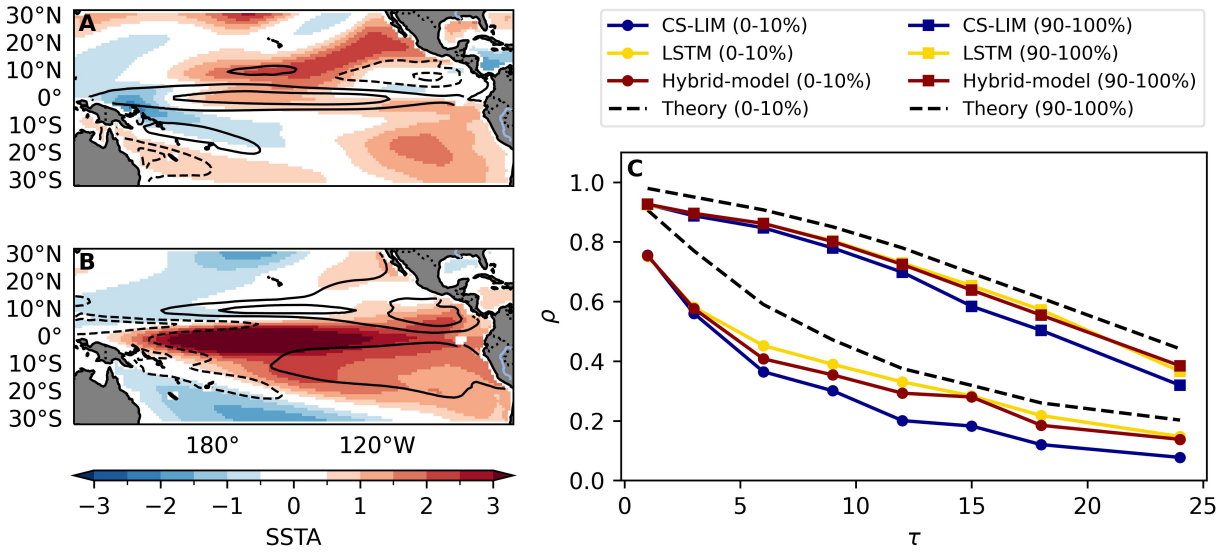


FIG. 8. **Predictability is determined by linear optimals.** The optimal initial condition (A) of the CS-LIM for a 12-month lag forecast initialized in April evolves into an El Nino-like pattern after 12 months (B). The projection of the data on the initial condition does not only indicates the potential skill of the CS-LIM but also of Hybrid-model and LSTM (C). We compute the pattern correlation over the CS-LIM, Hybrid-model and LSTM forecasts where the absolute projection of the data on the optimal initial pattern is the lowest (0-10%) and the highest (90-100%). The dashed line in (C) shows the theoretically expected pattern correlation of the CS-LIM, which is obtained using Eq. 15.

Besides the empirical skill, we also include the theoretically expected pattern correlation of the CS-LIM in Fig. 8. The theoretical expected skill is obtained from the analytic error covariance matrix and acts as an upper bound for the infinite ensemble member CS-LIM forecast.

The pattern correlation of the Hybrid-model forecast initialized from states with lowest and highest optimal initial growth exhibit both a clear difference in skill and also a substantial skill increase compared to the CS-LIM (Fig. 8 C). This results imply that optimal initial growth, a property derived from the CS-LIM, influences not only the linear predictability but also significantly impacts the predictability of nonlinear dynamics. This hypothesis is further supported by our analysis of the LSTM forecast, where a marked increase in skill is observed for initial states with high optimal initial growth as opposed to those with low.

d. ENSO asymmetry is nonlinearly predictable

The projection onto the optimal initial condition, as shown in 8 A, can yield either positive or negative values which evolve into warm and cold patterns. We examine the average $\tau = 12$ month hindcast of April initial states with the absolute highest optimal initial growth (>90%) for our CS-LIM, Hybrid-model and LSTM model.

Delineating the average hindcast pattern's magnitude and the asymmetry between warm and cold patterns, we present warm-cold patterns (Fig. 9, first column) and warm+cold patterns (Fig. 9, second column), respectively. A two-sided t-test is utilized to ascertain if there is a statistically significant difference between the means of the two distributions. We report only those values that surpass the 95% confidence interval threshold. The warm-cold pattern in the target data (Fig. 9A) closely resembles the evolved optimal pattern of the CS-LIM (Figure 8B) and its average empirical hindcast (Fig. 9C).

An evident east-west dipole structure is observed in the asymmetry of warm versus cold events of the target data, see Fig. 9B. Cold events exhibit higher SSTA and SSHA magnitudes in the western tropical Pacific, while warm events show greater magnitudes in the Eastern Pacific. This asymmetry is not present in the warm+cold patterns of CS-LIM hindcast (Fig. 9D), which is due to its linear and therefore symmetric evolution of cold and warm events. The remaining subtle differences between warm and cold patterns of CS-LIM hindcast likely originate from the asymmetrical distribution of initial conditions.

In contrast, the Hybrid-model and LSTM hindcasts accurately capture both the magnitude and asymmetry of warm and cold events. Both nonlinear model hindcasts exhibit the zonal dipole structure present in the target data, as shown in Fig. 9F and H. This finding highlights the ability of nonlinear models to predict the asymmetry between warm and cold patterns. They also underscore the Hybrid-model's potential in disentangling linear and nonlinear predictable dynamics, setting the stage for future systematic analysis of nonlinearities in subsequent work.

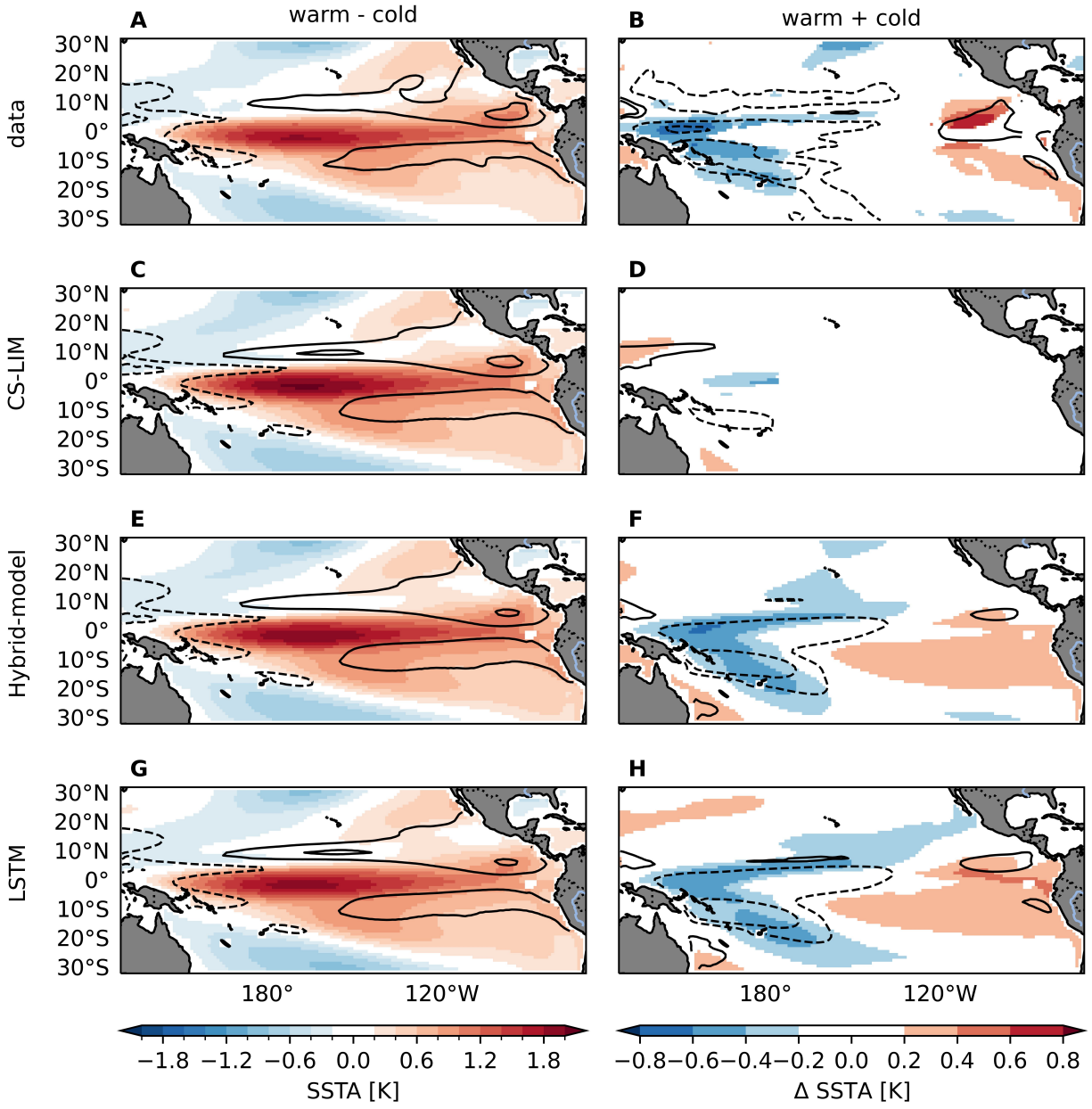


FIG. 9. **Nonlinear models capture ENSO asymmetrie.** The 12-month evolutions of states initialized in April with the absolute highest optimal initial growth (>90%) show warm and cold patterns. The average target state of warm-cold patterns (A) and warm+cold patterns (B) are depicted, as well as their hindcasts of the CS-LIM (C, D), Hybrid-model (E, F) and LSTM (G, H). While the warm-cold pattern delineates the average hindcast pattern's magnitude, the warm+cold pattern depicts their asymmetry. Using a two-sided t-test, we evaluate the significance of the difference between the means of the cold and the warm patterns. Only those values that surpass the 95% confidence interval threshold are shown.

5. Discussion

In this study, we introduce a Hybrid-model specifically tailored for forecasting SST and SSH in the tropical Pacific, critical factors in seasonal forecasting worldwide. We start from the LIM, an empirical model describing the dynamics of the slower-varying ocean as stochastically forced by the rapidly varying atmosphere with its deterministic dynamics assumed to be linear. However, while the LIM produces ENSO forecasts comparable to state-of-the-art numerical models, it is unable to capture observed asymmetries of ENSO that may also be important to its predictability.

We combine an LSTM with the LIM to capture predictable nonlinearities and non-Markovianity in the evolution of monthly tropical SSTA. This Hybrid-model is trained and tested on SSTA and SSHA data from the CESM2 preindustrial control run, where we observe that modeling nonlinearities significantly enhances the forecast accuracy, particularly in the western tropical Pacific within the 9 to 18-month range.

Our Hybrid model facilitates disentangling linear from nonlinear dynamics. Our findings provide initial evidence that the asymmetry between warm and cold events is a key source of nonlinearity that improves forecasting skill. This first insight lays the groundwork for a more comprehensive follow-up investigation of the potential sources of nonlinearity of ENSO forecasting.

Moreover, we demonstrate that the predictability of the Hybrid-model is strongly related to the theoretical expected skill of the LIM which allows us to reliably assess its predictability. In contrast, neural networks typically struggle to provide accurate predictability assessments on seasonal to annual scales, primarily hindered by their weak spread-to-skill relationship. While our Hybrid-model shows accurate predictability in the tropical Pacific, this potentially offers predictability for other climate oscillation and variables on sub-seasonal to seasonal scales, a promising avenue for future research.

A notable feature of our Hybrid-model is its relatively modest data requirements for training, particularly when compared to more data-intensive deep learning models like the LSTM network. This aspect is crucial given the limited span of available oceanic observational data. For a fair comparison, we utilized data from global circulation models in our training, acknowledging their inherent biases as discussed in our study. The use of domain adaptation methods from deep learning emerges as a promising strategy to close the gap between GCM data and observational

data. However, the field still needs more research to fully understand how neural network models can be adjusted to observational data when pre-trained on simulated data.

Acknowledgments. The authors would like to thank the reviewers for their helpful comments and suggestions. The authors thank the International Max Planck Research School for Intelligent Systems for supporting Jakob Schlör.

The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. Furthermore, we express our gratitude to the NOAA Physical Science Laboratory for making their resources available for this study.

Each author's contributions to the paper are listed below.

Conceptualization:	-
Methodology:	-
Investigation:	-
Visualization:	-
Supervision:	-
Writing—original draft:	-
Writing—review & editing:	-

Authors declare that they have no competing interests.

Data availability statement. All data used in this study are publicly available and are referenced in the main text or the supplementary materials. Our code is available at `https://github.com/jakob-schloer`.

APPENDIX A

Appendix

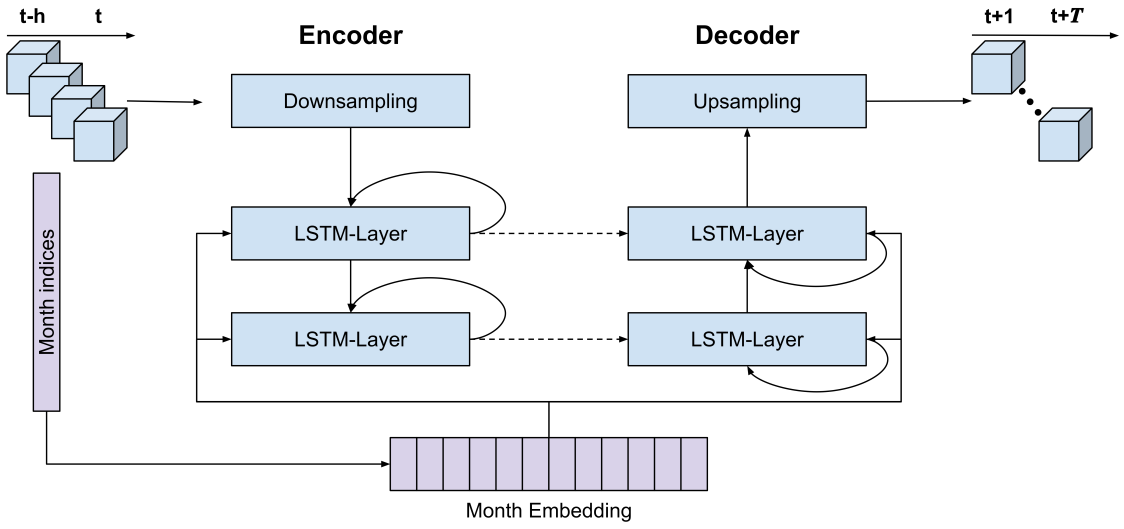
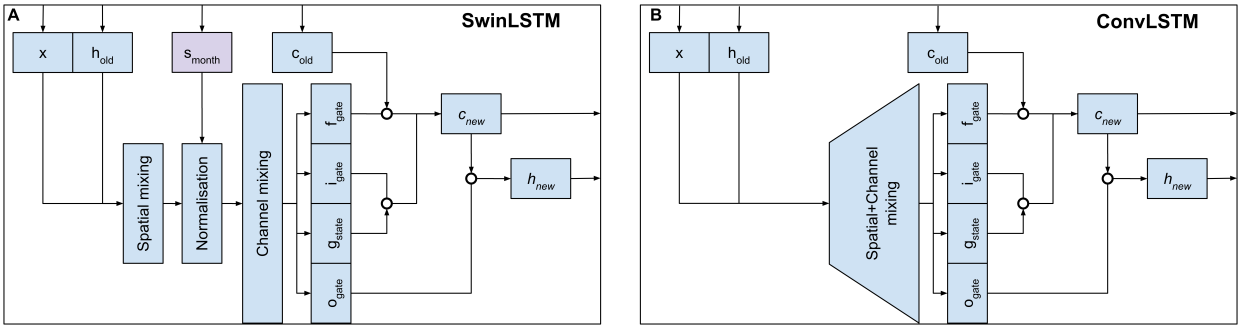


FIG. A1. **Encoder-Decoder architecture of the LSTM and SwinLSTM.** The encoder network starts with a downsampling layer, a linear layer for LSTM and a convolutional layer for SwinLSTM that condenses the historical input sequence into a latent representation. This is sequentially processed through two LSTM (ConvLSTM) layers, integrating the transformed input with the preceding hidden state. The resulting hidden state at time t is passed to the decoder network. Reflecting the encoder's design, the decoder is composed of two LSTM (ConvLSTM) layers, extending over the future prediction span t + T, transferring the hidden state across time without further input. A final upsampling phase, using individual upsampling layers for each m-ensemble members reverts the hidden state to the original input space. Monthly conditioning is incorporated into both encoder and decoder LSTM layers via affine transformations of the month embedding.



519 FIG. A2. **Schematic representation of the SwinLSTM and the ConvLSTM cell.** Input to our adapted
520 ConvLSTM cell, the SwinLSTM (A), and the ConvLSTM cell by Shi et al. (2015) (B) is the concatenated latent
521 state x and hidden state from the previous time step h_{old}. While in the ConvLSTM cell spatial and channel
522 mixing is performed at once, in the SwinLSTM, we separate spatial and channel mixing into two convolutions.
523 We further apply a layer normalization and conditioning on the monthly embedding (see Eq. 10) in between the
524 spatial and channel mixing. The remainder of the cells are equivalent to the ConvLSTM.

References

Albers, J. R., and M. Newman, 2021: Subseasonal predictability of the North Atlantic Oscillation. *Environmental Research Letters*, **16 (4)**, 044 024, https://doi.org/10.1088/1748-9326/abe781.

Cachay, S. R., E. Erickson, A. F. C. Bucker, E. Pokropek, W. Potosnak, S. Bire, S. Osei, and B. Lütjens, 2021: arXiv:2104.05089. The World as a Graph: Improving El Ni\~no Forecasts with Graph Neural Networks. arXiv, 2104.05089.

Callahan, C., and J. S. Mankin, 2023: Persistent effect of El Niño on global economic growth | Science. https://www.science.org/doi/10.1126/science.adf2983.

Capotondi, A., and Coauthors, 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., M. A. Cane, N. Henderson, D. E. Lee, D. Chapman, D. Kondrashov, and M. D. Chekroun, 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.

Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, **12 (2)**, e2019MS001 916, https://doi.org/10.1029/2019MS001916.

Geng, L., and F.-F. Jin, 2022: ENSO Diversity Simulated in a Revised Cane-Zebiak Model. *Frontiers in Earth Science*, **10**.

Gneiting, T., and A. E. Raftery, 2007: Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102 (477)**, 359–378, https://doi.org/10.1198/016214506000001437.

Goel, H., I. Melnyk, and A. Banerjee, 2017: arXiv:1709.03159. R2N2: Residual Recurrent Neural Networks for Multivariate Time Series Forecasting. arXiv, https://doi.org/10.48550/arXiv.1709.03159, 1709.03159.

Guo, C., G. Pleiss, Y. Sun, and K. Q. Weinberger, 2017: On Calibration of Modern Neural Networks. *Proceedings of the 34th International Conference on Machine Learning*, PMLR, 1321–1330.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573 (7775)**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, **95 (4)**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Martinez-Villalobos, C., D. J. Vimont, C. Penland, M. Newman, and J. D. Neelin, 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.

Newman, M., and P. D. Sardeshmukh, 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.

Okumura, Y. M., 2019: ENSO Diversity from an Atmospheric Perspective. *Current Climate Change Reports*, **5 (3)**, 245–257, https://doi.org/10.1007/s40641-019-00138-7.

Penland, C., and P. D. Sardeshmukh, 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.

Petersik, P. J., and H. A. Dijkstra, 2020: Probabilistic Forecasting of El Niño Using Neural Network Models. *Geophysical Research Letters*, **47 (6)**, https://doi.org/10.1029/2019GL086423.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño. *Monthly Weather Review*, **110 (5)**, 354–384, https://doi.org/10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2.

Rasp, S., and S. Lerch, 2018: Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review*, **146 (11)**, 3885–3900, https://doi.org/10.1175/MWR-D-18-0187.1.

Rodrigues, E., B. Zadrozny, C. Watson, and D. Gold, 2021: arXiv:2106.11111. Decadal Forecasts with ResDMD: A Residual DMD Neural Network. arXiv, 2106.11111.

Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of Probability Associated with El Niño. *Journal of Climate*, **13 (24)**, 4268–4286, https://doi.org/10.1175/1520-0442(2000)013<4268:COPAWE>2.0.CO;2.

Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, 2015: arXiv:1506.04214. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. arXiv, https://doi.org/10.48550/arXiv.1506.04214, 1506.04214.

Shin, S.-I., P. D. Sardeshmukh, M. Newman, C. Penland, and M. A. Alexander, 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.

Strnad, F. M., J. Schlör, C. Fröhlich, and B. Goswami, 2022: Teleconnection Patterns of Different El Niño Types Revealed by Climate Network Curvature. *Geophysical Research Letters*, **49 (17)**, e2022GL098 571, https://doi.org/10.1029/2022GL098571.

Sutskever, I., O. Vinyals, and Q. V. Le, 2014: arXiv:1409.3215. Sequence to Sequence Learning with Neural Networks. arXiv, https://doi.org/10.48550/arXiv.1409.3215, 1409.3215.

Takahashi, K., and B. Dewitte, 2016: Strong and moderate nonlinear El Niño regimes. *Climate Dynamics*, **46 (5-6)**, 1627–1645, https://doi.org/10.1007/s00382-015-2665-3.

Takahashi, K., A. Montecinos, K. Goubanova, and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters*, **38 (10)**, L10 704, https://doi.org/10.1029/2011GL047364.

Timmermann, A., and Coauthors, 2018: El Niño–Southern Oscillation complexity. *Nature*, **559 (7715)**, 535–545, https://doi.org/10.1038/s41586-018-0252-6.

van Straaten, C., K. Whan, D. Coumou, B. van den Hurk, and M. Schmeits, 2023: Correcting Subseasonal Forecast Errors with an Explainable ANN to Understand Misrepresented Sources of Predictability of European Summer Temperatures. *Artificial Intelligence for the Earth Systems*, **2 (3)**, https://doi.org/10.1175/AIES-D-22-0047.1.

Wang, S., L. Mu, and D. Liu, 2021: A hybrid approach for El Niño prediction based on Empirical Mode Decomposition and convolutional LSTM Encoder-Decoder. *Computers & Geosciences*, **149**, 104 695, https://doi.org/10.1016/j.cageo.2021.104695.

Zhou, L., and R.-H. Zhang, 2022: A Hybrid Neural Network Model for ENSO Prediction in Combination with Principal Oscillation Pattern Analyses. *Advances in Atmospheric Sciences*, **39 (6)**, 889–902, https://doi.org/10.1007/s00376-021-1368-4.

Zhou, L., and R.-H. Zhang, 2023: A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Science Advances*, **9 (10)**, eadf2827, https://doi.org/10.1126/sciadv.adf2827.