

Enhancing the predictability limits of ENSO with physics-guided deep echo state networks

Received: 8 November 2025

Accepted: 16 February 2026

Cite this article as: Zhang, Z., Meng, J., Qiu, Z. *et al.* Enhancing the predictability limits of ENSO with physics-guided deep echo state networks. *npj Clim Atmos Sci* (2026). <https://doi.org/10.1038/s41612-026-01360-5>

Zejing Zhang, Jun Meng, Zhongpu Qiu, Wansuo Duan, Jian Gao, Zixiang Yan, Jinghua Xiao, Xiaosong Chen, Wenju Cai, Jürgen Kurths, Shlomo Havlin & Jingfang Fan

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Enhancing the predictability limits of ENSO with physics-guided Deep Echo State Networks

Zejing Zhang¹, Jun Meng^{2,*}, Zhongpu Qiu³, Wansuo Duan², Jian Gao^{1,4}, Zixiang Yan¹, Jinghua Xiao^{1,*}, Xiaosong Chen³, Wenju Cai^{5,6,7,8}, Jürgen Kurths^{9,10}, Shlomo Havlin¹¹, and Jingfang Fan^{3,9,*}

¹School of Physical Science and Technology, Beijing University of Posts and Telecommunications, Beijing, China

²State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

³School of Systems Science/Institute of Nonequilibrium Systems, Beijing Normal University, Beijing, China

⁴State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, China

⁵Frontier Science Center for Deep Ocean Multispheres and Earth System (FDOMES) and Physical Oceanography Laboratory, Ocean University of China, Qingdao, China

⁶Laoshan Laboratory, Qingdao, China

⁷State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen, China

⁸State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China

⁹Potsdam Institute for Climate Impact Research, Potsdam 14412, Germany

¹⁰Institute of Physics, Humboldt-University, Berlin, Germany

¹¹Department of Physics, Bar-Ilan University, Ramat-Gan, Israel

*Corresponding authors: jun.meng.phy@gmail.com; jhxiao@bupt.edu.cn; jingfang@bnu.edu.cn

ABSTRACT

The El Niño–Southern Oscillation (ENSO) is a dominant mode of interannual climate variability, yet the mechanisms limiting its long-lead predictability remain unclear. Here we develop a physics-guided Deep Echo State Network (DESN) that operates on physically interpretable climate modes selected from the extended recharge oscillator (XRO) framework. DESN achieves skillful Niño 3.4 predictions up to 16–20 months ahead with minimal computational cost. Mechanistic experiments show that extended predictability arises from nonlinear coupling between warm water volume and inter-basin climate modes. Error-growth analysis further indicates a finite ENSO predictability horizon of approximately 30 months. These results demonstrate that physics-guided reservoir computing provides an efficient and interpretable framework for diagnosing and predicting ENSO at long lead times.

Introduction

The El Niño–Southern Oscillation (ENSO) is the dominant mode of interannual climate variability, arising from nonlinear ocean–atmosphere interactions in the tropical Pacific and exerting far-reaching impacts on global climate extremes, ecosystems, agriculture, and socio-economic systems^{1–6}. Classical conceptual models distill the essential physics of ENSO, from the Delayed Oscillator, which emphasizes delayed negative feedbacks associated with equatorial wave reflections^{7,8}, to the Recharge Oscillator (RO), which frames ENSO as a coupled evolution of sea surface temperature and equatorial subsurface heat content governed by Bjerknes feedback, delayed oceanic adjustment, stochastic forcing, and nonlinear atmospheric processes^{9–13}. State-of-the-art coupled general circulation models reproduce many observed ENSO characteristics and achieve skillful forecasts up to approximately one year^{14,15}.

However, accumulating evidence suggests that ENSO dynamics are embedded within a broader, multiscale inter-basin framework involving the Indian Ocean, the extratropical Pacific, and Atlantic variability^{16–22}. Data-driven approaches based on network theory and complexity science further demonstrate that dynamical coupling within the tropical Pacific and with remote oceans provides robust precursors of ENSO evolution, enabling forecasts beyond the seasonal timescale^{23,24}. These insights are formalized in the Extended Recharge Oscillator (XRO), which elevates ENSO from a basin-confined oscillator to a multi-basin coupled dynamical system and extends predictive skill to roughly 18 months²⁵.

While XRO represents a major advance toward physically interpretable long-lead ENSO prediction, existing conceptual models are formulated in a reduced state space with prescribed feedbacks, emphasizing physical interpretability and diagnostic clarity. In parallel, deep learning (DL) models have demonstrated substantial potential for long-range ENSO prediction by flexibly extracting complex spatiotemporal patterns from high-dimensional climate data. Landmark studies²⁶ demonstrated that purely data-driven convolutional networks could rival physical models, and recent transformer-based architectures such as 3D-Geoformer²⁷ have extended Niño3.4 forecast skill beyond 18 months. Despite their predictive success, these architectures are computationally demanding and often difficult to interpret physically, which complicates mechanistic diagnosis, robustness assessment, and regime transfer.

These contrasting strengths raise a central challenge: can machine learning, when guided by physically motivated conceptual frameworks such as XRO, achieve strong long-lead predictive skill while also providing a physically grounded understanding of ENSO dynamics? Here, “physically grounded understanding” refers to identifying the interaction pathways, timescales, and nonlinear structures that are dynamically essential for predictability, rather than establishing strict physical causality.

To address this challenge, we develop a physics-guided and computationally efficient forecasting framework based on Deep Echo State Networks (DESNs), a class of reservoir computing models well suited for nonlinear dynamical systems^{28–31}. Unlike conventional deep learning architectures, DESN combines hierarchical reservoir dynamics with physically interpretable inputs, enabling both long-lead prediction and systematic dynamical diagnosis. The model incorporates seasonal priors and ten ENSO-related climate indices selected by the XRO framework²⁵, allowing nonlinear cross-basin interactions to be represented without prescribing explicit coupling structures.

The proposed DESN framework achieves Niño3.4 forecast anomaly correlation coefficients exceeding 0.5 at lead times of 16–20 months, comparable to state-of-the-art physical and deep learning models^{25–27}, while requiring only seconds of training on a standard CPU. Beyond predictive skill, DESN supports mechanistic diagnosis through sub-model and controlled experiments, revealing that nonlinear, state-dependent couplings—particularly those involving warm water volume and cross-basin interactions—play a central role in sustaining predictability beyond canonical recharge–discharge dynamics. Building on the error-growth framework underlying nonlinear local Lyapunov exponent (NLLE) analyses^{32,33}, we examine the saturation behavior of forecast errors using an absolute-error metric. The resulting error growth dynamics indicate an intrinsic ENSO predictability horizon of approximately 30 months.

Rather than prioritizing forecast accuracy alone, this work demonstrates that incorporating physically interpretable structure into lightweight recurrent models enables interpretable and theory-consistent ENSO prediction. By guiding a Deep Echo State Network with climate modes motivated by the XRO framework, we show how nonlinear coupling structures and the intrinsic predictability limits imposed by ENSO dynamics jointly shape forecast skill across lead times.

Results

To explore long-lead ENSO predictability, we develop a physics-guided DESN framework that operates on a compact set of physically interpretable climate predictors. Guided by the extended recharge oscillator (XRO) framework²⁵, the model incorporates Niño3.4 (sea surface temperature anomalies averaged over 170°–120°W, 5°S–5°N)³⁴, warm water volume (WWV), defined as 20°C isotherm depth anomalies averaged over 120°E–80°W, 5°S–5°N³⁵, and a suite of key inter-basin climate modes (Fig. 1a). These include the North and South Pacific Meridional Modes (NPMM and SPM) ^{36,37}, the Indian Ocean Basin (IOB) mode³⁸, the Indian Ocean Dipole (IOD)³⁹, the Southern Indian Ocean Dipole (SIOD)³⁹, as well as Tropical North Atlantic (TNA) variability⁴⁰, the Atlantic Niño (ATL3)⁴¹, and the South Atlantic Subtropical Dipole (SASD)⁴², all derived from the ORAS5 ocean reanalysis (see Methods and Table S1). Together, these ten physically motivated climate indices, along with sinusoidal seasonal cycles, constitute the model inputs. Stochastic variability is incorporated as part of the model’s dynamical evolution and addressed through ensemble forecasting to extract the predictable signal (Fig. 1b). During inference, the model generates rolling monthly forecasts up to 21 months ahead, initialized at each calendar month with 20-member ensembles. Forecast skill is evaluated using in-sample, out-of-sample, and 24-year block cross-validation schemes (see Methods), providing a robust basis for assessing long-lead ENSO predictability and its dynamical origins.

DESN achieves high skill and long-term ENSO forecasts

We trained our DESN models on the ORAS5 dataset over the period 1958–1999 and evaluated them on independent data from 2002–2023, following an out-of-sample forecasting protocol (Supplementary Algorithms, and Table S2). Forecast skill was assessed using the anomaly correlation coefficient (ACC) of the Niño 3.4 index across a range of lead times. As benchmarks, we compared DESN against (1) a minimal Echo State Network (ESN), a standard single-layer reservoir computing model trained only on the Niño 3.4 index, (2) a full ESN trained on ten climate modes (Fig. 1a) and seasonal cycles, (3) the XRO model²⁵, (4) operational ensemble forecasts from the International Research Institute for Climate and Society (IRI), and (5) the state-of-the-art 3D-Geoformer deep learning model²⁷. While all models performed similarly at short lead times, DESN showed

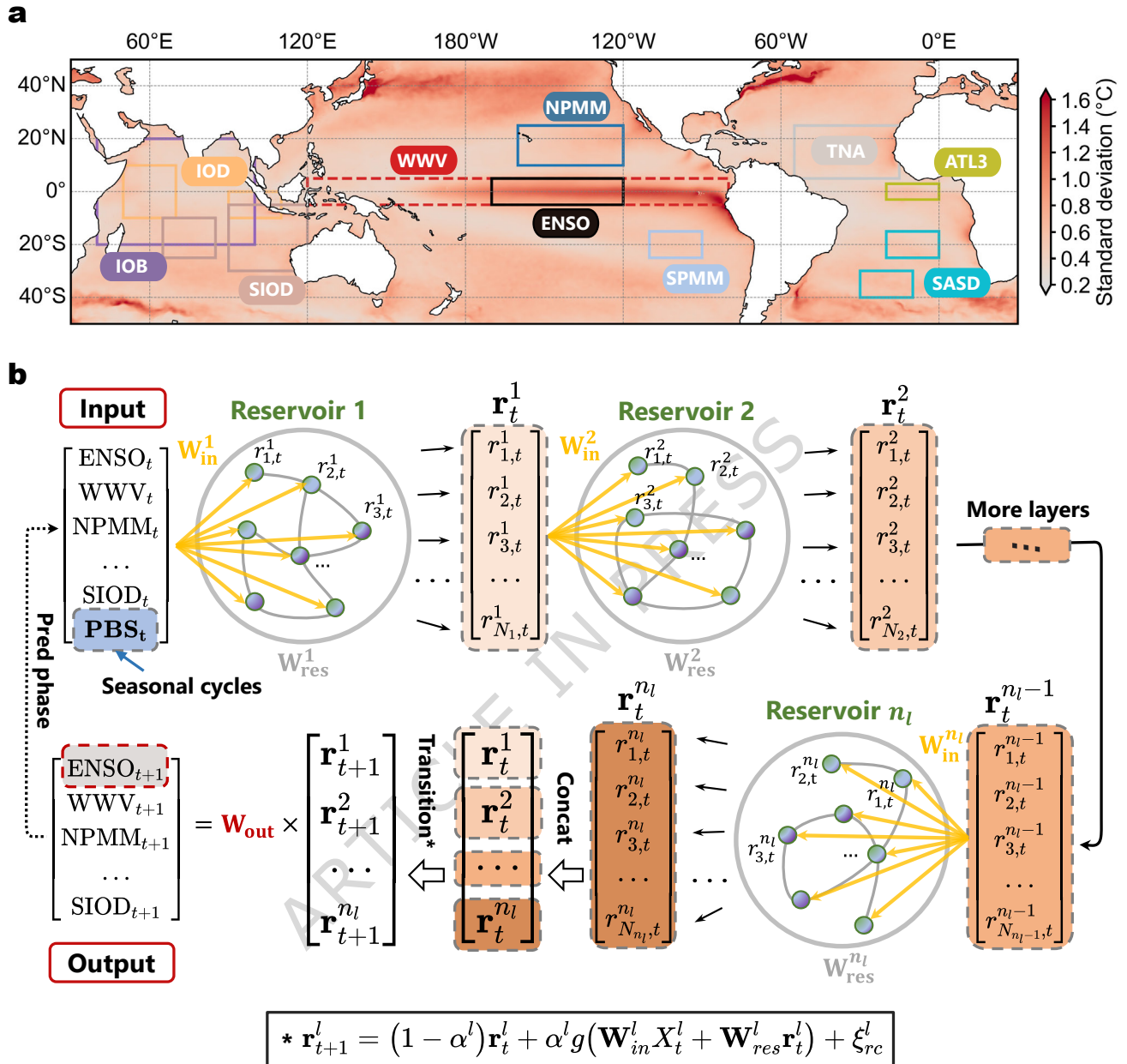


Figure 1. Regions of interest and architecture of the DESN model for ENSO prediction. **a**, Standard deviation of sea surface temperature anomalies (SSTA) from the ORAS5 reanalysis (1979–2023). Colored boxes denote regions used to compute area-averaged climate indices for ENSO and associated modes, following the XRO framework²⁵. **b**, Schematic of the DESN architecture. The input layer comprises multiple climate indices and seasonal cycles as periodic bootstrap sequences (\mathbf{PBS}_t). Inputs are passed through a hierarchy of n_l recurrent reservoirs, each generating a neuron state vector \mathbf{r}_t^l . Additive white noise ξ_{rc}^l is injected into each reservoir to emulate stochastic variability. Internal weights consist of input matrices \mathbf{W}_{in}^l and reservoir matrices \mathbf{W}_{res}^l . Reservoir outputs are concatenated and mapped to next-month targets via a linear readout matrix \mathbf{W}_{out} . Multi-step forecasts are obtained recursively using rolling inputs. When $n_l = 1$, the model reduces to a standard ESN. The reservoir transition function at the bottom illustrates neuron dynamics with activation function $g(\cdot)$ and leakage rate α . See Methods for definitions and implementation details.

clear advantages beyond 10 months, maintaining ACC values above 0.5 up to 16 months (Fig. 2a). It consistently outperformed the IRI dynamical ensemble beyond 9 months and surpassed the statistical ensemble after just 6 months.

To evaluate seasonal sensitivity, we computed target-month-dependent forecast skill using rolling initializations. The minimal ESN maintained ACC > 0.5 only at short leads across all months (Fig. 2c). The full ESN extended this horizon to about 15 months between January and June (Fig. 2d), but its performance dropped sharply around June and July, reflecting the Spring

Predictability Barrier (SPB)⁴³. In contrast, DESN preserved elevated skill through this season. For example, forecasts targeting June achieved a 10-month skillful lead with DESN compared to only 6 months for the ESN (Fig. 2e). These results indicate that extending the predictor set from the minimal to the full ESN highlights the importance of cross-basin coupling as a source of ENSO predictability, whereas the further improvement from the full ESN to DESN arises from the deep architecture's ability to represent higher-order, state-dependent couplings and multiscale interactions that provide additional predictive information.

We further assessed model generalization over extended periods, applying the same hyperparameters (Table. S2) as in the out-of-sample implementation to ensure consistent dynamical behavior. Evaluation was performed over two temporal periods: 1979 to 2023 for the in-sample experiment (Fig. 2b) and 1958 to 2023 for cross-validation (see Methods). In the in-sample test, the minimal ESN rapidly lost skill after 3 months, likely reflecting overfitting. The full ESN sustained ACC values > 0.5 up to 18 months, comparable to the XRO and 3D-Geoformer. By contrast, DESN outperformed all baselines, maintaining ACC > 0.5 across lead times up to 20 months, with particularly strong performance during major El Niño episodes (Figs. S1 and S2). Beyond ENSO, the DESN also delivered skillful long-lead predictions for nine other climate modes, achieving lead times of 9–20 months (Fig. S3). Robustness was further demonstrated through 24-year leave-out cross-validation across the full 1958–2023 window, where DESN retained ACC > 0.5 at lead times of 11–18 months (Fig. S4). In addition to forecast accuracy, DESN offers substantial computational efficiency. It trains in seconds on a standard CPU, with runtimes comparable to the XRO while achieving forecast skill that matches or exceeds both dynamical and deep learning approaches (Table. 1). These results show that DESN generalizes well across decades, avoids overfitting to particular ENSO regimes or climate states, and provides a computationally efficient pathway for advancing long-lead climate prediction.

Table 1. Comparison of training efficiency across ENSO forecasting models. Estimated training time and computational requirements for the DESN, ESN, XRO, and 3D-Geoformer models. ^a Out-of-sample evaluation (2002–2023); ^b Out-of-sample (1983–2021); ^c In-sample (1983–2023).

Models	Training data (size)	Training time (algorithms)	Training on	Months lead
3D-Geoformer ²⁷	Gridded monthly sea surface	~12 h	GPU	13 ^a
	wind stress and 3D ocean temperature anomalies (84 GB)	(Backpropagation)	(Nvidia V100)	18 ^b
XRO ²⁵	monthly climate indices (110 kB)	4 s	CPU	13 ^a
ESN	monthly climate indices (110 kB)	(Linear-regression)	(Intel i9-14900K)	18 ^c
		(Ridge-regression)	(Intel i9-14900K)	18 ^c
DESN (this study)	monthly climate indices (110 kB)	122 s	CPU	16^a
		(Ridge-regression)	(Intel i9-14900K)	20^c

Mechanisms shaping long-term ENSO predictability

Here we identify the dynamical mechanisms that sustain ENSO predictability at extended lead times. Using a suite of targeted and computationally efficient experiments enabled by the DESN framework, we show that long-lead ENSO predictability arises from nonlinear, seasonally modulated cross-basin interactions that are integrated into the subsurface ocean memory through WWV. By systematically varying predictor sets, input dimensionality, and model structure, and by directly comparing DESN with the physics-based XRO model, we disentangle the respective roles of seasonal forcing, cross-basin coupling, and nonlinearity. Together, these analyses reveal how physics-guided constraints and nonlinear interaction pathways jointly shape the limits and sources of long-term ENSO predictability.

First, we show that seasonal cycles can be interpreted as periodic bootstrap sequences. ENSO exhibits strong phase locking to the seasonal cycle, with events typically initiating in boreal spring and peaking in winter^{44,45}. To represent this periodicity, we introduced sinusoidal seasonal functions, or Periodic Bootstrap Sequences (PBS), as additional inputs, defined as,

$$\mathbf{PBS}_t = \begin{bmatrix} \sin \omega t \\ \cos \omega t \\ \sin 2\omega t \\ \cos 2\omega t \end{bmatrix}, \quad (1)$$

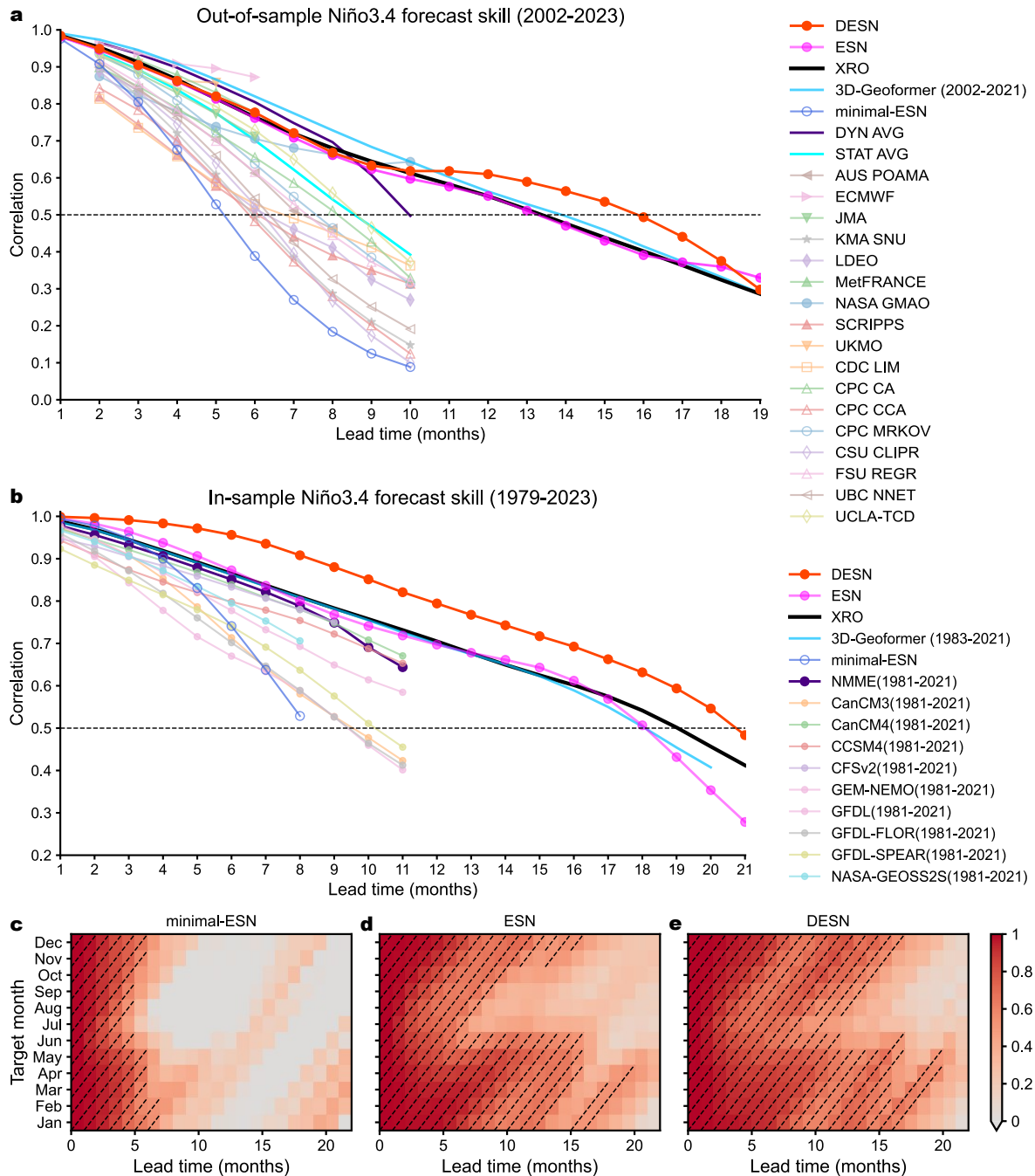


Figure 2. Forecast performance of ESN and DESN models for ENSO prediction. **a**, Forecast correlation skill (ACC) of the 3-month running mean Niño3.4 index as a function of lead time, evaluated out-of-sample for 2002–2023. Shown are DESN (red), ESN (magenta), minimal-ESN (dark blue), XRO (black; trained on 1958–1999), 3D-Geoformer (light blue), and the IRI operational ensemble forecasts—ensemble means of dynamical models (dark purple) and statistical models (dark cyan). **b**, Same as panel **a**, but evaluated in-sample over 1979–2023, including the ensemble mean of dynamical models from the North American Multi-Model Ensemble (NMME). Individual NMME model forecasts (1981–2021) are shown in various colors. Forecast periods for the 3D-Geoformer and NMME correspond to their respective training spans. **c–e**, Target-month-dependent ACC of Niño3.4 forecasts for the minimal-ESN (**c**), ESN (**d**), and DESN (**e**). Colors denote correlation skill as a function of target month (vertical axis) and lead time (horizontal axis). Black contours delineate regions where ACC exceeds 0.5, highlighting combinations of target months and lead times with high predictive skill.

where $\omega = \frac{2\pi}{T} = \frac{\pi}{6}$ is the angular frequency of the annual cycle ($T = 12$)⁴⁴, $t \in [0, L - 1]$, and L is the number of time steps in the climate indices. PBS terms are applied during both training and prediction phases, significantly altering the reservoir dynamics and enhancing prediction accuracy (Fig. S5). This demonstrates the value of embedding known seasonal structure directly into the model to capture periodic forcing more effectively.

We further interpret long-term ENSO predictability as arising from the nonlinear coupling between WWV and additional inter-basin climate modes, rather than by any single predictor acting in isolation. This interpretation is supported by targeted experiments (see Methods). To clarify the physical basis of these experiments, we recall that XRO restricts nonlinear terms to ENSO–WWV coupling and a conditional IOD quadratic contribution. In contrast, DESN does not impose explicit constraints on interaction structure, allowing higher-order and cross-basin nonlinear couplings to emerge implicitly from the data.

To diagnose the sources and characteristic timescales of predictability, we conduct controlled mode-decoupling experiments in which one climate mode is removed from the full predictor set while all remaining inputs are held fixed. Figs. 3a and b present the decoupling results for XRO and DESN, respectively. In both models, removing WWV leads to a pronounced collapse of long-lead forecast skill, confirming its indispensable role in sustaining extended ENSO predictability. By contrast, removing other individual modes while retaining WWV generally affects forecast skill but does not eliminate long-term predictability. Notably, the skill reduction associated with WWV removal is more pronounced in DESN than in XRO (Fig. S6), suggesting that DESN exploits WWV through higher-order nonlinear interactions with other climate modes that are not fully represented in the prescribed XRO framework.

To complement the decoupling analysis, we perform controlled mode-addition experiments that explicitly focus on interactions between the core ENSO variables and individual external climate modes. We define a baseline configuration consisting of Niño 3.4 and WWV, which encapsulates the classical recharge–discharge framework. One additional climate mode from the standard XRO set is then added at a time, and both XRO and DESN are trained under identical input conditions. Figs. 3c and d show the mode-addition results for XRO and DESN, demonstrating that retaining WWV alone is insufficient to sustain long-lead predictability and that extended forecast skill emerges only when WWV is coupled with additional climate modes. Among these, the NPMM yields consistent improvements in both XRO and DESN, highlighting its robust contribution to extended-range ENSO predictability.

Figs. 3e and f examine systems of progressively increasing dimensionality, defined by the number of physically distinct climate modes coupled to the core ENSO variables (Niño 3.4 and WWV). Predictive skill increases as additional climate indices are combined with WWV and Niño 3.4, confirming that cross-mode interactions substantially enhance ENSO predictability.

Taken together, these experiments provide complementary evidence for a unified mechanism. Retaining ENSO together with external climate modes while excluding WWV fails to recover long-lead forecast skill, indicating that WWV serves as the primary mediator through which ENSO integrates cross-basin influences at extended lead times. Conversely, retaining WWV alone within the mode-addition framework is insufficient to sustain extended predictability. Extended skill emerges only when WWV is nonlinearly coupled with additional inter-basin climate modes, with the NPMM exerting a particularly robust influence. These results demonstrate that long-lead ENSO predictability is governed primarily by the nonlinear coupling between WWV and inter-basin climate modes.

Our results further demonstrate that nonlinear cross-basin interactions underpin long-lead ENSO predictability. A first line of evidence for the importance of nonlinearity emerges from the mode-decoupling and mode-addition experiments, as revealed by the systematic comparison between XRO and DESN (Figs. S6 and S7). Although both models are driven by the same set of climate modes, DESN consistently outperforms XRO at long lead times, while their skills remain comparable at short leads. This divergence emerges despite XRO already incorporating multiple climate modes and instead reflects a structural difference: XRO contains only a limited number of prescribed nonlinear terms, whereas DESN can represent a much broader class of nonlinear interactions through its hierarchical reservoir architecture. The superior long-lead performance of DESN therefore provides indirect but robust evidence that higher-order nonlinear interactions contribute substantially to extended ENSO predictability.

A second, independent line of evidence is provided by the pronounced seasonality of forecast skill. We perform additional uninitialized forecast experiments in which the initial conditions of selected inter-basin modes are replaced by climatology while retaining all learned interactions (see methods and Fig. S8). These experiments reveal a strong seasonal modulation of inter-basin memory, consistent with previous findings²⁵. Notably, under specific combinations of initial and target months, removing the initial conditions of Atlantic or Indian Ocean modes can even improve forecast skill, indicating state-dependent and nonlinear interactions that are absent in linear frameworks. Such behavior further supports the view that long-lead predictability arises from nonlinear, seasonally modulated cross-basin influences rather than from linear superposition.

Finally, we obtain strong mechanistic support by introducing an augmented Sparse-Nonlinear XRO (SN-XRO) model (see methods). Starting from the standard XRO formulation, we systematically introduce quadratic coupling terms representing multiplicative interactions between Niño3.4, WWV, and external climate modes, with seasonally modulated coefficients estimated via sparse regression^{46,47}. When trained and evaluated under the same conditions as XRO, SN-XRO behaves

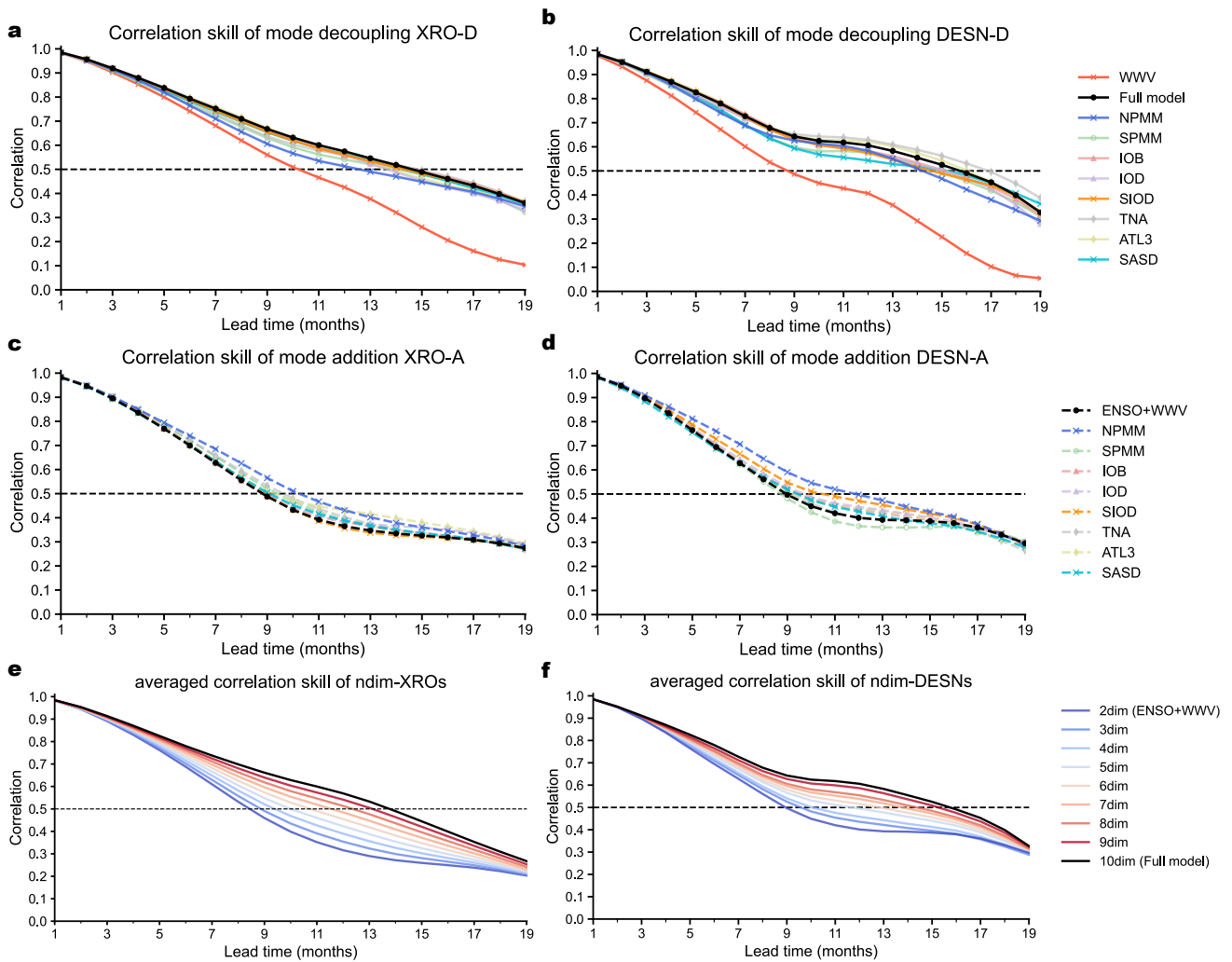


Figure 3. Role of WWV and cross-basin modes in long-lead ENSO predictability. **a,b**, Mode-decoupling experiments for XRO (**a**) and DESN (**b**). Each curve shows the anomaly correlation coefficient (ACC) of Niño 3.4 forecasts as a function of lead time when one climate mode is removed from the full predictor set while all remaining inputs are retained. The black curve denotes the full model, and colored curves correspond to decoupling individual modes (WWV, NPMM, SPMM, IOB, IOD, SIOD, TNA, ATL3, and SASD). Removing WWV leads to the strongest degradation of long-lead skill in both models, while removing other modes generally reduces skill without eliminating extended predictability. The horizontal dashed line marks ACC = 0.5. **c,d**, Mode-addition experiments for XRO (**c**) and DESN (**d**). Forecast skill is shown for a baseline configuration consisting of Niño 3.4 and WWV (black dashed curve), with one additional climate mode added at a time (colored curves). Extended forecast skill emerges only when WWV is coupled with additional inter-basin modes, with the North Pacific Meridional Mode (NPMM) yielding particularly robust improvements. The dashed horizontal line again indicates ACC = 0.5. **e,f**, Increasing input-dimensionality experiments for WWV-selected sub-models in XRO (**e**) and DESN (**f**). Curves show the average forecast skill across ensembles of sub-models with progressively increasing numbers of climate modes coupled to Niño 3.4 and WWV (from 2 to 9 dimensions). Forecast skill increases with input dimensionality, indicating the cumulative contribution of cross-mode interactions.

similarly to the standard XRO at short lead times but exhibits a pronounced slowdown in skill degradation at longer leads (approximately 8–19 months). This delayed improvement closely mirrors the behavior observed in DESN, providing strong supporting evidence that enhanced nonlinear cross-basin coupling underlies the extension of the forecast horizon (see Figs. S9-S10).

Taken together, these results indicate that WWV influences ENSO predictability through slow, nonlinear, and seasonally modulated cross-basin interactions. This interpretation is consistent with established ENSO theory, in which inter-basin influences are primarily mediated by atmospheric teleconnections and subsequently integrated into subsurface ocean memory^{48–50}. The convergence of evidence from model-structure comparisons, seasonal sensitivity experiments, and explicit nonlinear augmentation strongly supports the central role of nonlinearity in sustaining long-lead ENSO predictability.

Intrinsic limits of ENSO predictability

Although their methodological foundations differ, a range of contemporary ENSO prediction frameworks—including the physics-based XRO conceptual model, the 3D-Geoformer deep learning system, and our ESN/DESN frameworks converge on a forecast horizon of roughly 15–20 months. Given that ENSO cycles span 2–7 years, this convergence raises a fundamental question: does the plateau in forecast skill reflect model limitations, or does it represent an intrinsic barrier imposed by ENSO’s nonlinear dynamics?

To further quantify predictability, we analyze the saturation behavior of forecast errors using an absolute-error metric, inspired by the error-growth framework underlying the NLE approach. Specifically, we track the temporal evolution of absolute prediction errors following small perturbations to the initial conditions. Error growth in DESN saturates more slowly than in ESN or XRO (Fig. 4a), confirming enhanced dynamical stability. The practical predictability limit was defined as the time when absolute error reached 95% of its saturation value, estimated consistently across models as the mean absolute error between months 60–100 of the XRO trajectory. Using this criterion, we obtain predictability horizons of approximately 18 months for ESN, 22 months for XRO, and 34 months for DESN. Sensitivity tests further show that reducing the initial perturbation magnitude delays the approach to saturation but yields only limited extensions of the predictability horizon once the initial error norm falls below 0.1 (Fig. 4b), indicating that the estimated limits are robust (Fig. 4b).

These results indicate that the widely cited 15 to 20 month ceiling is not the absolute barrier but rather lies near the intrinsic predictability limit of ~ 30 months. DESN’s ability to approach this barrier arises not from statistical overfitting but from its physically informed design, stabilized internal dynamics, and improved representation of multiscale evolution. More broadly, the analysis also points to concrete pathways for improving forecast skill: reducing errors in initial conditions, incorporating more precise multivariate couplings across ocean–atmosphere modes, and embedding physically consistent constraints to better capture delayed or cross-basin feedbacks.

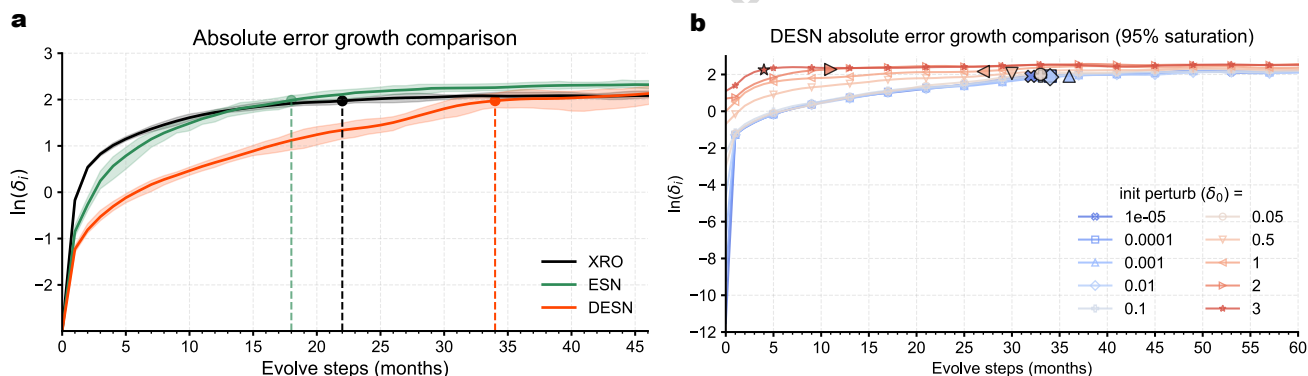


Figure 4. Error-growth analysis and nonlinear predictability limits of ENSO models. **a**, Evolution of absolute forecast error between perturbed and unperturbed trajectories for the XRO (black), ESN (green), and DESN (red) models, shown as $\ln(\delta_t)$ as a function of lead time. Dashed vertical lines mark the predictability limits for XRO (22 months), ESN (18 months), and DESN (34 months), defined as the time when absolute error reaches 95% of the XRO saturation value. **b**, Absolute error growth in DESN for a range of initial perturbation magnitudes δ_0 . While smaller initial perturbations delay the approach to saturation, the long-term saturation level and the inferred predictability limit remain largely unchanged once $\delta_0 < 0.1$.

Discussion

Our results demonstrate that a physics-guided DESN can extend skillful ENSO forecasts well beyond one year, maintaining ACC above 0.5 for 16–20 months. This performance matches or exceeds the skill of operational explainable and state-of-the-art deep learning models, while requiring only seconds of training on a standard CPU. Such efficiency enables systematic exploration of model configurations and physical mechanisms, bridging the gap between predictive skill and mechanistic understanding.

A central contribution of this study is clarifying the mechanisms that sustain long-lead predictability. Submodel experiments confirm that WWV is the dominant precursor of ENSO, consistent with the recharge–discharge paradigm, but also show that its predictive power is amplified by couplings with remote modes such as the NPMM. These delayed and cross-basin feedbacks extend predictability, and DESN’s hierarchical structure is particularly effective in capturing such multiscale interactions.

Despite the improved long-lead forecast skill achieved here, several limitations should be noted. First, the DESN acts as a physics-guided surrogate model and does not explicitly resolve coupled ocean–atmosphere dynamics, meaning that physical causality cannot be directly inferred from the learned representations. Second, model performance depends on the quality

and consistency of the ORAS5 reanalysis used for training, and uncertainties or biases in the underlying data may influence forecast skill. Third, while nonlinear error-growth analysis provides a robust estimate of the predictability horizon, the inferred saturation time represents a practical rather than a strict theoretical limit. Finally, although cross-basin interactions emerge as a key contributor to extended predictability, their detailed dynamical pathways remain to be confirmed through targeted numerical experiments with fully coupled climate models.

From a broader climate perspective, our results highlight physics-guided learning as a surrogate dynamical model, capable of delivering both forecast skill and mechanistic clarity. By showing how WWV-mediated interactions and multiscale feedbacks shape the intrinsic predictability limit of ENSO, this study advances understanding of ENSO as a nonlinear dynamical system. Looking ahead, integrating improved initialization, richer multivariate observations, and coupled Earth system models may further enhance long-lead forecasts. More broadly, this framework illustrates how machine learning can be embedded within dynamical theory to advance prediction and understanding of other critical components of the climate system.

ARTICLE IN PRESS

Methods

Data and preprocessing

We use sea surface temperature (SST) and thermocline depth data from the ORAS5 ocean reanalysis dataset (1958–2023) to construct climate indices relevant to ENSO dynamics. Specifically, we derive ten monthly indices from predefined regions (including the Niño 3.4 region and the equatorial Pacific 20 °C isotherm depth), covering nine ENSO-related climate modes. These modes were previously identified as physically linked to ENSO variability and are also featured in the extended recharge oscillator (XRO) model²⁵. The definitions and regions used to calculate these indices are summarized in Supplementary Table S1.

To isolate ENSO-related variability, we first remove the monthly climatology over the 1979–2010 reference period. We then apply a second-order polynomial detrending to eliminate long-term trends. The resulting standardized anomalies represent interannual variability and are used as model inputs.

Deep Echo State Networks

Echo State Networks (ESNs) leverage the richness of reservoir dynamics to replicate the temporal patterns expressed in training data. Deep Echo State Networks (DESNs), first proposed by Gallicchio et al.^{30,31}, extend classical ESNs by stacking multiple recurrent reservoir layers.

At each time step t , the DESN processes information hierarchically: the first reservoir layer receives external climate inputs $\mathbf{X}_{\text{in},t}$, while subsequent layers take as input the state vector from the preceding layer. Let n_l denote the *number of reservoir layers*, and N_l the *number of neurons in layer l* —both are architectural hyperparameters set prior to training (see Table S2 for specific configurations). We denote the reservoir state of layer l at time t as $\mathbf{r}_t^l \in \mathbb{R}^{N_l}$.

To emulate the stochastic nature of real climate dynamics, we introduce additive white noise ξ_{rc}^l with *noise level* σ_{rc} into each reservoir layer. The state update rule for layer l is given by:

$$\mathbf{r}_{t+1}^l = (1 - \alpha^l)\mathbf{r}_t^l + \alpha^l \tanh(\mathbf{W}_{\text{in}}^l \mathbf{X}_t^l + \mathbf{W}_{\text{res}}^l \mathbf{r}_t^l) + \xi_{\text{rc}}^l, \quad (2)$$

where $\alpha^l \in (0, 1]$ is the *leak rate* controlling the time-delay property of dynamics in layer l , and \tanh is the hyperbolic tangent activation function ensuring bounded neuron activations. The input weight matrix $\mathbf{W}_{\text{in}}^l \in \mathbb{R}^{N_l \times \dim(\mathbf{X}_t^l)}$ is randomly initialized from a uniform distribution and scaled by the *input scaling* parameter σ_{in}^l to control the strength of external inputs. The recurrent weight matrix $\mathbf{W}_{\text{res}}^l \in \mathbb{R}^{N_l \times N_l}$ is constructed as a sparse random matrix with *spectral radius* ρ^l (the largest absolute eigenvalue) and *connection density* d_{rc}^l (the fraction of nonzero elements) to maintain the echo state property²⁸. The layer input \mathbf{X}_t^l is defined hierarchically as:

$$\mathbf{X}_t^l = \begin{cases} \mathbf{X}_{\text{in},t}, & l = 1, \\ \mathbf{r}_t^{l-1}, & l > 1. \end{cases} \quad (3)$$

In this study, the external input $\mathbf{X}_{\text{in},t}$ to the first reservoir layer comprises a comprehensive set of climate indices and seasonal cycles:

$$\mathbf{X}_{\text{in},t} = \begin{bmatrix} \text{ENSO}_t & (\text{ENSO index}) \\ \text{WWV}_t & (\text{Warm Water Volume}) \\ \text{NPMM}_t & (\text{North Pacific Meridional Mode}) \\ \text{SPMM}_t & (\text{South Pacific Meridional Mode}) \\ \text{TNA}_t & (\text{Tropical North Atlantic}) \\ \text{ATL3}_t & (\text{Atlantic Niño}) \\ \text{SASD}_t & (\text{South Atlantic Subtropical Dipole}) \\ \text{IOB}_t & (\text{Indian Ocean Basin}) \\ \text{IOD}_t & (\text{Indian Ocean Dipole}) \\ \text{SIOD}_t & (\text{South Indian Ocean Dipole}) \\ \text{PBS}_t & (\text{Periodic bootstrap sequence. Eq. 1}) \end{bmatrix}, \quad (4)$$

Following reservoir initialization, we apply the Intrinsic Plasticity (IP) learning rule⁵¹ to each reservoir unit to enhance network expressiveness. IP optimizes the internal activation dynamics of individual neurons by adapting unit-specific gain (a) and bias (b) parameters, such that the distribution of activations approximates a target Gaussian distribution with *target mean* μ and *target standard deviation* σ .

This optimization is formulated as minimizing the Kullback–Leibler (KL) divergence between the empirical output distribution and the Gaussian target. For a reservoir unit receiving net input r_{net} , the activation is computed as:

$$\tilde{r} = \tanh(ar_{\text{net}} + b), \quad (5)$$

where the gain a controls the slope of the activation function and the bias b shifts the operating point. At each training step, these parameters are updated according to:

$$\begin{aligned}\Delta b &= -\eta \left(-\frac{\mu}{\sigma^2} + \frac{\tilde{r}}{\sigma^2} (2\sigma^2 + 1 - \tilde{r}^2 + \mu\tilde{r}) \right), \\ \Delta a &= \frac{\eta}{a} + \Delta b \cdot r_{\text{net}},\end{aligned}\quad (6)$$

where η is the *IP learning rate*. This adaptation process ensures that each reservoir unit maintains a high-entropy, information-rich activation profile, thereby improving the overall dynamical capacity and generalization performance of the DESN.

After the IP learning process stabilizes the reservoir dynamics, the output of the DESN is computed by a linear readout layer that combines states from all reservoir layers:

$$\hat{\mathbf{y}}_t = \mathbf{W}_{\text{out}} \begin{bmatrix} \mathbf{r}_t^1 \\ \mathbf{r}_t^2 \\ \vdots \\ \mathbf{r}_t^{n_l} \end{bmatrix}, \quad (7)$$

where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{N_y \times \sum_{l=1}^{n_l} N_l}$ is the readout weight matrix connecting all reservoir layers to the output. Unlike the reservoir weights which remain fixed after initialization, \mathbf{W}_{out} is the only component trained using supervised learning.

We employ ridge regression to determine \mathbf{W}_{out} by minimizing the regularized least-squares objective:

$$L(\mathbf{W}_{\text{out}}) = \|\mathbf{Y} - \mathbf{W}_{\text{out}}\mathbf{R}\|_2^2 + \lambda \|\mathbf{W}_{\text{out}}\|_2^2, \quad (8)$$

where $\mathbf{Y} \in \mathbb{R}^{N_y \times T}$ is the target matrix containing observed climate states, $\mathbf{R} \in \mathbb{R}^{(\sum_{l=1}^{n_l} N_l) \times T}$ is the concatenated reservoir state matrix across all time steps and layers, and $\lambda \geq 0$ is the *regularization parameter* preventing overfitting. This yields the closed-form solution:

$$\mathbf{W}_{\text{out}} = \mathbf{Y}\mathbf{R}^\top (\mathbf{R}\mathbf{R}^\top + \lambda\mathbf{I})^{-1}. \quad (9)$$

In summary, the DESN architecture combines three key components: (1) hierarchical reservoir layers that extract multiscale temporal features with architecture defined by hyperparameters $(n_l, N_l, \alpha^l, \rho^l, d_{\text{rc}}^l, \sigma_{\text{in}}^l, \sigma_{\text{rc}})$ detailed in Table.S2, (2) intrinsic plasticity that optimizes activation distributions with parameters (η, μ, σ) , and (3) a linear readout trained via ridge regression with regularization λ . This design maintains computational efficiency while achieving strong generalization for long-lead climate prediction.

Training and validation periods

We employed three complementary experiments to assess model performance,

For out-of-sample forecasting, the model was trained on data from 1958–1999 and validated on 2002–2023, ensuring that the validation period is temporally independent of the training period. and thus testing the model’s ability to generalize to new periods.

For in-sample forecast, the model was both trained and validated on the 1979–2023 dataset. This experiment evaluates the model’s ability to reproduce known data, providing insight into its fitting capacity.

For Cross-validation, the full 1958–2023 dataset was divided into 11 subsets of 6 consecutive years. For each iteration, 24 consecutive years were withheld for validation, with the remainder used for training. This rolling-block design ensures that all portions of the record contribute to both training and validation, yielding a robust evaluation less dependent on specific temporal divisions.

The Table. 2 summarizes these validation settings.

Table 2. Validation settings for model performance assessment.

Validation method	Training period	Validation period
Out-of-sample forecast	1958–1999	2002–2023
In-sample forecast	1979–2023	1979–2023
Cross-validation	Remaining 7 subsets(e.g., 1958–1975, 2000–2023)	4 consecutive subsets(e.g., 1976–1999)

Prediction skill metrics

Forecast skill was evaluated using two standard metrics: the ACC and the root mean square error (RMSE). The ACC, calculated as the Pearson correlation between forecasts (f) and observations (o),

$$\text{ACC} = \frac{\sum(f - \bar{f})(o - \bar{o})}{\sqrt{\sum(f - \bar{f})^2 \sum(o - \bar{o})^2}}, \quad (10)$$

where \bar{f} and \bar{o} are the means of the forecasts and observations, respectively. The RMSE measures the average forecast error magnitude,

$$\text{RMSE} = \sqrt{\overline{(f - o)^2}}, \quad (11)$$

where the overbar denotes the mean.

Quantitative Experiments for Predictability Source

To rigorously quantify the contribution of individual climate modes and their interactions to ENSO predictability, we conducted four complementary sets of systematic experiments. All experiments use the same training (1958–1999) and validation (2002–2024) periods.

In the mode-decoupling experiments, we consider nine supplementary climate indices beyond Niño 3.4 that are relevant to ENSO dynamics²⁵ (Fig. 1a), forming a ten-variable predictor set in the full model. Mode-decoupling experiments are performed by removing one supplementary mode j at a time from the full predictor set, while keeping all remaining inputs, model architecture, and training procedure unchanged.

The full model, trained using all ten climate indices (Eq. (4)), serves as the control experiment. For example, DESN- D_{NPMM} denotes a DESN experiment in which NPMM is excluded from both training and prediction, with all other inputs and model configurations identical to the full model. Analogous experiments are performed for the XRO framework, denoted as XRO- D_j (e.g., XRO- D_{WWV} , XRO- D_{NPMM}). The difference in forecast skill between the full model and D_j quantifies the sensitivity of ENSO predictability to the removal of climate mode j . Comparing DESN- D_j with XRO- D_j under identical predictor configurations reveals the role of nonlinear interactions in cross-basin interactions.

In the mode-addition experiments, we start from a baseline system composed of Niño 3.4 and WWV, corresponding to the core ENSO recharge–discharge subsystem. Mode-addition experiments are performed by adding one supplementary climate mode j to this baseline, while keeping all other model settings unchanged.

For example, DESN- A_{NPMM} denotes a DESN configuration in which Niño 3.4, WWV, and the North Pacific Meridional Mode (NPMM) are used as inputs. Identical experiments are conducted for the XRO framework, denoted as XRO- A_j (e.g., XRO- A_{NPMM} , XRO- A_{SPMM}).

The difference in forecast skill between A_j and the baseline configuration quantifies the sensitivity of ENSO predictability to the inclusion of climate mode j beyond the core recharge–discharge dynamics. Because XRO is governed primarily by linear dynamics with prescribed coupling structure, systematic skill differences between DESN- A_j and XRO- A_j under identical inputs highlight processes that require nonlinear representations beyond linear superposition.

In the increasing input-dimension experiments, the baseline configuration ($d = 2$) consists of Niño 3.4 and WWV, representing the minimal recharge–discharge system. Higher-dimensional configurations are constructed by sequentially adding supplementary climate modes to this baseline, yielding input dimensions $d = 3, 4, \dots, 10$. For each dimension d , all possible combinations of $d - 2$ modes are selected from the remaining eight supplementary climate indices, while Niño 3.4 and WWV are always retained. Consequently, the $d = 10$ configuration corresponds to the fully coupled model.

For each input dimension, we compute the mean forecast skill by averaging over all corresponding sub-model realizations. This approach isolates the systematic effect of increasing coupling dimensionality, independent of the influence of any single climate mode. Identical increasing-dimension experiments are conducted for both DESN and XRO frameworks.

To investigate state-dependent predictability and the role of initial-condition memory, we performed a set of uninitialized experiments, denoted as U_j . In these experiments, the initial condition of climate mode j is replaced by its climatological mean (zero in standardized units), while all other initial conditions are retained as observed.

Importantly, all uninitialized experiments are conducted using the fully trained DESN model, which is trained with the complete set of coupled climate modes. Thus, the learned interaction structure remains unchanged, and only the initial-state information of mode j is removed. The impact of the initial condition of mode j is quantified by the difference in forecast skill between the control experiment and U_j .

To further disentangle the role of basin-scale memory, we also conducted ocean-level uninitialized experiments by simultaneously replacing the initial conditions of all climate modes within a given ocean basin by climatology. Specifically, $U_{\text{NPMM}+\text{SPMM}}$ removes the initial-state information of both Pacific Meridional Modes, $U_{\text{IOB}+\text{IOD}+\text{SIOD}}$ removes all Indian Ocean modes, and $U_{\text{TNA}+\text{ATL3}+\text{SASD}}$ removes all Atlantic modes.

The resulting differences in forecast skill isolate the contribution of initial-condition information associated with individual modes or entire ocean basins to ENSO predictability. Unlike mode-decoupling experiments, which remove predictors entirely and alter the dynamical system learned during training, uninitialized experiments retain all learned nonlinear interactions and instead diagnose sensitivity to initial-state uncertainty.

Sparse-Nonlinear XRO (SN-XRO) model

For reference, the state vector of XRO is given by

$$\mathbf{X} = (T_{\text{ENSO}}, h_{\text{WWV}}, T_{\text{NPMM}}, T_{\text{SPMM}}, T_{\text{IOB}}, T_{\text{IOD}}, T_{\text{SIOD}}, T_{\text{TNA}}, T_{\text{ATL3}}, T_{\text{SASD}})^T. \quad (12)$$

In the original XRO formulation, the system evolves according to

$$\frac{d\mathbf{X}}{dt} = \mathbf{L}\mathbf{X} + \mathbf{N}_{\text{XRO}}(\mathbf{X}), \quad (13)$$

where the nonlinear terms are restricted to

$$\mathbf{N}_{\text{XRO}}(\mathbf{X}) = \begin{pmatrix} b_1 T_{\text{ENSO}}^2 + b_2 T_{\text{ENSO}} h_{\text{WWV}} \\ 0 \\ 0 \\ 0 \\ b_3 T_{\text{IOD}}^2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (14)$$

Thus, nonlinear interactions in XRO are confined to ENSO asymmetry, ENSO–WWV recharge–discharge coupling, and a conditional quadratic contribution associated with the Indian Ocean Dipole.

In the SN-XRO model, we extend the nonlinear tendency by introducing additional quadratic coupling terms that involve ENSO and WWV:

$$\mathbf{N}_{\text{SN}}(\mathbf{X}) = \mathbf{N}_{\text{XRO}}(\mathbf{X}) + \begin{pmatrix} \mathbf{N}_1 T_{\text{ENSO}} \mathbf{T}_M \\ \mathbf{N}_2 h_{\text{WWV}} \mathbf{T}_M \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (15)$$

where

$$\mathbf{T}_M = (T_{\text{NPMM}}, T_{\text{SPMM}}, T_{\text{IOB}}, T_{\text{IOD}}, T_{\text{SIOD}}, T_{\text{TNA}}, T_{\text{ATL3}}, T_{\text{SASD}})^T, \quad (16)$$

denotes the set of external climate modes. No nonlinear interaction terms are introduced among the external climate modes themselves.

Here, \mathbf{N}_1 and \mathbf{N}_2 are seasonally modulated coefficient vectors controlling the nonlinear coupling between ENSO, WWV, and inter-basin climate modes. Each coefficient is parameterized as a truncated Fourier series,

$$n(t) = n_0 + n_{1s} \sin(\omega t) + n_{1c} \cos(\omega t) + n_{2s} \sin(2\omega t) + n_{2c} \cos(2\omega t), \quad (17)$$

where $\omega = 2\pi/12$ represents the annual cycle.

The Sparse Identification of Nonlinear Dynamical systems (SINDy) method^{46,47} is applied to fit the parameters of the prescribed quadratic interaction terms using ORAS5 data over the training period (1979–2001). This procedure automatically eliminates statistically insignificant nonlinear terms, yielding a parsimonious representation of nonlinear coupling consistent with the data. The resulting seasonally modulated nonlinear coupling parameters are shown in Fig. S9.

To ensure a controlled comparison between XRO and SN-XRO, model parameters are estimated in two successive steps.

First, the standard XRO model is fitted using the SINDy framework, including all linear coefficients L_{ij} and the prescribed nonlinear terms $b_1 T_{\text{ENSO}}^2$, $b_2 T_{\text{ENSO}} h_{\text{WV}}$, and $b_3 T_{\text{IOD}}^2$. This step yields a physically consistent baseline representation of ENSO and its linear cross-basin couplings.

In the SN-XRO model, all linear coefficients L_{ij} are fixed to the values obtained from the XRO fitting and remain unchanged. Additional nonlinear terms are then estimated using SINDy.

This hierarchical fitting strategy ensures that differences between XRO and SN-XRO arise solely from enhanced nonlinear coupling involving ENSO and WV, rather than from changes in linear dynamics.

SN-XRO is integrated deterministically and evaluated over the same out-of-sample period (2002–2024) as XRO and DESN, ensuring that differences in forecast performance arise solely from the inclusion of physically constrained nonlinear interaction pathways.

By construction, SN-XRO provides a minimal nonlinear extension of XRO that allows us to directly test whether enhanced ENSO–WV–inter-basin coupling is sufficient to reproduce the delayed error growth and extended predictability identified in DESN.

Computation of Nonlinear Local Lyapunov Exponents

The XRO, ESN, and DESN are all autonomous dynamical systems whose forward evolution can be expressed as $\mathbf{z}_{t+1} = \mathbf{F}(\mathbf{z}_t)$, where \mathbf{z}_t is the full state vector and \mathbf{F} is the evolution operator. The state space differs across models:

$$\mathbf{z}_t = \begin{cases} \mathbf{X}_t \in \mathbb{R}^{10}, & \text{XRO} \\ [\mathbf{r}_t; \mathbf{X}_t] \in \mathbb{R}^{N+10}, & \text{ESN } (N = 20000) \\ [\mathbf{r}_t^1; \mathbf{r}_t^2; \mathbf{X}_t] \in \mathbb{R}^{N_1+N_2+10}, & \text{DESN } (N_1 = 20000, N_2 = 12000) \end{cases} \quad (18)$$

where $\mathbf{X}_t \in \mathbb{R}^{10}$ denotes the 10 climate indices, and \mathbf{r}_t , \mathbf{r}_t^1 , \mathbf{r}_t^2 are the reservoir neuron states.

Starting from an initial climate state \mathbf{X}_0 on the attractor, we introduce a small perturbation to the 10 climate indices: $\mathbf{X}'_0 = \mathbf{X}_0 + \boldsymbol{\delta}_0$, where $\boldsymbol{\delta}_0$ is a random unit vector scaled by initial amplitude ϵ_0 . For ESN/DESN, the corresponding reservoir states are initialized from the unperturbed climate state, ensuring perturbations only affect the climate variables. Both the reference trajectory \mathbf{X}_i and perturbed trajectory \mathbf{X}'_i are evolved independently using the full nonlinear dynamics for $T = 100$ months.

We test a range of initial perturbation amplitudes ϵ_0 . For each amplitude, we perform 500 realizations with different random perturbation directions, sampled from 500 initial conditions uniformly distributed across the test period (2002–2023).

The error at step i is defined as:

$$\delta_i = \|\mathbf{X}'_i - \mathbf{X}_i\|_2. \quad (19)$$

where $\|\cdot\|_2$ denotes the Euclidean norm (L2 norm). The absolute error growth is quantified by the logarithmic error:

$$E_i = \ln \delta_i. \quad (20)$$

To define a consistent saturation threshold across models with different error growth rates, we compute the mean logarithmic error over months 60–100 of XRO evolution:

$$E_{\text{sat}} = \frac{1}{40} \sum_{i=60}^{100} E_{\text{XRO},i}. \quad (21)$$

The predictability horizon T_p is defined as the first time when the 95 percentile of E_{sat} is reached:

$$T_p = \min \{i \mid \langle E_i \rangle \geq \gamma E_{\text{sat}}\}, \quad (22)$$

where $\langle \cdot \rangle$ denotes average over all realization, and $\gamma = 0.95$ is the saturation threshold.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 42575057, T2525011, 42450183, 12275020, 12135003, 12205025, 42461144209, 62333002), the Ministry of Science and Technology of China (2023YFE0109000), the National Key R&D Program of China (2025YFF0517203, 2025YFF0517304), and State Key Laboratory of Information Photonics and Optical Communications (IPOC2023ZJ02). J.F. is supported by the Fundamental Research Funds for the Central Universities.

References

1. Chen, D., Zebiak, S. E., Busalacchi, A. J. & Cane, M. A. An improved procedure for El Niño forecasting: Implications for predictability. *Science* **269**, 1699–1702, DOI: [10.1126/science.269.5231.1699](https://doi.org/10.1126/science.269.5231.1699) (1995).
2. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an integrating concept in earth science. *Science* **314**, 1740–1745 (2006).
3. Timmermann, A. *et al.* El Niño–Southern Oscillation complexity. *Nature* **559**, 535–545, DOI: [10.1038/s41586-018-0394-2](https://doi.org/10.1038/s41586-018-0394-2) (2018).
4. Zhang, R.-H., Rothstein, L. M. & Busalacchi, A. J. Origin of upper-ocean warming and El Niño change on decadal scales in the tropical Pacific ocean. *Nature* **391**, 879–883, DOI: [10.1038/36081](https://doi.org/10.1038/36081) (1998).
5. Cai, W. *et al.* Changing El Niño–Southern Oscillation in a warming climate. *Nat. Rev. Earth & Environ.* DOI: [10.1038/s43017-021-00199-z](https://doi.org/10.1038/s43017-021-00199-z) (2021).
6. Dijkstra, H. A. *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*. Atmospheric and Oceanographic Sciences Library (Springer Netherlands, 2005), 2 edn.
7. Suarez, M. J. & Schopf, P. S. A delayed action oscillator for ENSO. *J. Atmospheric Sci.* **45**, 3283–3287, DOI: [10.1175/1520-0469\(1988\)045<3283:ADAOFE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2) (1988).
8. Battisti, D. S. & Hirst, A. C. Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry and nonlinearity. *J. Atmospheric Sci.* **46**, 1687–1712, DOI: [10.1175/1520-0469\(1989\)046<1687:IVIATA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<1687:IVIATA>2.0.CO;2) (1989).
9. Bjerknes, J. Atmospheric teleconnections from the equatorial Pacific. *Mon. Weather. Rev.* **97**, 163–172, DOI: [10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2) (1969).
10. Cane, M. A. & Zebiak, S. E. A theory for El Niño and the southern oscillation. *Science* **228**, 1085–1087, DOI: [10.1126/science.228.4703.1085](https://doi.org/10.1126/science.228.4703.1085) (1985).
11. Cane, M. A., Zebiak, S. E. & Dolan, S. C. Experimental forecasts of El Niño. *Nature* **321**, 827–832, DOI: [10.1038/321827a0](https://doi.org/10.1038/321827a0) (1986).
12. Jin, F.-F. An equatorial ocean recharge paradigm for ENSO. part I: Conceptual model. *J. Atmospheric Sci.* **54**, 811–829, DOI: [10.1175/1520-0469\(1997\)054<0811:AEORPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2) (1997).
13. Jin, F.-F., Lin, L., Timmermann, A. & Zhao, J. Ensemble-mean dynamics of the ENSO recharge oscillator under state-dependent stochastic forcing. *Geophys. Res. Lett.* **34**, DOI: [10.1029/2006GL027372](https://doi.org/10.1029/2006GL027372) (2007).
14. Barnston, A. G., Tippett, M. K., L’Heureux, M. L., Li, S. & DeWitt, D. G. Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Am. Meteorol. Soc.* **93**, 631–651, DOI: [10.1175/BAMS-D-11-00111.1](https://doi.org/10.1175/BAMS-D-11-00111.1) (2012).
15. Guilyardi, E., Capotondi, A., Lengaigne, M., Thual, S. & Wittenberg, A. T. ENSO modeling: History, progress, and challenges. In McPhaden, M. J., Santoso, A. & Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*, 199–226, DOI: [10.1002/9781119548164.ch9](https://doi.org/10.1002/9781119548164.ch9) (American Geophysical Union, Washington, DC, 2020).
16. Izumo, T. *et al.* Influence of the state of the Indian Ocean Dipole on the following year’s El Niño. *Nat. Geosci.* **3**, 168–172, DOI: [10.1038/ngeo760](https://doi.org/10.1038/ngeo760) (2010).
17. Larson, S. M. & Kirtman, B. P. The Pacific meridional mode as an ENSO precursor and predictor in the North American Multimodel Ensemble. *J. Clim.* **27**, 7018–7032, DOI: [10.1175/JCLI-D-14-00055.1](https://doi.org/10.1175/JCLI-D-14-00055.1) (2014).
18. Zhang, H., Clement, A. & Di Nezio, P. The South Pacific Meridional Mode: A mechanism for ENSO-like variability. *J. Clim.* **27**, 769–783, DOI: [10.1175/JCLI-D-13-00082.1](https://doi.org/10.1175/JCLI-D-13-00082.1) (2014).
19. Ham, Y.-G., Kug, J.-S., Park, J.-Y. & Jin, F.-F. Sea surface temperature in the north tropical Atlantic as a trigger for El Niño/Southern Oscillation events. *Nat. Geosci.* **6**, 112–116, DOI: [10.1038/ngeo1686](https://doi.org/10.1038/ngeo1686) (2013).
20. Ding, H., Keenlyside, N. S. & Latif, M. Impact of the equatorial Atlantic on the El Niño Southern oscillation. *Clim. Dyn.* **38**, 1965–1972, DOI: [10.1007/s00382-011-1097-y](https://doi.org/10.1007/s00382-011-1097-y) (2012).
21. Rodríguez-Fonseca, B. *et al.* Are Atlantic Niños enhancing Pacific ENSO events in recent decades? *Geophys. Res. Lett.* **36**, L20705, DOI: [10.1029/2009GL040048](https://doi.org/10.1029/2009GL040048) (2009).
22. Cai, W. *et al.* Pantropical climate interactions. *Science* **363**, eaav4236 (2019).

23. Ludescher, J. *et al.* Improved El Niño forecasting by cooperativity detection. *Proc. Natl. Acad. Sci.* **110**, 11742–11745 (2013).
24. Meng, J. *et al.* Complexity-based approach for El Niño magnitude forecasting before the spring predictability barrier. *Proc. Natl. Acad. Sci.* **117**, 177–183 (2020).
25. Zhao, S. *et al.* Explainable El Niño predictability from climate mode interactions. *Nature* **630**, 891–898, DOI: [10.1038/s41586-024-07534-6](https://doi.org/10.1038/s41586-024-07534-6) (2024).
26. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572, DOI: [10.1038/s41586-019-1559-7](https://doi.org/10.1038/s41586-019-1559-7) (2019).
27. Zhou, L. & Zhang, R.-H. A self-attention-based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Sci. Adv.* **9**, eadf2827, DOI: [10.1126/sciadv.adf2827](https://doi.org/10.1126/sciadv.adf2827) (2023).
28. Jaeger, H. & Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80, DOI: [10.1126/science.1091277](https://doi.org/10.1126/science.1091277) (2004).
29. Jaeger, H. Echo state network. *Scholarpedia* **2**, 2330, DOI: [10.4249/scholarpedia.2330](https://doi.org/10.4249/scholarpedia.2330) (2007).
30. Gallicchio, C., Micheli, A. & Pedrelli, L. Deep reservoir computing: A critical experimental analysis. *Neurocomputing* **268**, 87–99, DOI: [10.1016/j.neucom.2016.12.089](https://doi.org/10.1016/j.neucom.2016.12.089) (2017).
31. Gallicchio, C., Micheli, A. & Pedrelli, L. Design of deep echo state networks. *Neural Networks* **108**, 33–47, DOI: [10.1016/j.neunet.2018.08.002](https://doi.org/10.1016/j.neunet.2018.08.002) (2018).
32. Li, X., Ding, R. & Li, J. A new technique to quantify the local predictability of extreme events: The backward nonlinear local lyapunov exponent method. *Front. Environ. Sci.* **10**, 825233 (2022).
33. Hou, Z. *et al.* Asymmetry of the predictability limit of the warm ENSO phase. *Geophys. Res. Lett.* **45**, 12947–12955, DOI: [10.1029/2018GL077880](https://doi.org/10.1029/2018GL077880) (2018).
34. Trenberth, K. E. The definition of el nino. *Bull. Am. Meteorol. Soc.* **78**, 2771–2778 (1997).
35. Jin, F.-F. An equatorial ocean recharge paradigm for enso. part i: Conceptual model. *J. atmospheric sciences* **54**, 811–829 (1997).
36. Chiang, J. C. & Vimont, D. J. Analogous pacific and atlantic meridional modes of tropical atmosphere–ocean variability. *J. Clim.* **17**, 4143–4158 (2004).
37. Zhang, H., Clement, A. & Di Nezio, P. The south pacific meridional mode: A mechanism for enso-like variability. *J. Clim.* **27**, 769–783 (2014).
38. Jin, Y. *et al.* The indian ocean weakens the enso spring predictability barrier: role of the indian ocean basin and dipole modes. *J. Clim.* **36**, 8331–8345 (2023).
39. Jo, H.-S. *et al.* Southern indian ocean dipole as a trigger for central pacific el niño since the 2000s. *Nat. Commun.* **13**, 6965 (2022).
40. Ham, Y.-G., Kug, J.-S., Park, J.-Y. & Jin, F.-F. Sea surface temperature in the north tropical atlantic as a trigger for el niño/southern oscillation events. *Nat. Geosci.* **6**, 112–116 (2013).
41. Ham, Y.-G., Kug, J.-S. & Park, J.-Y. Two distinct roles of atlantic ssts in enso variability: North tropical atlantic sst and atlantic niño. *Geophys. Res. Lett.* **40**, 4012–4017 (2013).
42. Ham, Y.-G. *et al.* Inter-basin interaction between variability in the south atlantic ocean and the el niño/southern oscillation. *Geophys. Res. Lett.* **48**, e2021GL093338 (2021).
43. McPhaden, M. J. A 21st century shift in the relationship between ENSO SST and warm water volume anomalies. *Geophys. Res. Lett.* **39**, L09706, DOI: [10.1029/2012GL051826](https://doi.org/10.1029/2012GL051826) (2012).
44. Jin, Y., Liu, Z., Lu, Z. & He, C. Seasonal cycle of background in the tropical pacific as a cause of enso spring persistence barrier. *Geophys. Res. Lett.* **46**, 13371–13378, DOI: [10.1029/2019GL085205](https://doi.org/10.1029/2019GL085205) (2019).
45. Stuecker, M. F., Timmermann, A., Jin, F.-F., McGregor, S. & Ren, H.-L. A combination mode of the annual cycle and the el niño/southern oscillation. *Nat. Geosci.* **6**, 540–544, DOI: [10.1038/ngeo1826](https://doi.org/10.1038/ngeo1826) (2013).
46. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **113**, 3932–3937, DOI: [10.1073/pnas.1517384113](https://doi.org/10.1073/pnas.1517384113) (2016).
47. de Silva, B. M. *et al.* Pysindy: A python package for the sparse identification of nonlinear dynamical systems from data. *J. Open Source Softw.* **5**, 2104, DOI: [10.21105/joss.02104](https://doi.org/10.21105/joss.02104) (2020).

48. Dijkstra, H. A. *Nonlinear physical oceanography: a dynamical systems approach to the large scale ocean circulation and El Niño*, vol. 532 (Springer, 2005).
49. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. Enso as an integrating concept in earth science. *science* **314**, 1740–1745 (2006).
50. Timmermann, A. *et al.* El niño–southern oscillation complexity. *Nature* **559**, 535–545 (2018).
51. Jaeger, H., Lukoševičius, M., Popovici, D. & Siewert, U. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks* **20**, 335–352, DOI: [10.1016/j.neunet.2007.04.016](https://doi.org/10.1016/j.neunet.2007.04.016) (2007).

Data availability

Datasets and models used in this paper are freely available. ORAS5: <https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview>; XRO model ENSO forecast: <https://github.com/senclimate/XRO>; 3D-Geoformer model ENSO forecast: <https://msdc.qdio.ac.cn/data/metadata-special-detail?id=1602252663859298305&otherId=1602252664035459074>; NMME: <https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/>; IRI ENSO forecast: <https://iri.columbia.edu/our-expertise/climate/forecasts/enso/current/>; Niño3.4 Index from the HadISST1.1: <https://psl.noaa.gov/data/timeseries/month/DS/Nino34/>.

Code availability

The Python codes used for the analysis are available on GitHub (<https://github.com/zhangzejing/RC-ENSO>).

Author Contributions

J.M, J.X and J.F designed the research. Z.Z performed the analysis. Z.Z, J.M, Z.Q, W.D, J.G, Z.Y, J.X, X.C, W.C, J.K, S.H and J.F prepared the manuscript, discussed results, and contributed to writing the manuscript. J.M and J.F led the writing of the manuscript.

Additional information

Supplementary Information is available in the online version of the paper.

Competing interests

The authors declare no competing financial or non-financial interests.

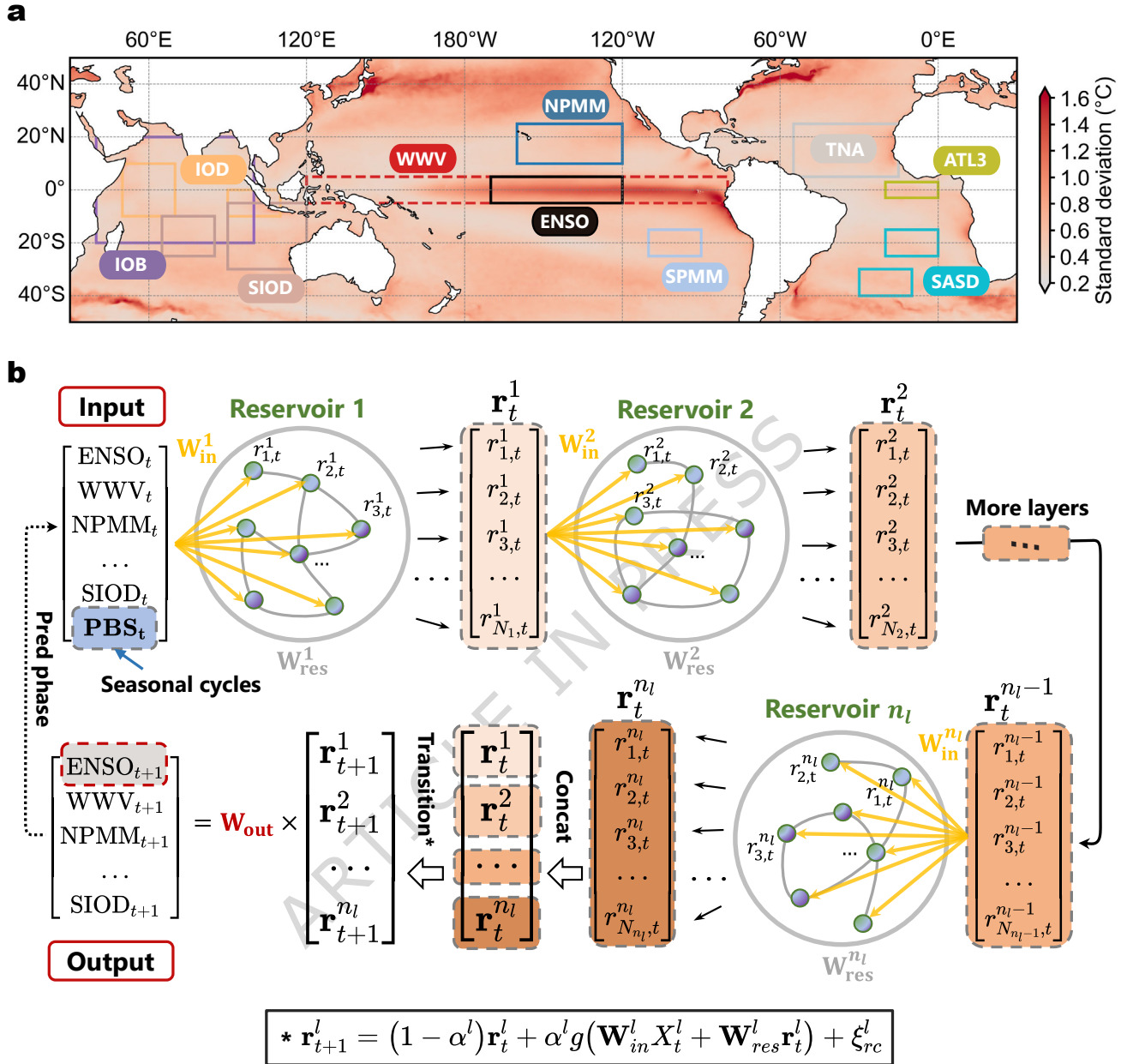


Figure 1. Regions of interest and architecture of the DESN model for ENSO prediction. **a**, Standard deviation of sea surface temperature anomalies (SSTA) from the ORAS5 reanalysis (1979–2023). Colored boxes denote regions used to compute area-averaged climate indices for ENSO and associated modes, following the XRO framework²⁵. **b**, Schematic of the Deep Echo State Network (DESN) architecture. The input layer comprises multiple climate indices and seasonal cycles as periodic bootstrap sequences (\mathbf{PBS}_t). Inputs are passed through a hierarchy of n_l recurrent reservoirs, each generating a neuron state vector \mathbf{r}_t^l . Additive white noise ξ_{rc}^l is injected into each reservoir to emulate stochastic variability. Internal weights consist of input matrices \mathbf{W}_{in}^l and reservoir matrices \mathbf{W}_{res}^l . Reservoir outputs are concatenated and mapped to next-month targets via a linear readout matrix \mathbf{W}_{out} . Multi-step forecasts are obtained recursively using rolling inputs. When $n_l = 1$, the model reduces to a standard ESN. The reservoir transition function at the bottom illustrates neuron dynamics with activation function $g(\cdot)$ and leakage rate α . See Methods for definitions and implementation details.

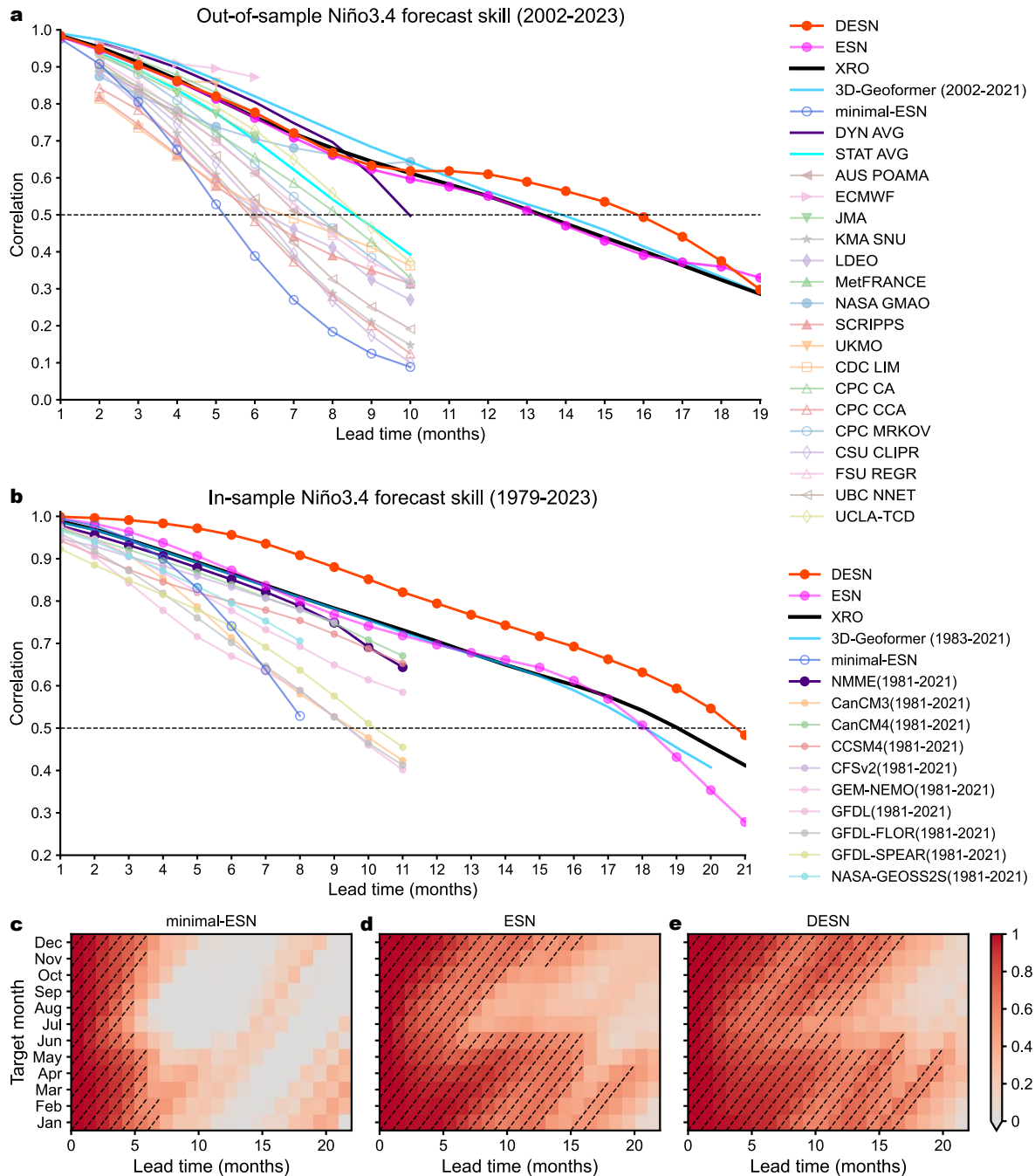


Figure 2. Forecast performance of ESN and DESN models for ENSO prediction. **a**, Forecast correlation skill (ACC) of the 3-month running mean Niño3.4 index as a function of lead time, evaluated out-of-sample for 2002–2023. Shown are DESN (red), ESN (magenta), minimal-ESN (dark blue), XRO (black; trained on 1958–1999), 3D-Geoformer (light blue), and the IRI operational ensemble forecasts—ensemble means of dynamical models (dark purple) and statistical models (dark cyan). **b**, Same as panel **a**, but evaluated in-sample over 1979–2023, including the ensemble mean of dynamical models from the North American Multi-Model Ensemble (NMME). Individual NMME model forecasts (1981–2021) are shown in various colors. Forecast periods for the 3D-Geoformer and NMME correspond to their respective training spans. **c–e**, Target-month-dependent ACC of Niño3.4 forecasts for the minimal-ESN (**c**), ESN (**d**), and DESN (**e**). Colors denote correlation skill as a function of target month (vertical axis) and lead time (horizontal axis). Black contours delineate regions where ACC exceeds 0.5, highlighting combinations of target months and lead times with high predictive skill.

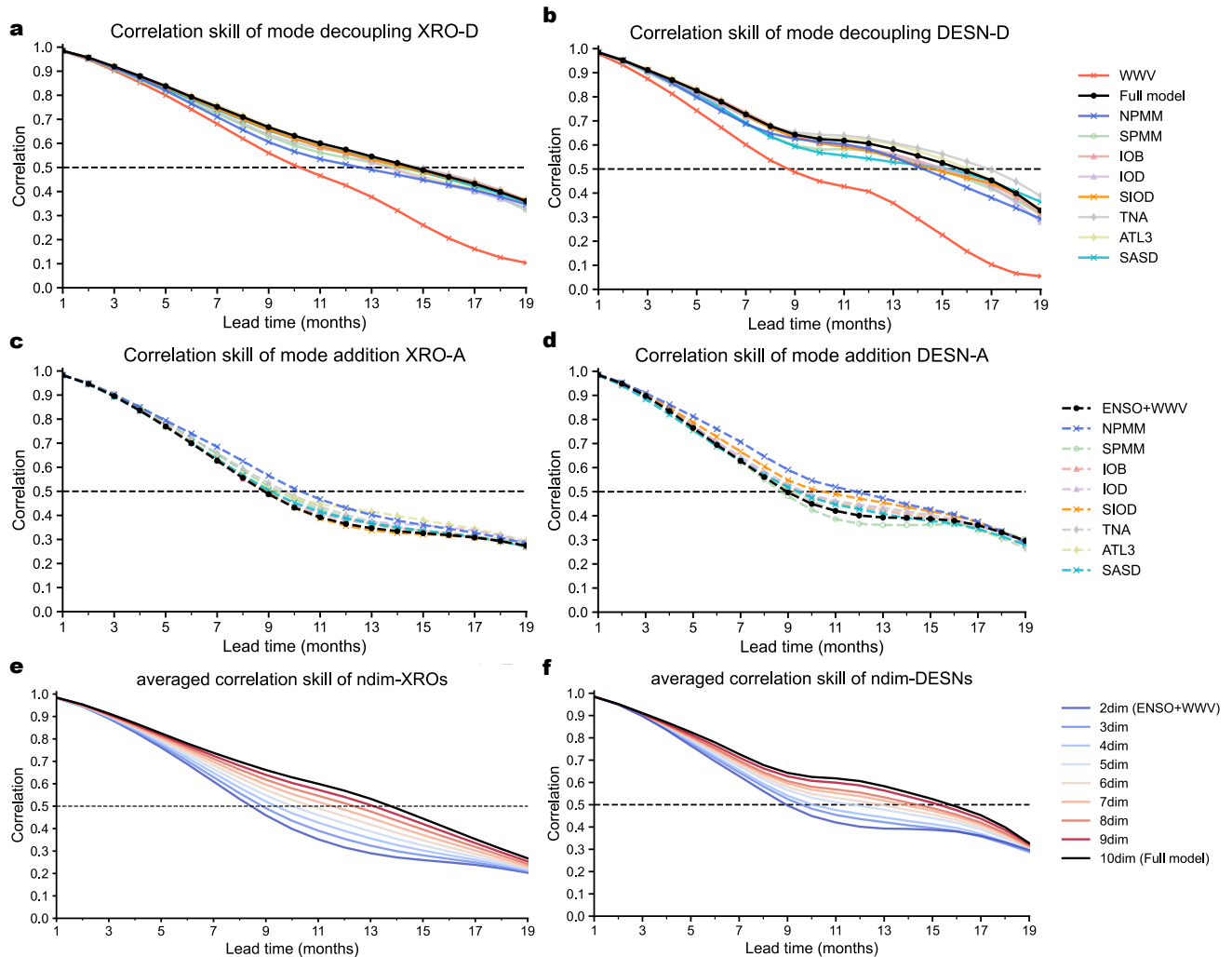


Figure 3. Role of WWV and cross-basin modes in long-lead ENSO predictability. **a,b**, Mode-decoupling experiments for XRO (**a**) and DESN (**b**). Each curve shows the anomaly correlation coefficient (ACC) of Niño 3.4 forecasts as a function of lead time when one climate mode is removed from the full predictor set while all remaining inputs are retained. The black curve denotes the full model, and colored curves correspond to decoupling individual modes (WWV, NPMM, SPMM, IOB, IOD, SIOD, TNA, ATL3, and SASD). Removing WWV leads to the strongest degradation of long-lead skill in both models, while removing other modes generally reduces skill without eliminating extended predictability. The horizontal dashed line marks ACC = 0.5. **c,d**, Mode-addition experiments for XRO (**c**) and DESN (**d**). Forecast skill is shown for a baseline configuration consisting of Niño 3.4 and WWV (black dashed curve), with one additional climate mode added at a time (colored curves). Extended forecast skill emerges only when WWV is coupled with additional inter-basin modes, with the North Pacific Meridional Mode (NPMM) yielding particularly robust improvements. The dashed horizontal line again indicates ACC = 0.5. **e,f**, Increasing input-dimensionality experiments for WWV-selected sub-models in XRO (**e**) and DESN (**f**). Curves show the average forecast skill across ensembles of sub-models with progressively increasing numbers of climate modes coupled to Niño 3.4 and WWV (from 2 to 9 dimensions). Forecast skill increases with input dimensionality, indicating the cumulative contribution of cross-mode interactions.

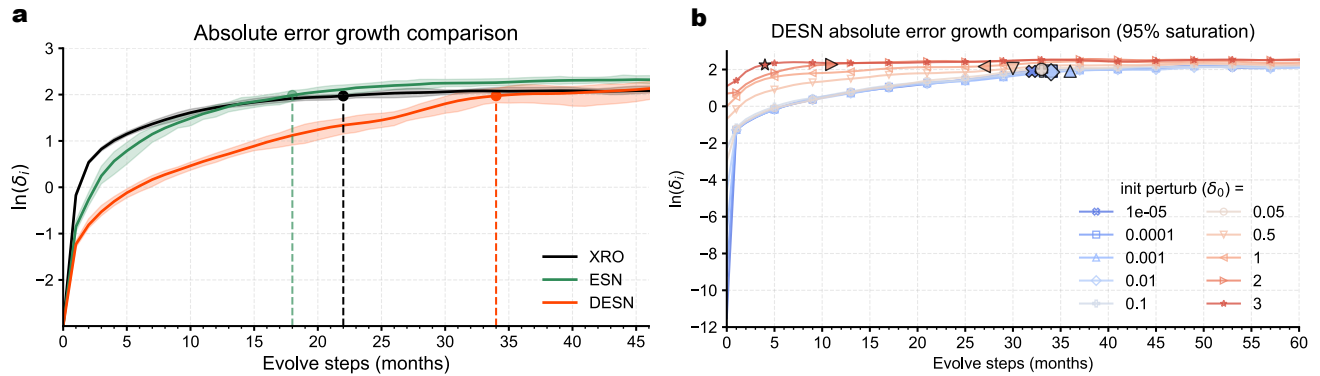


Figure 4. Error-growth analysis and nonlinear predictability limits of ENSO models. **a**, Evolution of absolute forecast error between perturbed and unperturbed trajectories for the XRO (black), ESN (green), and DESN (red) models, shown as $\ln(\delta_i)$ as a function of lead time. Dashed vertical lines mark the predictability limits for XRO (22 months), ESN (18 months), and DESN (34 months), defined as the time when absolute error reaches 95% of the XRO saturation value. **b**, Absolute error growth in DESN for a range of initial perturbation magnitudes δ_0 . While smaller initial perturbations delay the approach to saturation, the long-term saturation level and the inferred predictability limit remain largely unchanged once $\delta_0 < 0.1$.

Table 1. Comparison of training efficiency across ENSO forecasting models. Estimated training time and computational requirements for the DESN, ESN, XRO, and 3D-Geoformer models. ^a Out-of-sample evaluation (2002–2023); ^b Out-of-sample (1983–2021); ^c In-sample (1983–2023).

Models	Training data (size)	Training time (algorithms)	Training on	Months lead
3D-Geoformer ²⁷	Gridded monthly sea surface	~12 h	GPU	13 ^a
	wind stress and 3D ocean temperature anomalies (84 GB)	(Backpropagation)	(Nvidia V100)	18 ^b
XRO ²⁵	monthly climate indices (110 kB)	4 s (Linear-regression)	CPU (Intel i9-14900K)	13 ^a 18 ^c
ESN	monthly climate indices (110 kB)	26 s (Ridge-regression)	CPU (Intel i9-14900K)	13 ^a 18 ^c
DESN (this study)	monthly climate indices (110 kB)	122 s (Ridge-regression)	CPU (Intel i9-14900K)	16^a 20^c